ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ И ЕЕ ПРИЛОЖЕНИЯ

Т. СААТИ

ELEMENTS OF QUEUEING THEORY

With Applications

THOMAS L. SAATY
Office of Naval Research

McGRAW—HILL BOOK COMPANY, INC.
New York Toronto London 1961

ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ И ЕЕ ПРИЛОЖЕНИЯ

Т. Л. СААТИ

Перевод с английского *Е. Г. Коваленко*

Под редакцией *И. Н. Коваленко* и *Р. Д. Когана*

Предисловие академика АН УССР *Б. В. Гнеденко*

ИЗДАТЕЛЬСТВО „СОВЕТСКОЕ РАДИО"

МОСКВА—1965

УДК 519.152

В книге дается детальное изложение одного из важнейших разделов исследования операций — теории массового обслуживания. Широко освещаются основные задачи теории в тесной связи с приложениями к телефонии, движению транспортных потоков, управлению запасами и другим областям практической деятельности. Приводится обширная библиография.

Книга является фундаментальным руководством для широкого круга специалистов, занимающихся как теорией, так и практическими приложениями теории массового обслуживания.

Предисловие к русскому изданию

Советский читатель знаком с первой переведенной и изданной у нас книгой Томаса Саати «Математические методы исследования операций» (Воениздат, М., 1963).

Новая книга Саати, предлагаемая в русском переводе, посвящена детальному изложению одного из важнейших разделов исследования операций, получившего у нас наименование теории массового обслуживания, а в странах английского языка — теории очередей.

Теория массового обслуживания в последние годы развивалась исключительно быстро. За это время не только существенно продвинулось решение вопросов, возникших в период зарождения теории массового обслуживания, и появилось большое число новых постановок проблем, но и были также разработаны некоторые общие приемы решения широких классов задач, а также осмыслены специфические особенности самой теории. Более того, выяснилось, что постановки задач теории массового обслуживания имеют весьма широкий характер, поскольку к ним приводят почти все направления современной практики. Естественно, что расширение поля приложений неизбежно привело к увеличению числа исследователей и к расширению того круга изданий, в которых появляются работы интересующего нас направления. Лицо, пожелавшее теперь ознакомиться с задачами и методами теории массового обслуживания по журнальным статьям, вынуждено обратиться к научным изданиям, посвященным математике, транспорту, медицине, экономике, военному делу, промышленности, сельскому хозяйству, связи, физике и многим другим областям знания. Вот почему сейчас возникает настоятельная необходимость в издании как хороших обзоров, подводящих итог определенному этапу развития теории и практики применения, так и аннотированных библиографий.

Среди вышедших за последние годы монографий по теории массового обслуживания книга Саати заслуживает особого внимания по нескольким причинам. Во-первых, в ней достаточно широко изложены основные задачи теории, во-вторых, эти задачи тесно увязаны с практическими вопросами ряда направлений деятельности, в-третьих, математические методы теории изложены удачно и, в-четвертых, книга снабжена превосходно составленной библиографией, пополненной к тому же переводчиком. Указанные обстоятельства приведут к тому, что книга окажет существенную помощь читателям в их повседневной работе.

Понятно, что издание книги Саати на русском языке не снимает необходимости написания новых книг на ту же тему, которые осветили бы теорию массового обслуживания с разных сторон и способствовали бы как прогрессу теории и методологии науки, так и расширению поля ее приложений. Это тем более необходимо сделать, что ряд новых постановок задач, интересных с теоретической и прикладной точек зрения, возник уже после выхода в свет книги Саати. Несколько слов о такого рода вопросах, а также о вопросах, совсем еще не затронутых в литературе, будет мной сказано позднее.

Теперь, после того как дана общая оценка предлагаемой советскому читателю книги, можно перейти к краткому обзору современного состояния теории, изложению которой она посвящена.

В последние десятилетия развитие науки, техники, экономики, средств связи и транспорта привело к необходимости иметь дело с большими системами, обладающими определенной целостностью, хотя и состоящими из огромного числа частей, использование которых находится под влиянием иррегулярных воздействий. Особенности такого рода систем, естественно, требуют особого подхода к их изучению, а также специальных приемов управления ими. К таким системам можно отнести телефонный узел, крупное предприятие, систему противовоздушной обороны, пароходство или же крупный аэропорт. Естественно, что во всех этих системах возникают многочисленные задачи, отражающие специфику конкретной системы, а также тех целей, которые ставит перед собой руководство этой системой, если речь идет о системах, руководимых человеком. Однако в этом огромном разнообразии индивидуальных особенностей выявляются и задачи общего характера, свойственные многим сложным системам самой разнообразной природы. Некоторые из этих задач и создали базу теории массового обслуживания.

В качестве примера рассмотрим работу крупного морского порта. В порт прибывают суда из различных стран. Эти суда необходимо обработать, т. е. разгрузить, погрузить, произвести необходимый ремонт, снабдить горючим, вспомогательными материалами и т. д. Все это необходимо организовать так, чтобы добиться максимальной экономической эффективности. Из всего многообразия аспектов работы порта остановимся лишь на одном: погрузке и выгрузке судов. Суда в порт прибывают из многих портов мира, моменты их прихода не могут быть точно определены, поскольку для грузовых судов имеется огромное количество причин, в силу которых отклоняются от графика как моменты их прихода в порты и моменты их отправления, так и длительность их обработки. Чтобы проиллюстрировать, насколько значительное влияние оказывают различного рода причины на нарушение графика прихода судов в порт назначения, приведем две выборки из графиков, которые составляются заранее для планирования работы как порта, так и судов. Данные взяты за один из месяцев 1962 г. и за тот же месяц 1963 г. и заимствованы из статьи Б. В. Гнеденко и М. Н. Зубкова [990].

Таблица 1

График прихода судов и его выполнение

Название судна	Дата прихода в порт по графику	Дата фактического прихода в порт
	1962 год.	
„Авамери“	От 8 до 20	23
„Либерейтор“	5 или 6	Не пришло
„Маруба“	От 25 текущего месяца до 8 следующего	Не пришло
„Атлантида“	28	Не пришло
„Ив. Ползунов“	12	25
„Пулково“	Не ожидалось	27
„Алапаевск“	От 12 до 15	Не пришло
	1963 год	
„Арагви“	4	5
„Вормси“	6	10
„Коломна“	7	30
„Грибоедов“	13	Не пришло
„Медногорск“	15	Не пришло
„Ижевск“	22	18
„Кировск“	24	Не пришло
„Репино“	Не ожидалось	30

Подобная же картина наблюдалась не только в этом порту и не только в этом месяце. Так, в другой порт за месяц не прибыло 17 судов, приход которых намечался графиком, и пришло 22 судна, не предусмотренных графиком. Из приведенной таблицы мы видим, что нередко суда приходят в намеченный месяц, но в другие сроки, не по графику. При этом общее число судов, приходящих в порт, оказывается сравнительно близким к намеченному планом. Очевидно, что график является не законом движения судов, а лишь примерным указанием на интенсивность возможного их движения в намеченный период. Как бы точно мы ни стремились спланировать движение судов и сделать поступление их в порт назначения равномерным, нам это не удастся, поскольку в действительности действует множество причин, которые невозможно предусмотреть заранее: метеорологические условия на трассе, изменение ранее намечавшихся маршрутов, условия поступления грузов, новые договоры на срочную поставку грузов, положение с рабочей силой в портах погрузки и выгрузки и пр.

Сказанное заставляет нас ожидать значительной неравномерности в прибытии судов в порт. Фактические наблюдения сведем в две таблицы.

Таблица 2

Данные о приходе судов в порт за 42 дня

Число судов, прибывших за сутки	0	1	2	3	4
Наблюденное число суток с приходом указанного числа судов	16	13	10	3	0

Промежутки времени между приходом последовательных судов подвержены очень большим случайным колебаниям. В уже указанной работе Б. В. Гнеденко и М. Н. Зубкова приведены данные о 335 промежутках времени, которые были наблюдены между 336 судами, прибывшими в порт за определенный промежуток времени. Эти данные сведены в табл. 3, из которой видно, что были периоды, когда в порт не прибывало ни одного судна в течение более чем трех суток. В то же время было большое число случаев, когда суда прибывали в порт буквально одно за другим. Так, промежутков, меньших часа, наблюдалось 34.

Таблица 3

Промежутки времени между приходами судов в порт в часах

Величина промежутка времени между приходом судов	Число случаев	Величина промежутка времени между приходом судов	Число случаев
0—1	34	15—16	7
1—2	28	16—17	9
2—3	23	17—18	13
3—4	18	18—20	13
4—5	8	20—22	10
5—6	13	22—24	8
6—7	14	24—26	6
7—8	13	26—28	9
8—9	8	28—30	9
9—10	16	30—35	9
10—11	5	35—40	8
11—12	14	40—50	9
12—13	9	50—60	3
13—14	7	60—70	2
14—15	9	70—80	2

Нужно сказать, что при планировании работы порта приходится считаться не только со случайным разбросом моментов прихода судов, но и со значительным разбросом длительности погрузо-разгрузочных операций. Анализ 389 случаев обработки судов в одном из портов за одну навигацию, в течение которой никакой реконструкции причалов по приемке генеральных грузов не производилось, показал, что длительность обработки судна колебалась от 1 до 518 часов. Данные о длительности грузовых операций в порту сведены в табл. 4.

Таблица 4

Время, затрачиваемое на обработку судна в порту, в часах

Длительность погрузочно-разгрузочных операций	Число случаев	Длительность погрузочно-разгрузочных операций	Число случаев
0—5	17	80—85	8
5—10	23	85—90	6
10—15	17	90—95	12
15—20	24	95—100	4
20—25	25	100—110	12
25—30	17	110—120	10
30—35	14	120—130	11
35—40	13	130—140	6
40—45	9	140—150	10
45—50	20	150—160	5
50—55	17	160—180	10
55—60	19	180—200	7
60—65	14	200—250	10
65—70	16	250—300	7
70—75	15	300—400	2
75—80	8	400—600	1

Приведенные числовые данные достаточно убедительно показывают, что решение экономических и технических проблем, связанных с эксплуатацией флота и портов, не могут быть решены элементарными арифметическими приемами. Действительно, если при подсчете числа причалов, которые рационально соорудить в порту для реализации заданного грузооборота, мы станем исходить из таких средних цифр: средний тоннаж судна, среднее время обработки судна у причала, заданный грузооборот, то мы наверняка придем к неудачному решению.

Пусть, для примера, навигационный период порта равен 250 суткам, годовой грузооборот — 1 млн. тонн, средняя грузоподъемность судна — 5000 тонн и средняя длительность стоянки для обработки судна равна 5 суткам.

Ясно, что для обработки плановых грузов порт должен принять 200 судов, которые вместе потребуют 1000 суток для обработки. Таким образом, если в порту соорудить 4 причала, то порт справится с планом грузооборота. Однако этот чисто арифметический подход неудачен, поскольку он не учитывает случайных колебаний в подходе судов. В результате такого подсчета неизбежно будут создаваться в порту очереди судов под погрузку и выгрузку. Хотя причальные средства будут использованы с максимальной загрузкой, народнохозяйственный эффект будет неудовлетворительным, так как суда потеряют огромное количество времени в ожидании освобождения причалов. Поскольку как простой судов, так и простой причалов приводит к некоторым потерям, возникает задача определения экономически оптимального числа причалов. Решение этой задачи неизбежно приводит к необходимости привлечения теории массового обслуживания.

Приведенный пример является типичным для многих реальных ситуаций: расчет рационального запаса инструмента, числа ремонтных рабочих, пунктов выдачи инструмента, определение числа коек в больнице, числа посадочных полос на аэродроме, пропускной способности моста, шлюза и пр.

Естественно, что в конкретных задачах возникают и некоторые своеобразные постановки вопросов, учитывающие специфику той области деятельности, в которой они появились. Например, при организации пункта скорой помощи в том или ином городе необходимо выбрать число машин и дежурных бригад, способное обеспечить такое обслуживание населения, при котором процент вызовов, обслуженных несвоевременно, будет достаточно малым, не превышающим заданной величины. Точно так же при расчете числа станков, которые необходимо установить для производства последовательных операций при обработке определенного типа изделий, следует обеспечить бесперебойную работу всей линии даже при условии остановки какого-либо из ее звеньев.

Решение всех задач подобного типа требует в первую очередь изучения входящего потока требований (в нашем примере — моментов прихода судов в порт), а также длительности обслуживания отдельного требования. В последнее время в связи с тем, что на обслуживающую аппаратуру возлагаются все более и более ответственные задачи, возникла настоятельная необходимость учитывать также влияние на эффективность обслуживающей системы поломок ее приборов, а также разного рода отказов и сбоев. При этом возникают своеобразные многочисленные новые задачи, которые еще далеки от всестороннего разрешения. Эти задачи сближают задачи теории массового обслуживания с задачами теории надежности, и очень часто постановки вопросов той и другой теории лишь словесно отличаются друг от друга. Для примера отметим, что вопросы резервирования с восстановлением представляют собой типичную задачу теории массового обслуживания. Именно с этих позиций несколько задач теории резервирования были решены Т. П. Марьяновичем [1040, 1042] и рядом других исследователей.

Всякое исследование по теории массового обслуживания начинается с изучения того, чтó необходимо обслуживать, или, как говорят, входящего потока требований. За последние годы в этом направлении были выполнены многие интересные исследования. А. Я. Хинчин сосредоточил свое внимание на потоках без последействия и нашел широкие достаточные условия, при которых поток, являющийся суммой большого числа независимых между собой случайных потоков, каждый из которых «мал» по сравнению с суммой остальных, был бы близок к простейшему. Позднее Г. А. Ососков — ученик А. Я. Хинчина — показал, что по сути дела эти условия являются и необходимыми. Все исследования А. Я. Хинчина по теории массового обслуживания, в том числе и по теории входящего потока, недавно изданы на русском языке отдельной книгой [1098].

В последнее время обобщением его результатов занимались многие исследователи, среди которых я хотел бы указать Б. И. Григелиониса и Франкена. В работах Б. И. Григелиониса (см., например, [995]) были не только получены новым путем результаты Г. А. Ососкова, но и рассмотрены вместо стационарных слагаемых (точнее, слагаемых со стационарными приращениями) произвольные слагаемые случайные процессы и выяснены условия сходимости суммарных потоков к общим потокам Пуассона. Кроме того, он и Франкен изучили асимптотические разложения по степеням малого параметра распределений суммарного потока.

Всестороннее изучение условий, при которых поток приближается к пуассоновскому, было предметом работ Реньи [713], Р. Л. Добрушина [998], Ю. К. Беляева [929], Бреймана и др. Оказалось, что пуассоновский поток появляется не только при суммировании большого числа независимых стационарных равноправных в некотором смысле потоков, но и при других операциях. Так, Реньи рассматривал своеобразный процесс разряжения потока, которому соответствует сохранение не всех первоначально поступивших требований и одновременное сжимание времени. С такого рода разряжением первоначального потока приходится иметь дело во многих реальных практических задачах: при многократной проверке корректур, когда каждая проверка уничтожает значительную долю имевшихся опечаток набора; при изучении потока деталей, проходящих на станках последовательную обработку с одновременным отсеиванием испорченных на каждой операции изделий, и т. д.

Р. Л. Добрушин, а также Брейман показали, что при случайном блуждании частиц в пространстве при весьма общих условиях также получается пуассоновское распределение.

Большое число работ посвящается изучению реальных потоков, с которыми приходится встречаться на практике: приход судов в порты, прибытие самолетов, распределение деталей на конвейере, вызовы скорой помощи и пр.

Не меньшее значение, чем изучение входящего потока, представляет собой исследование выходящих потоков, т. е. потоков событий после системы обслуживания. Если нас интересует работа телефонной станции, то такими выходящими потоками могут быть моменты окончаний разговоров, потоки вызовов, получивших отказы, и т. д. Теория выходящего потока до сих пор разработана совсем недостаточно. Помимо работ Смита и некоторых других, здесь следует указать на работы Н. В. Яровицкого [1110—1114].

Входящий поток поступает в систему обслуживания. Самому понятию «система обслуживания» приходится придавать весьма широкий смысл, включающий число приборов, характер их использования поступающими требованиями (требование пользуется только одним из приборов, требование проходит последовательную обработку на ряде приборов, требование выбирает первый попавшийся свободный прибор, требование может выбирать лишь некоторые из имеющихся приборов и т. д.), поведение требований в системе обслуживания (при поступлении требования оно остается ожидать, пока один из приборов не освободится от обслуживания ранее поступивших требований; требование, которое застало все приборы занятыми обслуживанием ранее поступивших требований, уходит из системы; требование от момента поступления в систему до момента ухода из нее может потерять не больше чем заданное — постоянное или случайное — время и т. д.). Нужно сказать, что в книге Саати эта сторона теории выяснена достаточно хорошо.

В последние годы начали изучать влияние на качество обслуживания ненадежности приборов, выхода их из рабочего состояния и последующего восстановления. Эти вопросы совсем не затронуты в предлагаемой читателю книге. Это направление исследований очень существенно, и, к сожалению, ему уделяется до сих пор совсем недостаточное внимание. Все реальные системы обслуживания не только не обладают абсолютной надежностью, но они вдобавок стареют и не полностью восстанавливают свои свойства. Естественно возникают вопросы экономического характера: до каких пор целесообразно использовать установленную аппаратуру системы массового обслуживания, как сказывается на эффективности системы обслуживания ненадежность аппаратуры? Эти вопросы, как мне известно, до сих пор совсем не изучались.

В последние годы появился ряд интересных исследований, в которых решены разного рода экстремальные задачи: см., например, книгу Ховарда «Динамическое программирование и марковские процессы» («Советское радио», 1964), статьи И. Н. Коваленко, Г. П. Климова. Задачи этого типа также не рассмотрены в книге Саати.

Несомненно, что до последнего времени совсем не уделялось внимания разного типа статистическим задачам в теории массового обслуживания. Здесь огромное поле интересной и очень серьезной деятельности. Для примера я хотел бы указать на одну простую и притом естественную задачу: счетчик Гейгера—Мюллера в течение времени T регистрировал космические частицы. Спрашивается, каков истинный поток частиц, сколько частиц не было зарегистрировано счетчиком?

С позиций математика теория массового обслуживания представляет собой своеобразную задачу теории случайных процессов. Действительно, случайный поток, являющийся не чем иным, как целочисленным монотонным случайным процессом, подвергается некоторой трансформации (воздействию системы обслуживания). Требуется найти некоторые числовые характеристики результата воздействия этой трансформации. Такими характеристиками в зависимости от обстановки могут быть среднее время ожидания начала обслуживания, среднее число потерянных требований, средняя длительность периода непрерывной занятости обслуживающего прибора и т. д. Это замечание поясняет значение теории случайных процессов для теории массового обслуживания.

Если в начальный период развития теории массового обслуживания исследователи предполагали входящий поток простейшим и длительность обслуживания распределенной по показательному закону, то теперь стремятся найти такие методы, которые позволили бы отказаться от этих сильно ограничивающих предположений. На этом пути удалось добиться некоторых успехов. Особенно далеко удалось продвинуться при этом в случае одного прибора (как часто говорят, в случае однолинейных систем). Здесь работы Линдли, Бенеша, И. Н. Коваленко, Такача и ряда других существенно обогатили теорию массового обслуживания.

В случае многих приборов успехи пока достигнуты преимущественно при дополнительном ограничении, состоящем в том, что либо поток является простейшим, либо длительность обслуживания подчинена показательному закону распределения. В этих условиях серьезные результаты получены Коксом, Б. А. Севастьяновым, Такачом, Кифером и Вольфовицем. Из чисто математических методов, которые широко используются при решении конкретных задач, следует отметить метод добавочной переменной Кокса, метод интегро-дифференциальных уравнений Линдли—Такача—Севастьянова, метод вложенных цепей Кендалла, намеченный А. Я. Хинчиным еще в 1932 г. [1090].

Аналитическое решение обладает рядом положительных качеств: оно не привязано к определенным числовым значениям параметров потока и системы обслуживания, позволяет находить оптимальные решения и делать общие заключения. Однако во многих случаях аналитическое решение получить затруднительно, поскольку задача настолько сложна, что составление уравнений, к которым сводится задача, представляет собой практически неразрешимую задачу. В таких случаях неоценимую помощь способен оказать метод Монте-Карло, получивший наименование метода статистических испытаний. Этот метод был развит применительно к задачам теории массового обслуживания Н. П. Бусленко. С основами этого метода можно ознакомиться по хорошей монографии группы авторов [948].

Если 5—6 лет назад мировая литература по теории массового обслуживания была очень бедна и кроме монографии А. Я. Хинчина [1093], пожалуй, было трудно указать какую-нибудь другую, то теперь появилось большое число интересных книг очень различной направленности. Для того чтобы помочь читателю лучше ориентироваться в существующей литературе, я позволю себе привести очень короткий список иностранных монографий, а также список известных мне таблиц, которые необходимы при решении реальных задач.

ЛИТЕРАТУРА

1. Morse P. M. Queues, inventories and maintenance, N. Y., J. Wiley, 1958, p. 202.
2. Bodino G. A., Brambilla F. Teoria delle code, Milano, 1959, p. 219.
3. Syski R. Introduction to congestion theory in telephone systems, Oliver and Boyd, Edinburgh and London, 1960, p. 742.
4. Cox D. R. and Smith W. Queues, Methuen & Co., London, 1961, p. 180.
5. Pollaczek F. Théorie analytique des problemes stochastiques relatifs a un group de lignes téléphoniques avec dispositif d'attente, Paris Gauthier-villars, 1961, p. 114.
6. Le Gall P. Les systémes avec on sans attente et les processús stochastiques, t. I, Paris, Dunod, 1962, p. 482.
7. Takàcs L. Introduction to the theory of queues, Oxford University Press, 1962, p. 268.
8. Riordan J. Stochastic service systems. Wiley and Sons, 1962, p. 139.
9. Beneš V. E. General stochastic processes in the theory of queues, Addison-Wesley, Massachusets, 1963, p. 88.
10. Haight F. A. Mathematical theories of traffic flow. Academic Press, New York and London, 1963, p. 242.
11. Jensen A. Moe's principle. An econometric investigation intended an aid in dimensioning and menaging telephone plant. Theory and tables, Copenhagen, 1950.
12. Tables of the individual and cumulative terms of Poisson distributions; Defense Systems Department General Electric Company; van Nostrand Company, 1962, p. 202.
13. Peck L. G., Hazelwood R. N. Finite queueing tables, John Wiley, 1958, p. 210.
14. Descloux A. Delay tables for finite and infinite source systems, McGraw—Hill Co., 1962.
15. Palm C. Table of the Erlang loss formula, Stockholm, 1954.
16. Башарин Г. П. Таблицы вероятностей и средних квадратических отклонений потерь на полнодоступном пучке линий. Изд-во АН СССР, 1962, стр. 128.

Академик АН УССР Б. В. Гнеденко

МОЕЙ МАТЕРИ

Предисловие автора

Очередь связана с ожиданием. Обычно можно наблюдать, что очереди образуются у билетных касс, в кафетериях, на автобусных остановках. Следовательно, чтобы образовалась очередь, требуется, чтобы клиенты прибывали к такому обслуживающему устройству, у которого им нередко приходится ожидать, например у кабинета врача, чтобы попасть на прием в порядке записи. Телеграммы классифицируются по срочности доставки, при этом обычным телеграммам дается преимущество даже перед срочными письмами.

Теория массового обслуживания — отрасль прикладной математики, использующая методы теории вероятностных процессов. Стимулом к развитию теории массового обслуживания послужили попытки предсказывать случайно изменяющиеся потребности по результатам наблюдений и на основании этого организовывать обслуживание, характеризующееся приемлемым временем ожидания. Теория массового обслуживания позволяет раскрывать природу очередей, что обеспечивает возможность лучшего управления процессом. Возьмем такой пример. Если в гастрономическом магазине мало кассиров, покупатели будут охотнее ходить в другие магазины, что приведет к потерям. Следовательно, нужно обоснованно выбрать число кассиров; это снизит потери. Пожалуй, в этом деле можно обойтись без теории массового обслуживания. Но без нее уже нельзя обойтись, например, при организации управления движением самолетов в аэропорту с максимальной слаженностью, вследствие большого числа действующих факторов и сложности системы. В этом случае теория может помочь прогнозировать длительность ожидания, число пассажиров, ожидающих в любой момент времени, длительность периода занятости и т. д. Такого рода прогнозы и рекомендации помогают владельцу предприятия предвидеть ситуацию, принимать соответствующие меры и ослаблять перегрузку. Кроме того, эта теория побуждает как владельца предприятия, так и клиента осознавать постоянную необходимость в новых идеях для упрощения запутанности современной жизни.

Среди проблем, возникающих вследствие перенаселения (это понятие Мальтуса сейчас называют взрывообразным ростом населения), наиболее важными являются проблема питания и проблема жилища, большой спрос на предприятия массового обслуживания, обеспечение способности людей жить в условиях увеличивающейся скученности. Последний момент относится к эстетике, первый — к экономике. Решение многих практических задач требует применения теории массового обслуживания. Какие организационные мероприятия следует провести при ситуациях скученности, чтобы свести к минимуму общие потери времени, непроизводительные операции, и, в более общем смысле, непроизводительные затраты усилий и средств?

Теория массового обслуживания непосредственно не связана с оптимизацией. Она скорее пытается разработать, изучить и сравнить различные ситуации массового обслуживания и, таким образом, косвенно добиться приближенной оптимизации. Многим из тех, кто занят в сфере общественного обслуживания, следовало бы прилагать больше усилий для оценки тонкой природы и влияния очередей. Это видно, например, из следующего письма редактору газеты «Вашингтон Пост»: «Должно быть, высокопоставленные лица запланировали закрыть накануне Пасхи, когда наблюдается самое напряженное движение, один из проездов моста Мемориал [мост с шестью проездами]; из-за этого поток по ширине уменьшится на одну треть. Если отодвинуть одну из трех автомашин на длину ее корпуса и помножить на несколько тысяч, то последний автомобиль будет где-нибудь в районе Фоллз Чёрч [в семи милях от этого места].

Мне кажется, что тот, кому поручено нанесение белых полос, мог бы нанести их на мосту раньше, а субботу использовать для работы на какой-либо боковой улице, которую, я уверен, он нашел бы без труда» [1].

Методы теории массового обслуживания в большей степени основаны на анализе, чем на синтезе. Методы синтеза используются для обобщений и определения направлений аналитических исследований новых проблем.

Главная цель книги — дать общее изложение и обобщение содержания отдельных статей и книг по теории массового обслуживания. Кроме того, здесь содержится обширная библиография и указываются некоторые нерешенные проблемы. Книга включает также описательную вводную главу, большая часть которой предназначена для неспециалистов. Многие положения иллюстрируются примерами. Некоторые детали, опущенные в основном тексте, перенесены в многочисленные упражнения. Последние также имеют целью дать возможность читателю самостоятельно выводить формулы по намеченной схеме.

---

[1] Письмо редактору «Вашингтон Пост», 21 апреля 1960 г., от Джозефа М. Рей, Александрия, штат Виргиния.

Хотя книга предназначена для специалистов с высшим образованием, имеющих хорошую подготовку по математическому анализу, теории вероятностей, основам теории функций комплексного переменного, теории матриц, она может быть также использована исследователями, которым нужны готовые результаты теории массового обслуживания, ее идеи и методы.

По своему содержанию книга делится на четыре части. В первой части, которая включает три главы, даются основные понятия и методы теории массового обслуживания и теории вероятностей. Вторая часть содержит две главы. Она посвящена марковским моделям, аналитическое рассмотрение которых впервые было проведено Эрлангом. Третья часть, состоящая из пяти глав, касается немарковских моделей, которые исследуются, главным образом, в стационарном случае. Специально длительности обслуживания отведена гл. 9. Четвертая часть также содержит пять глав и посвящена влиянию различных дисциплин очереди на характеристики системы, сложным системам массового обслуживания и многочисленным частным задачам. В эту же часть входят глава о приложениях теории и глава о теории восстановления. Библиографии предшествует краткий обзор некоторых проблем, которые нуждаются в формулировке.

Чувствовалось, что разнообразный материал гл. 2 будет лучше воспринят читателем, если его поместить перед более формализованной гл. 3. Кроме того, ознакомление любознательного читателя с многообразием идей в начале книги укрепит его желание продолжать чтение.

В тексте встречаются понятия «требование», «вызов», «клиент», которые, в сущности, являются равнозначными. Совокупности, из которых требования поступают в систему обслуживания, называются источниками входящего потока.

В большинстве случаев принятые обозначения будут уточняться при употреблении. Многие обозначения применяются довольно часто, так что читатель легко с ними освоится.

Выражаю благодарность д-ру Джорджу Моргенталлеру за критический просмотр первого варианта рукописи, д-ру Ришарду Сиски и д-ру Эрику Уолмену за их полезные советы и читку корректуры, д-ру Лейле Брем и м-ру Александру Кроу за помощь и замечания при читке корректуры.

Томас Л. Саати

Основные обозначения

X, Y, Z — произвольные случайные величины.
S_n — n-я частичная сумма $X_1 + \ldots + X_n$.
M — символ, обозначающий, что поток — пуассоновский либо длительность обслуживания распределена по экспоненциальному закону.
G — произвольное распределение.
GI — обозначение последовательности независимых одинаково распределенных (по произвольному закону) случайных величин.
D — вырожденное распределение.
$\varphi(t)$ — характеристическая функция случайной величины.
$M(t)$ — производящая функция случайной величины.
$E[\ ]$ — математическое ожидание.
C_t — коэффициент вариации случайной величины t (отношение дисперсии к квадрату математического ожидания).
σ — среднее квадратическое отклонение.
σ_ρ — среднее квадратическое отклонение ρ.
μ'_r — r-й момент относительно произвольной точки.
μ_r — r-й начальный момент.
k_r — r-й семиинвариант.
$\equiv$ — знак определения.
$*$ — обозначение преобразования Лапласа.
$\sim$ — обозначение преобразования Лапласа—Стилтьеса.
τ — в большинстве случаев — временной интервал.
δ_{in} — символ Кронекера.
X_t — случайный процесс.
$F(x, t)$ — функция распределения случайной величины X в момент времени t.
E_j — j-е состояние системы.
P — матрица вероятностей перехода.
$P_{ij}^{(n)}$ — вероятность перехода из состояния E_i в состояние E_j в n-й момент перехода.
$f_{ij}^{(n)}$ — вероятность того, что система, находясь в момент времени $t=0$ в состоянии E_i, в момент времени n в первый раз окажется в состоянии E_j.
μ_{ij} — среднее время первого перехода из состояния E_i в состояние E_j.
$p(j, t; i, s)$ — вероятность того, что в момент времени t система находится в состоянии E_j, если в момент времени s система находилась в состоянии E_i.

$A(t)$ — функция распределения длительности промежутков времени между моментами поступления требований.

$a(t)$ — соответствующая плотность вероятности.

$B(t)$ — функция распределения длительности обслуживания.

$b(t)$ — соответствующая плотность вероятности.

$\beta(t)$ — преобразование Лапласа—Стилтьеса функции $B(t)$.

λ_n, μ_n — параметры процесса рождения и гибели.

q — число требований в очереди (случайная величина).

$P_{in}(t)$ — условная вероятность того, что в момент времени t в системе (в очереди и на обслуживании) находится n требований, если в момент времени $t=0$ в системе было i требований.

$P_n(t)$ — то же, что и выше (при опущенном i), а также вероятность того, что в момент времени t занято ровно n каналов.

p_n — вероятность того, что в стационарном состоянии в системе находится ровно n требований; употребляется также для обозначения вероятностей частных значений дискретной случайной величины.

π_n — вероятность того, что в установившемся режиме в момент времени, предшествующий началу обслуживания, или в момент времени, следующий непосредственно после окончания обслуживания, одноканальная система находится в состоянии E_n.

$P(z)$ — производящая функция последовательности, обычно P_n, но иногда и π_n.

$P(z, t)$ — производящая функция последовательности $\{P_n(t)\}$.

c — число каналов обслуживания.

λ — $\left(\text{иногда } \frac{1}{a}\right)$ интенсивность поступления требований.

μ — $\left(\text{иногда } \frac{1}{b}\right)$ интенсивность обслуживания.

s — число требований в группе (главным образом, в гл. 7); параметр преобразования Лапласа в других главах.

w_n — время ожидания n-го требования в очереди (случайная величина).

s_n — время обслуживания n-го требования (случайная величина).

t_n — n-й промежуток времени между моментами поступления требований (случайная величина).

$$u_n = s_n - t_n.$$

$U(u_n)$ — функция распределения u_n.

ρ — загрузка, или коэффициент использования (обычно отношение интенсивности поступления требований к интенсивности обслуживания).

L — среднее число требований, находящихся в системе.

L_q — среднее число требований в очереди.

$W(t)$ — длительность ожидания в очереди в момент времени t (это фиксированная реализация, а не распределение вероятностей).

$w_n(t)$ — плотность распределения длительности ожидания $(n+1)$-го требования.

$W_n(t)$ — соответствующая функция распределения.

$P(w, t)$ — функция распределения времени ожидания w в очереди в момент времени t; $P(w)$ — предел этой функции при $t \to \infty$ (если он существует).

$P(<t)$, $P(w)$ — функция распределения времени ожидания в стационарном режиме. [$P(<t)$ — неудачное, ранее сложившееся обозначение. Так, $P(<t)$ — то же самое, что и $P(w)$. Заметим это, так как при вычислении выражений вида $P(t=0)$ и $P(t>0)$ оставлены старые обозначения].

$P(t=0)$ — вероятность обслуживания без ожидания.

$P(t>0)$ — вероятность положительной длительности ожидания.

$w(t)$ — плотность времени ожидания, т. е.

$$w(t) = \frac{dP(<t)}{dt}.$$

$\gamma(s)$ — преобразование Лапласа—Стилтьеса $P(<t)$.

$\gamma(s, t)$ — преобразование Лапласа—Стилтьеса $P(w, t)$.

W — среднее время пребывания в системе.

W_q — среднее время ожидания в очереди.

W_p — среднее время ожидания требования с p-м приоритетом.

$G(x)$ — функция распределения длительности периода занятости для системы $M/G/1$ в гл. 8.

$\Gamma(k)$ — гамма-функция.

$\Gamma(s)$ — преобразование Лапласа—Стилтьеса функции $G(x)$.

$p_m(t)\,dt$ — вероятность того, что период занятости для системы $GI/M/1$ находится в пределах $(t, t + dt)$.

$f_n(t)$ — вероятность того, что данное из числа n требований, ожидающих в любой момент времени, будет впоследствии ожидать в течение времени t, если выбор на обслуживание производится случайным образом.

$P_{nm}(t)$ — вероятность того, что в момент времени t в системе с прерыванием обслуживания находится n требований с первым приоритетом и m требований со вторым приоритетом.

$P(n_1, \ldots, n_k;\ t)$ — вероятность того, что в момент времени t в системе с k последовательными каналами в j-м канале ожидает n_j $(j = 1, \ldots, k)$ требований.

$p(n_1, \ldots, n_k;\ t)$ — та же вероятность для стационарного случая.

Часть I

СОДЕРЖАНИЕ ПРЕДМЕТА, МЕТОДЫ И ОСНОВЫ ТЕОРИИ

В первой главе части I этой книги дается качественное описание структуры систем массового обслуживания — как существующих, так и перспективных; приведены многочисленные примеры. Во второй главе дается представление о различных моделях, о применении статистических методов, приводятся количественные критерии принятия гипотезы и излагается методика анализа систем массового обслуживания с помощью метода Монте-Карло. В третьей главе даются основы теории вероятностей и теории случайных процессов, в частности, рассматриваются цепи и процессы Маркова.

Глава I

ОПИСАНИЕ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

1.1. Введение

В этой книге дается изложение существующей теории массового обслуживания, ограниченное рамками предмета. Особое внимание уделяется различным математическим моделям и применяемым методам решений с приложением большого числа понятий к некоторым областям практической деятельности, таким, как телефония, управление запасами, ремонт машин, режим работы плотины, управление движением воздушного транспорта.

В этой главе в доступной описательной форме излагается содержание теории массового обслуживания. Гл. 2 — вводная техническая глава с многочисленными понятиями для читателя, у которого появится интерес к предмету. В последующих трех главах даются понятия из теории вероятностей и математического анализа, связанные с решением задач массового обслуживания. Гл. 4 и 5 посвящены марковским моделям массового обслуживания. Остальные вопросы теории рассматриваются затем в гл. 6—13; следующая, гл. 14, посвящена приложениям. Гл. 15 касается теории восстановления. Приводится обширная библиография.

Математических методов, применяемых в теории массового обслуживания, сравнительно немного; поэтому естественно, что материал гл. 1 более насыщен понятиями, чем излагаемый в остальной части книги. Однако прогресс в исследовании систем массового обслуживания позволяет находить широкие обобщения, намного выходящие за пределы первоначально поставленных задач.

Для образования очереди необходим поток требований, которые ожидают поступления в устройство, производящее обслуживание. Предположим, что таким устройством является касса в бакалейно-гастрономическом магазине. Если очередь длинна, то нетерпеливые покупатели могут уходить из магазина, что приводит к убыткам владельца. Владелец магазина может решить, что выгодно вложить средства в другую кассу, так как затраты на нее окупятся прибылью от нетерпеливых покупателей, большинство из которых теперь уже останется и будет обслужено. Таким образом, в теорию массового обслуживания вводится оптимизация. Вообще, в отличие от теории оптимизации, в которой главное значение имеет максимум или минимум функции цели при наложенных связях, теория массового обслуживания имеет своей основной чертой математическую формализацию процесса. Она выражает в виде формул, объясняет и предсказывает ситуации массового обслуживания, обеспечивая лучшее понимание их и принятие соответствующих мер.

Может быть, вы задумывались над тем, как часто и как долго в своей повседневной деятельности вам приходится задерживаться в очередях. Вы ожидаете транспортные средства, еду, доктора; в самолете вы ожидаете посадки; вы ожидаете в театре и т. д. Можно произвести приближенный подсчет, сколько в среднем за неделю вы простаиваете во всех очередях, с которыми сталкиваетесь. Распространив в итоге этот срок на всю жизнь, вы убедитесь, что продолжительность времени, теряемого в очередях, будет ужасающей. Если у кого-либо из вас и есть свободное время, то от простаивания во многих очередях в значительной степени омрачается радость жизни. Но при предвидении и хорошем планировании неприятности и потери из-за очередей могут быть сведены к минимуму.

То, что очереди неизбежны и будут оставаться, является естественным следствием прироста населения. Чем больше людей требует определенного вида обслуживания (в театре или же на городском транспорте), тем длиннее очередь. Эти потребности возрастают с ростом населения; следовательно, общая длительность времени, теряемого в очередях в течение всей жизни, есть функция, возрастающая с течением времени[1]. Конечно, можно пропорционально расширять и обслуживание, но нередко этому препятствует величина затрат, как, например, при расширении городских улиц для увеличения движения. Правильное планирование, учитывающее прирост населения, может предупреждать некоторые из этих трудностей.

Большинство систем обслуживания связано с людьми. Исследование причин очередей и принятие определенных мер нельзя полностью отделить от рассмотрения влияния человека на эту проблему. Нельзя принимать меры без учета того обстоятельства, что существуют люди, использующие средства обслуживания и оказывающие влияние на результат конечного анализа.

Запланированные очереди — хорошо знакомая черта городской жизни, они обязательно присутствуют и при установившемся порядке массового производства на различных его фазах. Наличие очереди может быть хорошей или плохой рекомендацией для общественных средств обслуживания. Например, ясно, что постоянные

[1] Этот вывод, на наш взгляд, ничем не обоснован. — *Прим. ред.*

очереди у театра или кино являются признаком того, что спектакль или фильм стоит посмотреть и что это — гвоздь программы, на который не так-то легко попасть. Таким образом, чтобы найти хорошее представление, ищите подлиннее очередь [1]. С другой стороны, одна из самых постоянных и самых раздражающих очередей — очередь на автобус. Иногда она лишь незначительно отличается от того хаоса, во избежание которого устанавливается. Многие остановки обслуживают несколько маршрутов. В результате, когда одновременно прибывает два или более автобусов, начинается свалка и общее разрушение очереди. Как только автобус наполняется, выброшенный из очереди пассажир с большим трудом находит свое законное место, и нередко вся очередь должна устанавливаться заново.

Было бы весьма желательно ожидать врача, основываясь на расписании приема, которое должно быть согласовано как для врача, так и для пациента. Ведь нередко потеря времени для пациента связана с бо́льшим общественным и экономическим ущербом, чем потеря времени для врача. Если расписание приема будет тщательно составлено на основании определенного исследования, то это может только укрепить престиж врача. По крайней мере, это будет свидетельствовать о том высоком уважении, с которым врач относится ко времени пациента.

Как те, кто образует очередь, так и те, к кому она образована, должны знать и понимать природу массового обслуживания и предвидеть его хитрости. Нередко ожидающий мог бы выиграть время, зная, что очередь, в которой он ожидает, запланирована оптимально с тем, чтобы уменьшить время ожидания или лично его или группы людей. Тот, к кому образуется очередь, должен знать, как лучше ее спланировать, связав очередь со своей деятельностью. Наиболее важной из всех является очередь жизни, т. е. цикл рождение—жизнь—смерть. Жизнь рассматривается с точки зрения ожидания. Следовательно, в известном смысле, понимание очереди становится пониманием планирования лучшей жизни.

Некоторые основные виды массового обслуживания, к которым применимы понятия этой книги, относятся к следующим областям: средства связи (телефон, телеграф, почта), транспортные средства (воздушные, наземные, морские), средства и предприятия обслуживания (театры, рестораны, автобусы, больницы, клиники), управление запасами и производственные процессы (эксплуатация механизмов, сборочные линии, обслуживание станков), физические процессы (управление системой плотин, движение частиц через отверстие), эпидемические процессы в биологии, прирост населения, даже обсуждение статей для публикации, физиологический поток нервных импульсов и т. д.

---

[1] Довольно ограниченный критерий. — *Прим. ред*

1.2. Поясняющий пример и числовая иллюстрация

Многие понятия, употребляемые в теории массового обслуживания, можно проиллюстрировать на одном важном примере: взлет и посадка самолетов в столичном аэропорту — операция, представляющая интерес для многих людей, пользующихся этим видом транспорта. Термины теории массового обслуживания, поясняемые на этом примере, выделяются курсивом.

Допустим, что аэропорт имеет несколько взлетно-посадочных полос (*параллельных каналов*). Эти полосы ведут к большему или меньшему количеству дорожек, оканчивающихся на аэровокзале (*последовательные каналы*). После того как самолет, прибывший в соответствии с определенным *распределением входящего потока*, приземляется, он присоединяется к очереди самолетов, ожидающих обслуживания (продвижения по дорожке к месту выгрузки). Таким образом, *выходящий поток* одной очереди становится *входящим потоком* для другой. Очередь существует как на земле (взлет самолета), так и в воздухе (посадка самолета). Обе эти очереди имеют свое распределение входящего потока. Приземляющиеся самолеты могут прибывать *группами*, при этом члены каждой группы должны кружить над аэропортом и приземляться по порядку. (Если полоса очень широкая, то нетрудно представить посадку самолетов группами.) *Длительность операции обслуживания* (время приземления или взлета) около минуты. В любом случае имеется *распределение времени обслуживания*. Если для различных типов самолетов отведены различные взлетно-посадочные полосы, которые могут быть длиннее, например, для реактивных самолетов, то распределение времени обслуживания может изменяться от одной полосы к другой.

При выборе самолетов для посадки важно определить соответствующий *показатель эффективности*. Например, если желательно *минимизировать общее время ожидания* пассажиров, то вначале нужно производить посадку самолетов с бо́льшим количеством людей.

Здесь же имеет место элемент обслуживания *с приоритетом*, когда разрешается посадка снижающемуся самолету раньше, чем взлет ожидающему. Эта система с приоритетом распространяется на случай аварийной обстановки, когда вследствие крайней необходимости разрешается посадить первым самолет, прибывший позже. Нередко приоритет на посадку дается реактивным самолетам из-за ограниченного запаса топлива.

Иногда вследствие определенного характера расположения самолетов в зоне ожидания прибывающий самолет присоединяется к очереди эшелонированных самолетов, ожидающих посадки, а затем выбор самолета на посадку производится случайным образом (одна из форм обслуживания с приоритетом). Так, например, если самолет находится ближе других к точке, в которой он может покинуть зону ожидания, то ему будет дана команда на посадку. В промежутке времени между получением приоритета на посадку и командой «посадку разрешаю» самолет выходит из эшелона и направляется к аэродрому. Это время известно как время захода на посадку. Время приземления затрачивается на операцию приземления и продолжается до того момента, когда самолет свернет со взлетно-посадочной полосы.

Самолет, ожидающий посадки, может находиться в положении, близком к критическому (в это время другие самолеты будут действительно в критическом положении). Он примет решение *присоединиться к более короткой очереди* в ближайшем аэропорту и приземлиться там. Прибывающий самолет может не выстраиваться в эшелон, а уходить в другой аэропорт. В этом случае говорят, что аэропорт «потерял» этот самолет. Случается, что самолет отправляется в соседний аэропорт после того, как, присоединившись к очереди, он прождал больше, чем предполагалось. Можно рассматривать приземляющийся самолет участвующим в *цикле*, если он присоединяется к очереди самолетов, ожидающих взлета, и снова включается во входящий поток системы. Если приземляющийся самолет имеет информацию о размерах очереди эшелонированных самолетов, ожидающих посадки в соседнем аэропорту, то он может присоединиться к этой очереди. Если у него есть информация еще об одном аэропорте, то он может отправиться и туда (редкий случай). Это движение туда и обратно при наличии нескольких очередей называется переходом из одной очереди в другую (*возможность выбора очереди*).

Аэропорт может временно закрываться, и прибывший самолет будет вынужден отправиться в другой аэропорт, если количество эшелонированных самолетов, ожидающих посадки, достигнет заданной величины. Операция обслуживания может быть ускорена путем оборудования специальных гасителей скорости, которые позволяют самолетам приземляться на главной полосе с большой скоростью.

Основной проблемой при управлении аэропортом является связь. Если входящий поток как на земле, так и в воздухе велик, то аэропорт должен быстро связываться с самолетами и получать ответ. При организации связи важной проблемой является определение числа операторов и *каналов связи*, необходимых для регулирования различных состояний перегруженности, которые могут возникнуть. В данном случае необходим выбор *оптимального числа каналов* для обслуживания требований, поступающих в соответствии с данным распределением. Можно произвести сравнение *стоимости дополнительного канала* со стоимостью возросшего объема обслуживания существующими каналами.

Важной проблемой является наличие соответствующего места для ожидания в очереди. Например, при проектировании аэропорта существенным моментом является наличие наземной рулежной дорожки для самолетов, готовых ко взлету.

Во многих задачах теории массового обслуживания для определения необходимого показателя эффективности достаточно знать распределение входящего потока, дисциплину очереди (например, случайный выбор, обслуживание в порядке поступления или с приоритетом) и распределение времени обслуживания. В других задачах нужно иметь дополнительную информацию. Например, в случае отказов в обслуживании нужно определить вероятность того, что поступившее требование получит отказ сразу или после прибытия и, следовательно, покинет очередь до или после присоединения к ней.

С теоретической точки зрения очереди могут рассматриваться при условии, что поток проходит через систему пунктов обслуживания, соединенных последовательно или параллельно. На поток оказывают влияние различные факторы; они могут замедлять его, приводить к насыщению и т. д.

Рассмотрим следующую простую ситуацию для системы с одним обслуживающим устройством, приведенную в табл. 1.1. В первой строке дано время прибытия клиента A, а во второй — время, истекшее от момента прибытия предыдущего клиента до момента прибытия клиента A. В третьей строке дано время обслуживания клиента A, а в четвертой — время ожидания клиента A, равное сумме времени обслуживания и времени ожидания предыдущего клиента за вычетом промежутка времени между моментом прибытия предыдущего клиента и моментом прибытия клиента, чье время ожидания вычисляется. Так, например, клиент A будет ожидать в течение промежутка, равного времени ожидания и времени обслуживания предыдущего клиента, минус промежуток времени между моментом прибытия предыдущего клиента и моментом прибытия клиента A. Если результат равен нулю или отрицателен, то время ожидания равно нулю.

Таблица 1.1

Выборка данных обслуживания

Текущее время . . .	0	2	6	11	12	19	22	26	36	38	45	47	49	52	61
Промежуток времени между требованиями .	0	2	4	5	1	7	3	4	10	2	7	2	2	3	9
Время обслуживания	5	7	1	9	2	4	4	3	1	2	5	4	1	2	1
Время ожидания . .	0	3	6	2	10	5	6	6	0	0	0	3	5	3	0

Для обоснованного изучения реального процесса потребовалась бы значительно бо́льшая выборка, но некоторые важные характеристики массового обслуживания могут быть вычислены и в этом случае. В данном примере десять клиентов из пятнадцати ожидают. Среднее время ожидания для этих клиентов равно $\frac{49}{10}$; среднее время ожидания по всем клиентам равно $\frac{49}{15}$.

Можно вычислить также общее время простоя канала. Канал простаивает в ожидании клиента A, если промежуток времени между моментом прибытия клиента A и моментом прибытия предыдущего клиента превышает время ожидания в очереди и время обслуживания предыдущего клиента. Таким образом, общее время простоя равно сумме положительных разностей между промежутком времени от момента прибытия предыдущего клиента до момента прибытия данного клиента и временем ожидания плюс время обслуживания предыдущего клиента. Доля времени, в течение которого канал простаивает, равна отношению последней величины к общему времени обслуживания.

Если бы выборка была достаточно велика, то, подсчитав число случаев, когда ожидает один клиент, группа из двух клиентов и т. д., можно было бы получить относительную частоту ожидания для одного клиента, для двух клиентов и т. д., т. е. получить вероятность появления таких групп. Читателю предлагается ответить на вопрос: какую пользу можно извлечь из этих вероятностей?

Другой важной величиной является вероятность того, что в любой момент времени ожидает данное число клиентов. Заметим, к примеру, что четвертый клиент ожидает вместе с третьим в течение единицы времени, а затем он остается и ожидает в течение единицы времени вместе с пятым клиентом. Здесь образовались две ожидающие группы по два клиента в каждой.

Читателю предлагается построить диаграмму, состоящую из горизонтальных параллельных линий. Каждая линия связана с определенным клиентом, а длина их соответствует длительности ожидания каждым клиентом. Сверху проводится базисная линия, на которой откладывается текущее время. Каждая линия должна начинаться в момент прибытия клиента и оканчиваться в момент начала обслуживания. Таким образом, можно определить количество ожидающих клиентов и общую длительность ожидания. Например, линия, соответствующая второму клиенту, протянется от второй до четвертой единицы времени, а линия, соответствующая третьему клиенту, — от шестой до двенадцатой единицы. Эти две линии не перекрываются. Однако линия, соответствующая четвертому клиенту, протянется от одиннадцатой до четырнадцатой единицы времени и перекроется на единицу времени с линией, соответствующей предыдущему клиенту. Она перекроется также и с линией, соответствующей следующему клиенту. Чтобы получить относительную частоту ожидания группой из двух клиентов для любого момента времени, берется отношение общей длительности ожидания группами из двух клиентов ко всему времени.

При повторении такого построения для новых данных возникнет новая ситуация. При достаточном числе повторений для практического случая можно оценить вероятность того, что в данный момент времени ожидает данное число клиентов. Эта вероятность отличается от предыдущей, которая вычисляется для любого момента времени. Она находится подсчетом частоты случаев, когда в данный момент времени ожидает один клиент, группа из двух клиентов и т. д.

Итак, существует три вида вычисляемых характеристик для числа ожидающих клиентов: 1) частота появления данной группы клиентов, ожидающих совместно; 2) частота появления данной группы клиентов в любой момент времени и 3) частота появления данной группы в данный момент времени.

Заметим, что ситуация массового обслуживания с точки зрения перегрузок должна изучаться с учетом того периода времени, в который требуется данное действие. Например, перегрузка ресторанов происходит в полдень и вечером. Иногда целесообразно рассмотреть обе перегрузки совместно вследствие различной интенсивности потока и различного колебания его в каждый из этих двух периодов. Если бы перегрузка не зависела от времени суток, т. е. была бы однородной во времени, дело обстояло бы проще. Однако необходим тщательный подход к вопросам правильного изучения и разделения периодов, в которые происходит перегрузка в данной операции.

В качестве последнего замечания укажем, что из приведенных данных можно найти интенсивность поступления клиентов и интенсивность обслуживания. Если бы данные были более обширными, то можно было бы получить как распределение моментов прибытия клиентов, так и распределение времени их обслуживания.

1.3. Элементы систем массового обслуживания

Система массового обслуживания включает четыре элемента: входящий поток, очередь, обслуживающее устройство и выходящий поток. С каждым из них связан ряд возможных допущений относительно протекания процессов массового обслуживания. Некоторые из этих допущений, как указывается в историческом обзоре (§ 1.6), специально исследовались. Другие допущения приводят к еще нерешенным задачам обслуживания, требующим исследования. Дадим общее описание различных вариантов систем массового обслуживания при некотором повторении понятий, приведенных в предыдущем параграфе.

1. Виды распределения входящего потока и времени обслуживания

В начальный момент в системе может быть некоторое число ожидающих требований. Следующее требование поступает через случайное время, обладающее определенной плотностью вероятности. Случайны также и длительности промежутков между моментами поступления последовательных требований. Эти промежутки времени для многих приложений могут считаться взаимно независимыми; однако, например, в случае потока транспорта, проезжающего у светофора, они являются зависимыми. Эти же замечания относятся к моментам поступления на обслуживание и к длительности обслуживания.

2. Начальное число требований и входящий поток

Число требований, находящихся в системе к началу процесса обслуживания, может быть задано законом распределения, так как оно может быть переменным для каждого начала операции (например, в тот или иной день). В систему массового обслуживания могут поступать требования из конечной или бесконечной совокупности, которая может состоять из различных категорий требований. Требования каждой из категорий могут поступать с различным распределением, поодиночке или в составе группы и занимать место в очереди в установленном порядке. Распределение входящего потока может зависеть от распределения выходящего потока, как, например, в больнице, куда пациенты принимаются при наличии освободившихся коек.

3. Поведение клиента

Отказ клиента становиться в очередь. Поведение клиента может изменяться. Прибывающие клиенты могут не становиться в очередь вследствие размеров очереди или просто потому, что они вообще не могут ожидать. Эти клиенты для системы теряются. Иногда потеря требования происходит оттого, что ожидание не имеет смысла, как в случае занятой телефонной линии, хотя вызов можно повторить. Можно также перенести этот вызов, отложив его до того времени, когда линия будет свободна. Может быть одна очередь, в которой клиенты ожидают обслуживания, или несколько очередей, как в банках или магазинах самообслуживания. Клиент может присоединиться к ближайшей очереди, независимо от ее длины. Хотя прибытие клиентов ожидается через равные промежутки времени, они могут все же поступать позже или раньше в соответствии с определенным распределением отклонений относительно заданных моментов поступления как средних значений.

Влияние неполной информации. Во многих задачах может потребоваться принять решение — к какой из нескольких очередей системы присоединиться, если в данный момент времени в распоряжении клиента имеется информация только о некоторых из них. Это случай неполной информации. При перегруженности уличного движения отсутствие сведений о том, какой дорогой лучше проехать, не проверяя всех их, — также случай «неполной информации».

Приспособление клиента к условиям очереди. Основываясь на опыте, пассажир может знать, когда — позже или раньше — ему нужно выезжать, чтобы избежать простаивания в очереди. Такие меры, если они достаточно изучены, могут ослаблять перегрузки. Например, можно наблюдать, как суда, приближающиеся к Суэцкому каналу, замедляют ход до тех пор, пока очередь у Порт-Саида сократится до приемлемых размеров. Клиент может присоединиться к длинной очереди тогда, когда обслуживание еще не начиналось, из опасения, что обслуживание в короткой очереди может неожиданно прекратиться — это известно из опыта. Существуют ситуации, когда клиент, прибывший раньше другого, должен быть и обслужен раньше. Имеют место случаи, когда каждое обслуживающее устройство имеет свою специфику и, следовательно, свою собственную очередь, как, например, окно продажи марок, окно почтовых переводов или окно заказных писем на почтамте.

Соглашение между клиентами. Переход клиентов из одной очереди в другую. Уход из очереди до начала обслуживания. Несколько клиентов могут договориться о том, что только один из них будет ожидать в очереди, а остальные в это время будут свободны и могут заняться другими делами. В некоторых случаях может быть организовано поочередное ожидание в очереди. Клиенты могут переходить из одной очереди в другую, как в банке. Клиент может потерять терпение и уйти из очереди.

4. Варианты систем и каналов массового обслуживания

Полная и ограниченная доступность. Обслуживающие каналы могут быть доступны любому требованию, ожидающему в системе (полнодоступная система), или могут быть доступны только некоторым из них. Другие требования задерживаются и вынуждены ожидать до тех пор, пока канал, который производит требуемое обслуживание, не станет доступным. В системах телефонной связи наличие незанятого соединения зависит от того, может ли незанятый вход в следующую линию сочетаться с любым незанятым выходом. Сама идея полной доступности связана с необходимостью в целях экономии допустить все возможные сочетания. Это особенно важно для вызовов при дальней связи, когда они должны проходить более чем через один коммутатор.

Способ обслуживания, или дисциплина очереди. Выбор из очереди на обслуживание и распределение клиентов по каналам может производиться в порядке прибытия, случайным образом, или же за клиентом может быть закреплен приоритет (при этом могут допускаться ошибки, когда вначале неясно, за кем закреплен приоритет, или когда закрепление приоритета может меняться с течением времени). И, наконец, выбор клиентов на обслуживание может производиться по принципу «прибыл последним — обслужен первым». Обычно допускается, что как только освобождается канал, способный производить обслуживание ожидающего требования, оно немедленно, без потери времени поступает на обслуживание.

Объединение очередей. Существуют различные способы объединения очередей. При этом достигается некоторое сокращение среднего времени ожидания, особенно когда велик разброс времени ожидания у обслуживающего устройства, перед которым образуется отдельная очередь.

Обслуживающие устройства с последовательными и параллельными каналами. Обслуживающее устройство может состоять из нескольких параллельных каналов; при этом некоторые из них могут соединяться последовательно с другими каналами или же несколько параллельных каналов могут вести к нескольким последовательным каналам. В системе с последовательными каналами очередь может разрешаться перед каждым каналом или только перед некоторыми из них. В магазине самообслуживания прибывший клиент обслуживает себя сразу же после прибытия, и, таким образом, число каналов обслуживания меняется в зависимости от числа клиентов (хотя число кассовых аппаратов остается неизменным). Все клиенты выстраиваются в очередь у касс для вторичного обслуживания. При обслуживании различных категорий клиентов канал может иметь различное распределение времени обслуживания. Когда очередь отсутствует, свободные обслуживающие устройства могут выполнять другие задания; например, при ремонте машин в этом случае выполняются вспомогательные работы. Часто это зависит от требуемого объема обслуживания, пропускной способности обслуживающего устройства и размеров очереди. Сам канал обслуживания может перемещаться, как, например, лента конвейера с устройством, производящим обработку деталей, установленных на ленте.

Специализированные обслуживающие каналы. Некоторые обслуживающие каналы могут специализироваться, в то время как другие каналы остаются общими, как это имеет место, например, при обслуживании пассажиров в аэропорту, где у некоторых окон производится обслуживание только тех пассажиров, чье время отлета находится в заданном интервале. Потребность прибывшего клиента в специальном обслуживании может изменяться в зависимости от длительности промежутка времени между моментом прибытия его и моментом отправления самолета. Параллельные каналы могут соединяться вместе для выполнения нескольких видов обслуживания. Клиенты могут возвращаться в очередь для дополнительного обслуживания, образуя таким образом цикл.

Взаимодействие очередей. Две очереди могут воздействовать друг на друга, как это наблюдается в том случае, когда один узкий проезд на части шоссе должен использоваться автомобилями, движущимися в обоих направлениях. На одной из сторон проезда транспорт должен ожидать, чтобы пропустить автомобиль, движущийся навстречу. Если же прежде, чем этот автомобиль проедет, с той же стороны подойдут другие автомобили, то и они также будут пропущены, весьма вероятно, группами. Все это можно отнести и к Суэцкому каналу, хотя там в настоящее время дело обстоит иначе — суда проходят по расписанию.

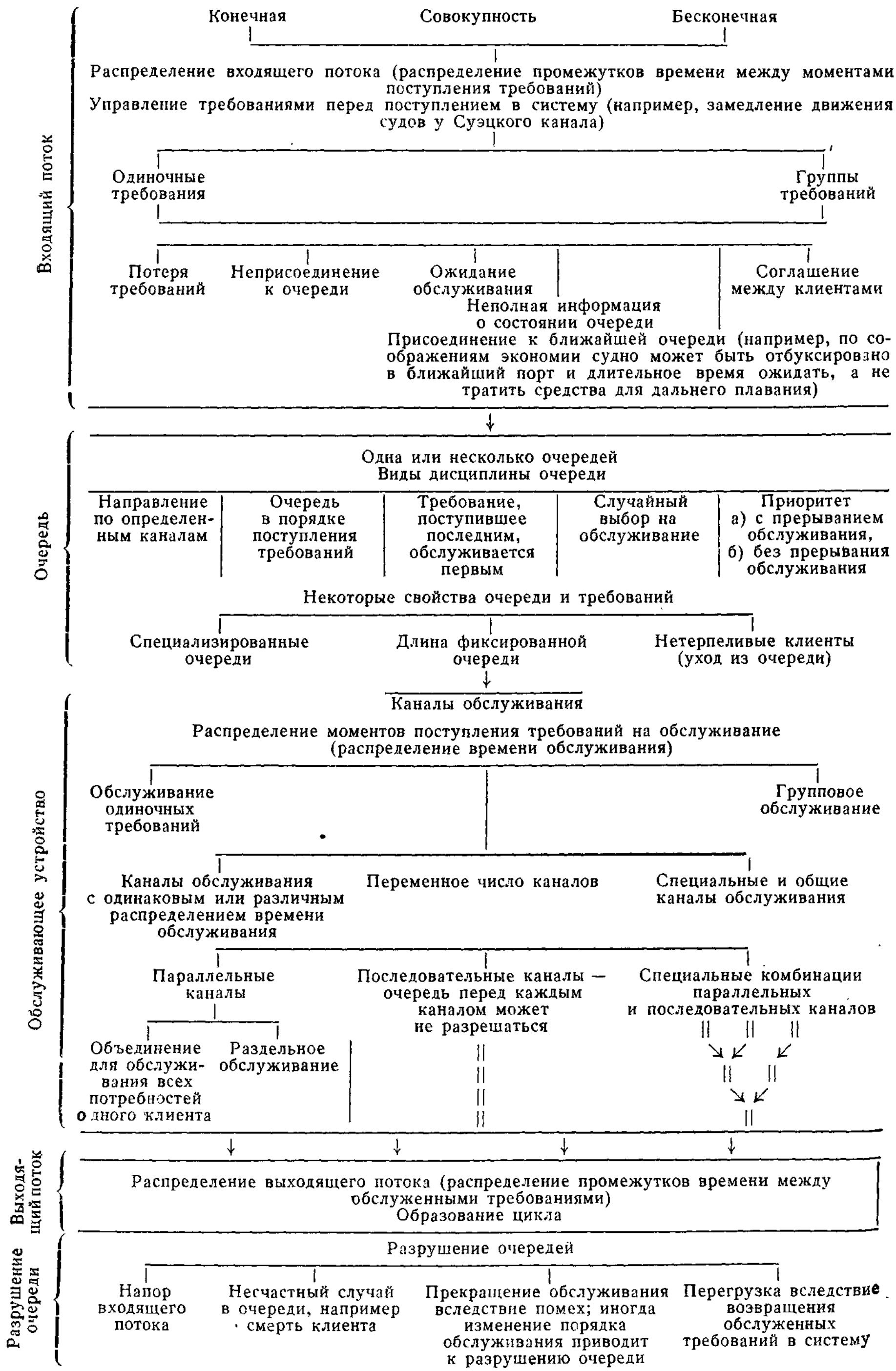
Входящий поток
Конечная
Совокупность
Бесконечная
Распределение входящего потока (распределение промежутков времени между моментами поступления требований)
Управление требованиями перед поступлением в систему (например, замедление движения судов у Суэцкого канала)
Одиночные требования
Группы требований
Потеря требований
Неприсоединение к очереди
Ожидание обслуживания
Неполная информация о состоянии очереди
Соглашение между клиентами
Присоединение к ближайшей очереди (например, по соображениям экономии судно может быть отбуксировано в ближайший порт и длительное время ожидать, а не тратить средства для дальнего плавания)
Очередь
Одна или несколько очередей
Виды дисциплины очереди
Направление по определенным каналам
Очередь в порядке поступления требований
Требование, поступившее последним, обслуживается первым
Случайный выбор на обслуживание
Приоритет а) с прерыванием обслуживания, б) без прерывания обслуживания
Некоторые свойства очереди и требований
Специализированные очереди
Длина фиксированной очереди
Нетерпеливые клиенты (уход из очереди)
Обслуживающее устройство
Каналы обслуживания
Распределение моментов поступления требований на обслуживание (распределение времени обслуживания)
Обслуживание одиночных требований
Групповое обслуживание
Каналы обслуживания с одинаковым или различным распределением времени обслуживания
Переменное число каналов
Специальные и общие каналы обслуживания
Параллельные каналы
Последовательные каналы — очередь перед каждым каналом может не разрешаться
Специальные комбинации параллельных и последовательных каналов
Объединение для обслуживания всех потребностей одного клиента
Раздельное обслуживание
Выходящий поток
Распределение выходящего потока (распределение промежутков времени между обслуженными требованиями)
Образование цикла
Разрушение очереди
Разрушение очередей
Напор входящего потока
Несчастный случай в очереди, например смерть клиента
Прекращение обслуживания вследствие помех; иногда изменение порядка обслуживания приводит к разрушению очереди
Перегрузка вследствие возвращения обслуженных требований в систему

Рис. 1.1.

5. Выходящий поток

Выходящий поток также может играть важную роль, особенно когда он сам образует входящий поток для другой очереди, последовательно соединенной с первой. Распределение входящего потока и распределение времени обслуживания могут зависеть друг от друга. Например, такая зависимость существует, когда обслуживание влияет на поток требований, и наоборот. Так, если несколько конкурирующих предприятий одинакового характера рассматривать как ряд параллельных каналов, то больший наплыв клиентов будет там, где производится экстренное или хотя бы более быстрое обслуживание.

Таковы некоторые примеры различных ситуаций, к которым может быть применена теория массового обслуживания. Впоследствии многие из этих положений будут исследованы.

На рис. 1.1 в краткой форме представлены многие из приведенных выше понятий. Сделана попытка показать возможные факторы, оказывающие влияние на входящий поток, на время поступления требований и т. д.; показаны различные дисциплины и особенности очередей, каналы обслуживания и, наконец, выходящий поток. В качестве процесса, противоположного обслуживанию, указывается разрушение очереди, поскольку не всегда операция обслуживания будет протекать нормально. Приведенная связь понятий не является жесткой. Читатель может попытаться найти другую схему.

1.4. Вводный описательный анализ

В следующей главе будет рассмотрен случай, когда поступление требований и обслуживание их происходит через постоянные промежутки времени. Хотя мы считаем, что промежутки времени между моментами поступления требований имеют одинаковую длительность и промежутки времени обслуживания также равны, очевидно, что это допущение не является обязательным при нахождении числа ожидающих требований и времени ожидания.

Обычно в системах массового обслуживания не существует такой запланированной регулярности. Требования могут поступать случайным образом, и поэтому нельзя быть заранее уверенным в точном времени их прибытия. Это справедливо также и для времени обслуживния. Когда требования ожидают, то существует вероятность того, что в любой момент времени в очереди находится определенное число их, так как поступление требований и время обслуживания случайны. Поэтому при теоретическом рассмотрении может быть получена только общая оценка процесса. Из нее для конкретного требования можно, например, получить представление о среднем времени ожидания и отклонении от него.

Очевидно, что с каждым требованием связан ряд случайных величин. Например, одна случайная величина описывает время поступления n-го требования, другая — его время ожидания, третья — время обслуживания и т. д. Каждое заданное состояние группы требований, например поступление их, описывается семейством случайных величин. Каждая случайная величина в произвольный момент времени может иметь различное распределение вероятностей (характеристику n-го требования). Вследствие зависимости их от времени и вследствие того, что, по крайней мере в теоретическом плане, их может быть бесконечное множество, эти семейства случайных величин принимают вид вероятностных процессов. Таким образом, вероятностный процесс есть семейство случайных величин, зависящих от параметров. Здесь рассматривается единственный параметр — время. Изучение теории массового обслуживания связано с нахождением распределений различных комбинаций сумм и разностей этих случайных величин при наличии достаточной информации о них.

Основная проблема заключается в том, чтобы 1) связать эти величины в правильной последовательности для описания задач массового обслуживания; 2) найти соответствующие функции распределения, основываясь на статистических наблюдениях; 3) использовать эти функции распределения для нахождения характеристик процесса обслуживания и 4) на основании полученных характеристик принять необходимые меры для улучшения процесса обслуживания.

Интересной особенностью теории массового обслуживания является зависимость различных величин от рассматриваемого момента времени. Например, число покупателей в очереди у кассы магазина самообслуживания через 10 минут после начала работы будет иным, чем через 2 часа. В каждый момент времени существует определенная вероятность для любого значения длины очереди или времени ожидания. Таким образом, функция распределения будет изменяться с течением времени.

Следовательно, для n-го требования существует случайная длительность ожидания, которая принимает различные значения с вероятностями, зависящими от того, когда рассматривается процесс. В момент времени $t=0$ в системе может находиться начальное число требований, ожидающих в очереди. Позже система может освободиться от влияния этих требований или влияния отсутствия требований к началу обслуживания и будет рассматриваться только переменный поток новых требований, и в конечном счете устанавливается состояние, не зависящее от числа начальных требований. Когда влияние начального состояния системы «стирается», можно получить модель системы, в которой распределение вероятностей в фиксированный момент времени будет таким же, как и в любой другой момент. Таким образом, закон распределения состояний системы перестает изменяться со временем; устанавливается равновесный, или стационарный режим с определенными вероятностями состояний.

Система никогда не войдет в стационарный режим, если, например, время обслуживания больше, чем промежутки времени между моментами поступления требований. Всегда необходимо установить, действительно ли существует стационарный режим, т. е. при каких условиях распределения состояний системы становятся независимыми от времени. В некоторых случаях этому вопросу уделяется главное внимание. На практике во многих процессах это равновесное состояние с некоторым приближением наступает, и поэтому распределение времени ожидания и распределение числа требований, находящихся в системе, изучается для равновесного состояния системы.

Заметим, что равновесие означает, что вероятности не зависят от времени, но это не значит, что система становится детерминированной. Система продолжает изменяться, но распределения, описывающие ее, будут постоянными во времени. Можно вычислять средние значения, отклонения от них и т. д.

Пояснение понятия стационарного состояния можно провести на примере работы установки для кондиционирования воздуха, которой мы пользуемся в жаркое время года. Это устаревший прибор, и поэтому он с трудом поддается управлению. Начальная нагрузка буквально сотрясает его. Он начинает работать с пронзительным шумом, вибрациями и грохотом. Через некоторое время создается впечатление, что все встало на свое место, но затем снова прибор возвращается в начальное состояние. Спустя длительный промежуток времени он начинает разогреваться, шум утихает, колебания гасятся.

В таком приборе все же никогда не может установиться стационарное состояние, под влиянием начальных условий он все время будет находиться в переходном состоянии. В жаркое время, когда он должен работать с бо́льшим напряжением, это все, на что он способен. Так может случиться и в некоторых системах массового обслуживания; вследствие зависимости от распределения входящего потока, распределения времени обслуживания и начальных условий система может не достигнуть стационарного состояния.

Нередко решение задачи можно получить гораздо проще, если допустить, что состояние системы не зависит от времени, в этом случае задача решается для стационарного состояния. Но не всегда можно ограничиться решением только для стационарного состояния, поэтому нередко нужно иметь решения, зависящие от времени.

Вследствие однородности требований нередко принимается, что они имеют одинаковое распределение моментов поступления в систему и одинаковое распределение времени обслуживания, не зависящие от того, поступило ли требование n-м или m-м по счету. Во многих задачах принимается, что в малом временно́м интервале появление более одного требования может произойти лишь с вероятностью меньшего порядка малости, чем длина этого интервала. Эти допущения делаются в зависимости от практической целесообразности и от того, насколько трудно получить решение задачи в более общих условиях.

В качестве иллюстрации зависимости некоторых величин заметим, что время ожидания (если имеется очередь) $(n+1)$-го требования, w_{n+1}, равно сумме времени ожидания n-го требования, w_n, и его времени обслуживания s_n минус промежуток времени t_n между моментами поступления n-го и $(n+1)$-го требований. Время поступления первого требования не оказывает влияния, поскольку это требование сразу же поступает на обслуживание, следовательно,

$$w_{n+1} = w_n + s_n - t_n.$$

Теперь каждая из случайных величин s_n и t_n будет иметь заданное распределение; следовательно, w_n и w_{n+1} также будут обладать функциями распределения. Эти неизвестные распределения можно определить. Поэтому необходимо изучить некоторые элементарные понятия, применяемые в теории вероятностей, и использовать их в дальнейшем.

Заметим, что если нет мест для ожидания, то требования, которые не могут сразу поступить на обслуживание, будут потеряны. Каналы обслуживания могут иметь сложную структуру, они могут, например, соединяться последовательно или параллельно. В этом случае задача становится еще более важной и более интересной.

В качестве примера важной характеристики можно указать время ожидания. Если обслуживание производится по принципу «первым пришел — первым обслужен», то время ожидания в очереди k-го требования равно сумме времени обслуживания $k-1$ предыдущих требований и времени, необходимого для окончания обслуживания того требования, которое уже находилось в системе в момент поступления k-го требования. Каждая из этих величин есть случайная величина, принимающая значения в соответствии со своей функцией распределения. Задача состоит в определении вероятности того, что сумма этих величин примет любое заданное значение.

Распределение времени ожидания получим, умножив вероятность того, что в очереди будет n требований, на распределение времени ожидания n-го требования и просуммировав результат по всем возможным значениям n. Затем из распределения времени ожидания путем усреднения по времени можно получить среднее значение времени ожидания. Эти вопросы поясняются в гл. 3. Попутно заметим, что если требования выбираются на обслуживание случайным образом, то распределение времени ожидания требования, которое поступило после того, как в очереди уже было $n-1$ требований, будет отличаться от распределения, которое было получено для системы с обслуживанием требований в порядке поступления.

Важной проблемой является использование результатов аналитических решений в практических ситуациях. Например, для практического случая может возникнуть задача оценки интенсивности поступлéния требований и интенсивности обслуживания по методу наибольшего правдоподобия. Как следует производить фактические измерения, чтобы получить минимальную погрешность? Если уравнения не решаются аналитически, то можно попытаться получить численные решения.

До сих пор рассматривался аналитический подход, т. е. имелось в виду, что ситуации массового обслуживания описываются, по возможности, аналитически; при этом находятся такие показатели, как среднее время ожидания и средняя длина очереди, а из них выводятся важные показатели для оценки эффективности систем массового обслуживания. Нередко случается, что задача массового обслуживания является сложной и не может быть описана аналитическими уравнениями, оперируя с которыми можно получить необходимые решения. Тогда приходится пользоваться другим методом, а именно моделированием системы. Применение этого метода было показано ранее на числовом примере, в котором численные значения времени ожидания и т. д. получены непосредственно. Эксперимент должен повторяться достаточно большое число раз, чтобы получить большие выборки, и большое число статистических результатов, которые затем определенным образом усредняются для получения необходимых величин. Этот метод имеет большое значение для практики и применяется для нахождения непосредственного решения сложных задач. Однако у него есть и недостатки, и поэтому при построении правильной модели с достаточным объемом выборок необходима особая тщательность; нужно уметь правильно объединять результаты для нахождения решения.

1.5. Показатели эффективности обслуживающих устройств и некоторые меры их улучшения

Излагаемая нами теория основывается на исследовании различных моделей образования очереди. Подобный подход является промежуточным между пассивным наблюдением перегрузок систем и синтезом оптимальных систем. При последнем подходе рассматриваются конкретные состояния системы, возникающие на практике, решение находится путем пересмотра всех возможных направлений, какими бы странными они ни казались, при этом наиболее нереальные отбрасываются, а затем, в манере Шерлока Холмса, принимается то, что остается, каким бы необычным оно ни казалось.

Такой подход рекомендуется для практических целей, однако он может оказаться рискованным. Чтобы получить относительное улучшение обслуживания на основе новых идей, нужен аналитический подход к проблеме в целом. Для эффективного решения трудных задач массового обслуживания эти два метода должны комбинироваться.

Большое практическое значение теории массового обслуживания уже доказано. Это видно из того множества задач, для решения которых она успешно применяется. Это, в частности, относится и к тому обстоятельству, что существует большое число важных показателей эффективности средств обслуживания. Необходима большая тщательность при выборе соответствующих показателей, а следовательно, и модели, описывающей процесс, на основании которого такой показатель выводится.

Эффективность некоторых систем массового обслуживания повышается слабо потому, что не замечают лучших показателей, способствующих улучшению работы системы. Например, может быть принято решение регулировать входящий поток вместо увеличения интенсивности обслуживания или же и то и другое; таким образом добиваются уменьшения числа требований, находящихся в системе. В отдельных случаях в периоды сильной интенсивности входящего потока может оказаться более целесообразным иметь соответствующее место для ожидания поступающих требований, чем ускорять обслуживание.

Вопрос о том, насколько полно теоретическая модель соответствует реальному процессу массового обслуживания или части его (если он может быть разделен на части, каждая из которых удовлетворяет, скажем, условиям равновесия), определяется главным образом тем, какой показатель необходимо принять. Для многих практических целей требуются такие показатели качества обслуживания, которые обеспечивают возможность сравнения. Например, может производиться сравнение влияния распределения времени обслуживания на распределение времени ожидания. В другом случае может сравниваться влияние на время ожидания различной дисциплины очереди.

Какой бы подход ни применялся при решении таких задач, необходимо выбирать такие показатели, с помощью которых можно принять правильное решение. Например, владелец системы обслуживания может путем сравнения расходов на увеличение объема обслуживания и убытков вследствие потери клиентов решить, стоит ли ему увеличивать место для ожидания или же лучше увеличить число обслуживающих устройств.

Среднее время ожидания для тех клиентов, которым приходится ожидать, может быть слишком большим, и, следовательно, среднее время ожидания всеми клиентами может оказаться неподходящим показателем. Владелец предприятия может использовать эти величины для определения необходимых размеров места для ожидания, и принять меры, чтобы сделать ожидание более приятным и, может быть, более коротким за счет увеличения числа каналов, более быстрого обслуживания и т. д.

Важно также рассматривать показатели эффективности, относящиеся и к тем клиентам, которые решают уйти из очереди или вообще не становиться в очередь, если длина ее или время ожидания слишком велики.

Назовем некоторые важные показатели, которые находятся при исследовании систем в стационарном состоянии и в переходном состоянии (зависящем от времени): интенсивность входящего потока; интенсивность обслуживания; загрузка системы, т. е. отношение среднего времени обслуживания к среднему промежутку времени между двумя последовательными требованиями; вероятность того, что в очереди или в системе находится определенное число требований n; среднее число требований, находящихся в очереди (или в системе, т. е. сюда включается еще и среднее число требований, находящихся на обслуживании); распределение времени ожидания в очереди или распределение времени пребывания в системе; среднее время ожидания и его дисперсия. В том случае, когда приходится иметь дело с нетерпеливыми клиентами или скоропортящимися продуктами, важно знать, чему равна вероятность того, что ожидание продлится больше допустимого времени.

Важными показателями являются также вероятность обслуживания без ожидания (доля требований, ожидающих обслуживания, находится путем вычитания этой величины из единицы); вероятность того, что требования будут ожидать положительное время; среднее число требований, ожидающих обслуживания; средняя длительность периода занятости; среднее число обслуженных требований. Для тех требований, на обслуживание которых затрачивается мало времени, особенно важно знать отношение среднего времени ожидания к среднему времени обслуживания. Важно вычислить также вероятность того, что занято ровно $0 \leqslant n \leqslant c$ каналов; среднее число незанятых каналов; коэффициент простоя каналов; коэффициент занятости (т. е. долю времени, в течение которого канал занят); среднее число требований, не поступивших в систему, применительно к случаю конечной совокупности; вероятность потери требования и т. д.

Во многих задачах теории массового обслуживания существенную роль играет анализ стоимости. Может возникнуть необходимость знать стоимость всей операции, а также определить, какое влияние на стоимость обслуживания окажет оборудование дополнительного канала по сравнению с увеличением объема обслуживания в канале, выбранная дисциплина очереди, ожидание клиента в течение определенного времени, содержание большего числа каналов по сравнению с использованием места для ожидания в других целях, размещение очереди различной длины и наличие места для ожидания различной вместимости, управление распределением входящего потока (например, можно выбрать распределение времени прибытия самолетов и затем сравнить стоимость процесса при заданном распределении времени обслуживания для случайного и регулярного входящих потоков). Кроме того, можно рассматривать показатели использования обслуживающего устройства и производительность (число требований, обслуженных в единицу времени). Важно также знать возможные убытки вследствие потери клиентов, а также представлять себе, во что обходится клиентам ожидание в очереди.

По различным причинам обслуживающее устройство может разрушиться. Например, стремительный входящий поток может оказать большое давление на систему, принуждая прекратить процесс обслуживания. Примером разрушения обслуживающего устройства может служить временное повреждение телефонной линии во время урагана. Коренные изменения порядка обслуживания могут полностью отпугнуть клиентов, и весь процесс прекратится. Неожиданный несчастный случай в банке (например, убийство клерка бандитом) может повлечь за собой прекращение всех операций. Перегрузка в конце обслуживания вследствие того, что обслуженные требования не покидают систему, или вследствие других причин может в конце концов остановить обслуживание.

Решительные меры по повышению эффективности систем массового обслуживания могут быть осуществлены путем сокращения времени обслуживания или уменьшения отклонений от среднего значения, применения дополнительного обслуживания в моменты наибольшего наплыва и управления распределением входящего потока. Выигрыш получается, например, при использовании упорядоченного входящего потока вместо случайного, упорядоченный поток, как мы увидим позже, приводит к меньшему времени ожидания при данном времени обслуживания. Ясно, что такая мера может привести только к частичному эффекту. Требования будут поступать с отклонениями относительно регулярных промежутков времени. Желательно также сглаживать нерегулярности поступления требований. Важно производить такое управление входящим потоком, чтобы загрузка системы (отношение интенсивности поступления требований к интенсивности обслуживания) оставалась меньше единицы. Упорядочение входящего потока и использование обслуживающих устройств в течение более длительного промежутка времени ослабляет тяжелые перегрузки в часы пик.

Позже будет показано, как с помощью применяемых моделей получить выражения для различных показателей эффективности функционирования систем массового обслуживания. Имеются примеры, когда удовлетворительное решение задачи массового обслуживания можно получить только при замене данного процесса более эффективным. Примером заботы о клиенте в некоторых обслуживающих устройствах является внедрение самообслуживания. Эстакады и современные шоссе с многорядным движением устраняют транспортные перегрузки, которые могут приводить к затруднениям даже в том случае, если движение в узких проездах хорошо регулируется. По телефону можно сделать предварительный заказ или назначить время приема. Таким образом, потери времени на обслуживание существенно уменьшаются.

Заметим, что, решив практическую задачу массового обслуживания путем увеличения числа лиц, производящих обслуживание, можно создать другую очередь, а именно очередь тех, кто производит обслуживание. Теперь они должны ожидать клиентов, нуждающихся в их услугах. В этом смысле может создаться перегрузка иного рода, когда каналы простаивают, хотя, возможно, перегрузка первого рода менее желательна.

1.6. Исторический обзор

1. Телефонные системы

Теория массового обслуживания получила развитие при решении проблем, связанных с перегрузкой телефонных линий. Это направление развивается также и в настоящее время; им занимаются многие способные исследователи в различных странах. В кратком историческом обзоре развития теории массового обслуживания целесообразно изложить некоторые основные понятия телефонной связи, представляющие особый интерес для исследователей.

Сразу же отметим два основных положения. Когда в группе все аппараты включены, то говорят, что группа занята. Если требование поступает, когда группа занята, то оно может или получить отказ или ожидать обслуживания. Первый случай характерен для систем с потерями, а второй — для систем с ожиданием; возможны также и системы смешанного типа.

Полнодоступные системы. Полнодоступный пучок линий представляет собой систему аппаратов (телефоны, междугородные или внутренние линии связи и т. д.), каждый из которых доступен любой группе источников (абонентов).

$\underline{A}.\ \underline{a}\ \underline{b}\ \underline{c}$

Рис. 1.2. Полная доступность.

На рис. 1.2 прописной буквой обозначена группа источников, а строчными буквами — соответствующие им аппараты. Заметим, что если абоненту A нужно установить связь, то это может быть сделано с помощью любой свободной линии a, b, c и т. д. Если же все линии заняты, то абонент либо ожидает, либо его вызов теряется, т. е. в этом случае он получает сигнал, что линия занята. В любом случае важной величиной является вероятность ожидания вызова в течение определенного времени или вероятность отказа.

Для случая ограниченного источника, включающего N абонентов, n из которых уже обслуживается, Эрланг нашел, что вероятность поступления нового вызова в промежутке времени $(t,\ t+\Delta t)$ равна $(N-n)\lambda dt$, где λ — среднее число вызовов, поступающих в единицу времени. Если число обслуживаемых независимых абонентов неизвестно, то обычно принимают предположение, что вероятность поступления в интервале $(t,\ t+\Delta t)$ одного вызова равна λdt, а интервалы между последовательными вызовами имеют экспоненциальное распределение с плотностью $\lambda e^{-\lambda t}$. Это приводит к пуассоновскому распределению, для которого вероятность поступления n вызовов за время t составляет $\frac{(\lambda t)^n e^{-\lambda t}}{n!}$, где λt — математическое ожидание числа вызовов за время t. Это распределение было проверено на практике для периодов интенсивной нагрузки. Оно удовлетворительно описывает распределение входящего потока. Аналогично, если $\frac{1}{\mu}$ — средняя продолжительность разговора, то вероятность того, что разговор окончится в промежутке $(t, t+\Delta t)$, равна $\mu\Delta t$ и не зависит от длительности разговора и наличия других вызовов. Это свидетельствует о том, что длительность разговора распределяется в соответствии с выражением $\mu e^{-\mu t}$.

Используя эти допущения, Эрланг [89] развил теорию телефонной связи и получил формулы для вероятностей различного числа ожидающих абонентов, распределения времени ожидания при равновесном состоянии системы и вероятности отказов для системы с потерями. Он опубликовал свои труды в 1917 г. вначале на датском языке, а затем на английском, немецком и французском языках. Эрланг рассматривал пуассоновский входящий поток из бесконечного источника при экспоненциальном и постоянном времени обслуживания.

Он получил решения для системы с ожиданием при постоянном времени обслуживания в случае одного, двух и трех каналов, а также для произвольного числа каналов при экспоненциальном времени обслуживания.

Труды Эрланга послужили толчком для других работ в этом направлении. В этой области работали Фрай (см. его блестящую книгу) [271], Молина [575] и О'Делл [618]. Изучение первых статей по теории массового обслуживания приводит к выводу, который сделал Сиски [791], что эти работы были связаны с одобрением или опровержением результатов Эрланга. Сиски называет этот период «периодом Эрланга и О'Делла». О'Делл опубликовал краткое изложение работ по теории скученности, вышедших до 1920 г. Полячек вывел известную формулу для одноканальной системы с пуассоновским входящим потоком и произвольным временем обслуживания (формула Полячека—Хинчина); он исследовал также случай постоянного времени обслуживания для системы с обслуживанием требований в порядке поступления и при разветвлении очереди перед каждым обслуживающим устройством. Кроммелин исследовал многоканальную систему с ожиданием в состоянии равновесия при постоянном времени обслуживания. Его результаты совпали с результатами некоторых задач, исследованных Эрлангом. Беркели, а также Шаутен и Джилтей проверили эти результаты экспериментально.

Совсем недавно Полячек [679] аналитически исследовал многоканальную систему с произвольным временем обслуживания, в которую поступает поток с последействием.

Пальм рассмотрел влияние изменения интенсивности потока обслуженных требований. Костен произвел статистические расчеты моментов для времени ожидания, для числа отказов и т. д. Вилкинсон и Хейуорд изучили влияние надежности поиска телефонных линий на распределение числа занятых линий.

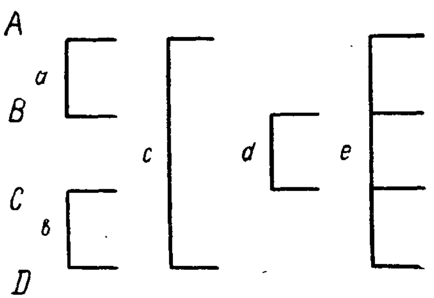

Рис. 1.3. Ограниченная доступность.

Неполнодоступные системы. В неполнодоступных системах только некоторые аппараты доступны абонентам. На рис. 1.3 показана типичная, хотя и сложная неполнодоступная система с так называемым ступенчатым включением. Например, линия a доступна только для вызовов, поступающих от абонентов A и B, линия e доступна всем абонентам, а линия c доступна абонентам A и D. Для этой системы также исследовались вопросы отказов (главным образом) и ожидания (в меньшей степени).

Допуская, что новые вызовы находят заданный аппарат среди большого числа абонентских телефонных аппаратов (так называемое идеальное ступенчатое включение), Эрланг получил выражение для вероятности того, что система окажется занятой. Строгий вывод этой формулы получен Брокмейером. В зависимости от типа соединений внутри системы были исследованы и другие вопросы.

Системы связи. Система связи — понятие, появившееся в более позднее время. В системе связи группа источников имеет ограниченную доступность к группе соответствующих им аппаратов; на рис. 1.4 представлена двухступенчатая система связи.

Существует зависимость распределения вызовов, поступающих в устройство a, от распределения вызовов в A. Задача формулируется в следующем виде: задано распределение потока вызовов, определить вероятность того, что

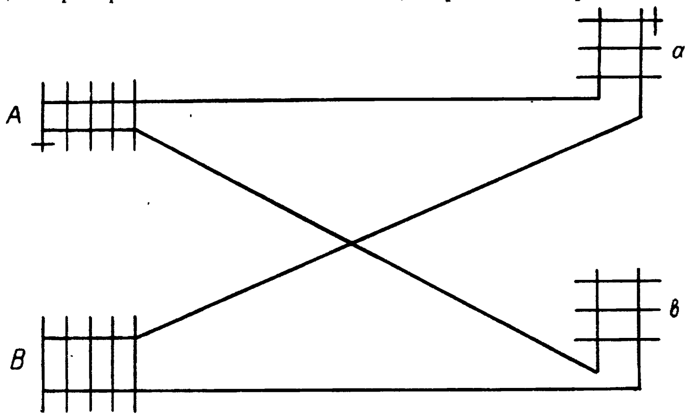

Рис. 1.4. Системы связи.

данный вызов не поступит на свободный вывод. Существует связь между распределениями вызовов на всех фазах. Якобеус, Иенсен и Форте решили эту задачу, приняв допущение, что эти функции распределения взаимно независимы, и определенная зависимость между ними существует только на последних двух ступенях. Для случая полной зависимости эта задача все еще остается нерешенной, но Эллдин предпринимает обнадеживающие попытки решить ее.

Двухкаскадная система. Рассмотрим теперь систему с потерями, в которой все требования перед поступлением на обслуживание направляются на общий пункт управления. Если пункт занят, то требование теряется, даже если обслуживающий канал свободен. Задача была решена для вероятности потери требования в системе с пуассоновским входящим потоком и экспоненциальным временем обслуживания во всех каналах для случаев, когда время обслуживания на пункте управления экспоненциальное (Воло и Леруа), постоянное (Форте) и произвольное (Полячек). Эту же задачу для системы с ожиданием решил Сиски как для случая одного канала, так и для бесконечного множества каналов.

Сиски [795] дает обширное изложение этого вопроса в главах, посвященных системам с потерями, системам с ожиданием, неполнодоступным системам, системам связи, а также взаимосвязанным коммутаторам.

2. Замечания по развитию общей теории массового обслуживания

Системы с обслуживанием в порядке поступления требований. В книге Феллера, изданной в 1950 г. [236] и впоследствии переработанной, исследовались задачи, связанные с очередями. С теоретической точки зрения значительный интерес представляет статья Кендалла, опубликованная в 1951 г. [432], в которой было введено широко используемое в настоящее время понятие вложенных цепей Маркова. За ней последовала его статья 1953 г., посвященная классификации систем массового обслуживания [433]. Следует заметить, что используемый им метод моментов регенерации был впервые предложен Пальмом. В 1952 г.

Кларк [121] впервые получил решение уравнений процесса рождения и гибели с постоянными коэффициентами для переходного состояния, опубликованное позже. Различные способы решения этих уравнений получили Ледерманн и Ройтер [493], Бейли [20], Морз [590], Чемпернаун [109]. Карлин и Мак-Грегор [409] предложили решение этой задачи с помощью коэффициентов, являющихся функциями числа требований в очереди. Саати [748] с помощью преобразований Лапласа получил решение для переходного состояния марковской модели многоканальной системы. Кларк [124], а затем Лучак исследовали задачу, в которой коэффициенты являются функциями времени. В 1952 г. Линдли [518] получил выражение для среднего времени ожидания в одноканальной системе для стационарного состояния, с произвольным входящим потоком и произвольным временем обслуживания при обслуживании требований в порядке поступления («первым пришел — первым обслужен»).

Случайный выбор на обслуживание. В 1942 г. Меллор [563] получил распределение времени ожидания для системы со случайным выбором из очереди. В 1946 г. Воло [853] дал точную формулировку этой задачи для системы с большим числом каналов при пуассоновском входящем потоке и экспоненциальном времени обслуживания. Позже эта задача была решена Полячеком [663] и Пальмом [640].

В 1953 г. Риордан разработал метод нахождения распределения времени ожидания, а Вилкинсон [882] получил кривые распределения. В 1959 г. Бёрк [100] исследовал задачу случайного выбора на обслуживание для одноканальной системы с пуассоновским входящим потоком и постоянным временем обслуживания.

Системы с приоритетом. В 1954 г. Кобхем опубликовал первую статью, посвященную системе с приоритетом, не прерывающим обслуживания, при пуассоновском входящем потоке и экспоненциальном времени обслуживания для одного и нескольких каналов. Он получил выражения для среднего времени ожидания требования, обладающего данным приоритетом. Холли упростил результаты Кобхема. Фиппс обобщил решение Кобхема для непрерывного числа приоритетов с приложением к ремонту машин. Морз [591] получил производящую функцию вероятностей стационарного состояния системы с двумя приоритетами, имеющими экспоненциальное время обслуживания с различными параметрами. В 1957 г. Кестен и Ранненберг исследовали задачу Кобхема более детально и получили преобразования Лапласа—Стилтьеса, а также первые два момента для времени ожидания в системе с произвольным временем обслуживания при стационарном состоянии.

В последнее время Дж. Джексон рассматривал системы с динамическим приоритетом [368]. Барри и Стефан исследовали систему с двумя приоритетами, прерывающими обслуживание, для случая пуассоновского входящего потока и экспоненциального времени обслуживания. Уайт и Кристай рассмотрели ту же задачу для двух каналов и получили выражения для среднего времени обслуживания и среднего времени ожидания в системе с любым приоритетом, а также производящие функции распределения времени ожидания. Они рассмотрели также сходство между этой системой и системой с разрушением очереди.

Миллер-младший [567] в обширной статье привел несколько полученных ранее результатов для системы с двумя приоритетами, имеющими экспоненциальное время обслуживания с различными параметрами, а также нашел преобразования Лапласа—Стилтьеса и производящие функции для новых случаев. Хиткоут в последнее время сделал значительный вклад в изучение систем с приоритетом, прерывающим обслуживание, исследовав многоканальную систему с несколькими приоритетами в переходном состоянии.

Отказ клиентов становиться в очередь. Уход из очереди до начала обслуживания. Нетерпеливые клиенты. Образование цикла. Переход клиентов из одной очереди в другую. Клиент может принимать решение относительно допустимой длины очереди, которая может быть и бесконечной, отказываясь присоединиться к более длинной очереди. В различных ситуациях приемлемая длина очереди различна. Общая совокупность ее размеров приводит к различным распределениям вероятности того, что поступающее требование отказывается становиться в очередь. В 1957 г. Хейт исследовал эту задачу для одноканальной системы с пуассоновским входящим потоком и экспоненциальным временем обслуживания для различных вероятностей того, что клиент отказывается становиться в очередь. Он определил среднее число требований, находящихся в очереди, при котором требование, которое становится в очередь, остается в системе.

Подобную задачу, в которой длина очереди не может превосходить фиксированного конечного значения, исследовал Финч [244] для случая входящего потока с ограниченным последействием. Он получил распределение длины очереди для стационарного состояния, доказав его существование.

В более ранних статьях Баррер [28] рассмотрел задачу о нетерпеливых клиентах, которые после определенного времени ожидания (детерминированное нетерпение) покидают находящуюся в стационарном состоянии одноканальную систему с пуассоновским входящим потоком и экспоненциальным временем обслуживания при случайном выборе требований на обслуживание. Он также исследовал задачу [29] для системы с обслуживанием по принципу «первым пришел — первым обслужен» в случае, когда требования покидают систему через определенный промежуток времени ожидания, и получил распределение длины очереди. Финч исследовал эту задачу для многоканальной системы с произвольным входящим потоком. Он также решил задачу для случая недетерминированного нетерпения в многоканальной системе с пуассоновским входящим потоком и экспоненциальным временем обслуживания, получив распределение длины очереди. Хейт рассмотрел комбинацию оставления очереди (нетерпения) с неприсоединением к очереди. Пальм исследовал случай раннего оставления очереди, а Коэн обобщил эту задачу для случая повторного поступления требований.

Процесс образования цикла требований в системе массового обслуживания был исследован Кенигсбергом в 1958 г. Глязер рассмотрел схему, когда возможен переход клиентов из одной очереди в другую в случае нескольких очередей, образованных перед системой, как, например, в банке.

Многофазовое обслуживание. В 1954 г. О'Брайен исследовал две системы многофазового обслуживания (обе фазы — в состоянии равновесия, с пуассоновскими входящими потоками и экспоненциальным временем обслуживания); им получены выражения для распределения длины очереди и среднего времени ожидания. В том же году Р. Джексон [369] решил задачу для ограниченного и неограниченного входящих потоков при двух и трех фазах, получив также распределение длины очереди. Хант исследовал двухфазные системы 1) с потерями, 2) с очередью, ограниченной конечной величиной, и 3) с неограниченной очередью и вычислил, в частности, среднее число требований в системе. Маршалл и Рейч исследовали время ожидания для многофазных систем. Бёрк [99], Рейч [708], Коэн и в последнее время Финч исследовали выходящий поток и установили независимо друг от друга, что выходящий поток для системы с пуассоновским входящим потоком и экспоненциально распределенным временем обслуживания также будет пуассоновским.

Немарковские модели. В немарковской модели либо поток непуассоновский, либо время обслуживания распределено по закону, отличному от экспоненциального, либо имеет место и то и другое. Классический вывод выражения для среднего числа требований, находящихся в очереди, и среднего времени ожидания для стационарного состояния системы с пуассоновским входящим потоком и произвольным временем обслуживания при обслуживании требований в порядке поступления сделал А. Я. Хинчин (формула Полячека—Хинчина). Как уже указывалось, Линдли вывел интегральное уравнение для распределения времени ожидания при произвольном входящем потоке и произвольном времени обслуживания. Дополнительно к ранним работам Кроммелина, который вычислил время ожидания для стационарного состояния многоканальной системы с пуассоновским входящим потоком и постоянным временем обслуживания, Эверетт предложил другой способ получения этих результатов, основанный на применении метода последовательных приближений.

Использовав метод Кендалла (метод вложенных цепей Маркова), Бейли [21] и Даунтон [188, 189] исследовали систему с групповым обслуживанием (обслуживается группа, состоящая из нескольких требований) в состоянии равновесия и нашли выражения для среднего времени ожидания и среднего числа требований, находящихся в очереди, для случая распределения времени обслуживания по эрланговскому закону. Даунтон рассмотрел предельный случай неограниченного увеличения объема группы. Совсем недавно Миллер-младший [568] подошел к вопросам группового поступления требований и группового обслуживания с точки зрения теории восстановления и смог добиться результатов, которые нельзя было получить с помощью вложенных цепей Маркова. Он рассмотрел более реальный случай, в котором возможны периоды, когда не происходит обслуживания. Этот автор также получил среднее время ожидания для периодов занятости.

Мейслинг [561], рассматривая дискретный аналог этой задачи и используя вложенные цепи Маркова, нашел среднее число требований, находящихся в очереди, и среднее время ожидания для стационарного состояния одноканальной системы с биномиальным входящим потоком и произвольным временем обслуживания при обслуживании требований в порядке поступления, а также рассмотрел специальные случаи, когда время обслуживания имеет геометрическое, а также вырожденное распределение. Перейдя к пределу, он получил известные результаты. Этот способ использовал также и Полячек.

Р. Джексон и Николс рассмотрели случаи равновесного состояния одноканальной системы при обслуживании требований в порядке поступления для входящего потока с распределением Эрланга для промежутков времени между моментами поступления требований и экспоненциального распределения времени обслуживания и нашли распределение вероятностей стационарного состояния и распределение времени ожидания тех требований, которые должны ожидать, а затем применили эти результаты для случаев экспоненциального и вырожденного распределений входящего потока.

В своей книге, опубликованной в 1957 г. [679], Полячек глубоко исследовал одноканальную систему с различными входящими потоками, различным временем обслуживания и различными дисциплинами очереди с приложением преимущественно к задачам, связанным с организацией воздушного сообщения.

Такач [800] и позже Бенеш [44] вывели и исследовали интегродифференциальное уравнение для распределения времени ожидания при переходном состоянии системы с обслуживанием требований в порядке поступления, с пуассоновским нестационарным входящим потоком и с произвольным распределением времени обслуживания. Аналогичный подход предпринял недавно и Декан. Решения уравнений для стационарного состояния получены им в виде преобразований (например, Лапласа). Смит [771] показал, что если распределение времени обслуживания есть функция из класса K_n, определяемого ниже, то распределение времени ожидания также принадлежит K_n при весьма общем характере входящего потока. Б. В. Гнеденко исследовал время ожидания в системах с нетерпеливыми клиентами и в системах с прерыванием обслуживания. Кифер и Вольфовиц [442] показали для случая многоканальной системы с произвольным входящим потоком и произвольным распределением времени обслуживания при обслуживании в порядке поступления, что распределение времени ожидания для стационарного состояния не существует, коль скоро загрузка системы ρ не меньше единицы. В одном частном случае оно может существовать также при $\rho=1$. Получено также общее выражение для распределения времени ожидания. Результат, полученный Линдли, является частным случаем данного результата, когда рассматривается один канал.

О. А. Вольберг [968], продолжая работу Полячека, получил распределение времени ожидания для переходного и стационарного состояний многоканальной системы с пуассоновским входящим потоком и произвольным временем обслуживания, когда каждое требование последовательно направляется в различные каналы, а каналы для обслуживания требований, поступающих в дальнейшем, снова назначаются в последовательном порядке.

Уишарт исследовал время ожидания для систем $GI/E_k/1$ и $D/E_k/1$ (см. ниже классификацию Кендалла).

Существует много других интересных ситуаций массового обслуживания, которые рассматриваются далее в нашей книге, но не включены в этот обзор.

Кратко остановимся на классификации систем массового обслуживания, предложенной Кендаллом. Если M обозначает, что распределение числа событий в фиксированном промежутке времени является пуассоновским, а распределение промежутков между ними — экспоненциальным, то $M/M/c$ означает, что в такой системе имеется c каналов. Символ D используется для обозначения вырожденного распределения, K_n — для распределения по закону хи-квадрат с четным числом степеней свободы, M_s — для группового обслуживания, E_n — для эрланговского распределения с математическим ожиданием, не зависящим от n, GI — символ последовательности независимых, произвольным образом (но одинаково) распределенных случайных величин (например, промежутки времени между моментами поступления требований независимы и имеют произвольное распределение, одно и то же для всех промежутков), а G обозначает произвольное распределение, в котором не требуется наличия независимости. Таким образом, $D/K_n/1$ обозначает одноканальную систему с регулярным входящим потоком и распределением времени обслуживания по закону Эрланга.

Задачи

1. Оцените задержки вследствие ожидания в системах обслуживания, с которыми вы встречаетесь в повседневной деятельности и определите, какими они могут оказаться за 40 лет. С материальной точки зрения подсчитайте потери творческой и иной деятельности, которой вы могли бы заняться в это время. Очевидно, что произвести такую оценку нелегко, но все же стоит попытаться.

2. Дайте пример работы системы массового обслуживания (как это было сделано для аэропорта), который включал бы большинство приведенных понятий. Придумайте по крайней мере две новых особенности функционирования системы массового обслуживания, не упоминавшиеся в этой главе.

3. Используйте последовательность случайных чисел (объем выборок равен 100) для моделирования системы массового обслуживания так, как было показано в тексте, и вычислите различные величины, о которых там упоминалось.

4. Каждому известно, к какому беспорядку приводят очереди, образующиеся перед многими параллельно работающими кассовыми аппаратами в магазинах самообслуживания в дни оживленной торговли. Очереди перед каждой кассой тянутся до проходов между продуктовыми прилавками, мешая движению. Большое пространство между этими очередями не используется. Из-за медлительности продавца можно долго простоять и в короткой очереди. Составьте схему обслуживания без потери времени и места по принципу «первым пришел — первым обслужен» всех клиентов, образующих общую очередь к кассам магазина. Предложите вашу схему ближайшему к вам магазину.

Глава 2

НЕКОТОРЫЕ МОДЕЛИ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

2.1. Введение

В этой главе будут довольно детально рассмотрены основные подходы к решению задач массового обслуживания. Цель главы — помочь читателю связать описательный материал первой главы с более аналитическими аспектами. Однако мы не ограничиваемся одними формулами. Для того чтобы читатель в самом начале получил более полное представление об основном математическом аппарате, необходимом для анализа реальных задач массового обслуживания, будут приведены также некоторые статистические задачи.

Глава делится, в основном, на три части. Первая часть посвящена построению моделей и нахождению решений. Вначале рассматривается детерминированная задача, на которую в прошлом не обращали достаточного внимания, несмотря на то, что вследствие ее простоты она помогает легко уяснить многие понятия теории массового обслуживания. Здесь будут рассмотрены только некоторые из них, остальные же войдут в упражнения.

За этой задачей следует задача промежуточного типа, которая представляет собой начальный переход от детерминированной задачи массового обслуживания к вероятностной. (Теория вероятностей и есть тот фундамент, на котором базируется развитие теории массового обслуживания.) Затем рассматриваются пуассоновские процессы. Они находят много важных приложений в теории массового обслуживания и поэтому должны быть изучены. Вторая модель системы массового обслуживания, основанная на свойствах пуассоновского процесса и связанного с ним экспоненциального распределения, построена для случая, когда имеется временна́я зависимость. Получены различные свойства пуассоновских процессов; на основании этих свойств выводятся некоторые свойства процессов массового обслуживания.

Аналогичный анализ применяется для исследования стационарного, т. е. независимого от времени, состояния одноканальной системы с пуассоновским входящим потоком и экспоненциальным временем обслуживания при обслуживании требований в порядке поступления. Затем дается вывод известной формулы Полячека—Хинчина для стационарного состояния (решение для переходного режима не рассматривается). Исследуя результаты, полученные для стационарного состояния, можно получить важные формулы, с помощью которых проверяется чувствительность системы к изменению параметров.

Вторая часть главы связана с проверкой гипотезы относительно наблюденного распределения и с оценкой параметров, полученных из наблюденных значений. Кратко обсуждаются проблемы, связанные с принятием гипотезы.

В последней части главы рассматривается решение одной простой задачи массового обслуживания методом Монте-Карло. Затем следуют численные методы решения некоторых уравнений для случаев, когда трудно получить решение в замкнутом виде.

2.2. Детерминированная система

Начнем с рассмотрения простой задачи массового обслуживания. В этой задаче мы принимаем, что поступление требований в систему происходит в фиксированные моменты времени, а время обслуживания постоянно. В этом случае оказывается возможным проследить за всем процессом обслуживания. Нашей задачей будет определение числа ожидающих требований и общего времени ожидания.

Допустим, что процесс массового обслуживания организован так, что поступление требований и обслуживание их происходит через равные промежутки времени. Задача является детерминированной и может быть решена элементарным способом.

Если требования поступают в одноканальную систему через равные промежутки времени a (т. е. интенсивность поступления требований равна $\frac{1}{a}$) и обслуживаются через равные промежутки времени b $\left(\text{т. е. интенсивность обслуживания равна } \frac{1}{b}\right)$, то при $b<a$ или, что то же самое, при $\frac{b}{a}<1$ ни одно требование ожидать обслуживания не будет. Если же $\frac{b}{a}>1$, то число ожидающих требований будет возрастать неограниченно. При $b=a$ требования ожидать не будут, если процесс обслуживания начался при отсутствии очереди. В противном случае будет очередь постоянной длины.

Допустим, что $\frac{b}{a}<1$ и к началу обслуживания в очереди имеется i требований (заметим, что $i\geqslant 2$, поскольку при $i=1$ это требование будет обслужено прежде, чем поступит новое, и очереди не будет). Все i требований будут обслужены к концу временно́го интервала, равного ib. За это же время поступит еще

$\left[\frac{ib}{a}\right]$ требований, которые будут ожидать. Квадратные скобки означают, что берется наибольшее целое число, не превосходящее $\frac{ib}{a}$ (требования не могут поступать в промежуточные моменты времени). Требование, поступившее последним, будет ожидать, пока обслуживающее устройство освободится.

Время обслуживания этих $\left[\frac{ib}{a}\right]$ требований равно $b\left[\frac{ib}{a}\right]$, а за это время поступит еще $\left[\frac{b\left[\frac{ib}{a}\right]}{a}\right]+1$ требований. Заметим, что это число меньше $\left[\frac{ib}{a}\right]$, так как $b<a$. Это происходит до того момента, когда поступающие требования уже не будут ожидать.

Поясним это на примере. Примем время обслуживания b равным одной временно́й единице, а длительность интервала между требованиями a равной трем временны́м единицам. Допустим, что $i=50$. Тогда время обслуживания i ожидающих требований равно 50 временны́м единицам. За это время поступит $\left[\frac{50}{3}\right]=16$ требований, и для их обслуживания потребуется еще 16 временны́х единиц. За это время поступит еще 6 требований, а за время обслуживания их — еще два. В момент окончания обслуживания последнего из этих двух требований поступит еще одно, но оно уже не будет ожидать. Таким образом, общее число требований, ожидающих обслуживания, будет равно 74. Сюда же входит и первое из начального числа i требований, поступившее на обслуживание сразу же после начала процесса.

Если обозначить через $P_n(t)$ вероятность того, что в момент времени t в очереди находится ровно n требований, то

$$P_n(0)=\begin{cases} 1 & \text{при } n=i, \\ 0 & \text{при } n\neq i. \end{cases}$$

Эти два возможных значения можно выразить одним символом $P_n(0)=\delta_{in}$, который называется символом Кронекера. Его значение равно 1 при $n=i$ и равно нулю при $n\neq i$. В данном случае число требований можно определить однозначно. Все величины определяются точно, следовательно, для любого момента времени вероятность $P_n(t)$ равна или нулю, или единице. Это не будет иметь места в случае недетерминированной системы, рассматриваемой в следующем параграфе.

Очевидно, что через промежуток времени, в течение которого все ожидающие требования будут обслужены, очередь исчезнет. Пусть в очереди ожидает A вновь поступивших требований. Тогда $(A+1)$-е требование поступит на обслуживание через промежуток времени, необходимый для обслуживания $i+A$ требований, т. е.

$(i+A)b \leqslant a(A+1)$ или $ib-A \leqslant A(a-b)$. Но $A=\left[\frac{ib-a}{a-b}\right]+1$, поэтому время, необходимое для обслуживания всех ожидающих требований, равно

$$b\left(i+\left[\frac{ib-a}{a-b}\right]+1\right)=b\left[\frac{ia-b}{a-b}\right].$$

И в этом случае сюда входит первое из i начальных требований.

Заметим, что чем ближе b к a, тем большее число требований будет ожидать, пока через продолжительный промежуток времени очередь не исчезнет.

Рассматривая несколько измененный вариант, можно получить более важные и легко вычисляемые выражения. При последующем рассмотрении примем, что a делится на b, а t делится на a без остатка.

Предположим, что процесс начинается с того, что одно из первоначально ожидавших требований поступает на обслуживание. Если обслуживание его окончено и новые требования не поступают, то на обслуживание допускается еще одно. Если же во время обслуживания этого или последующего начального требования поступает новое, то оно обслуживается сразу же, как только освобождается обслуживающее устройство. Возможно, что во время обслуживания этого требования поступит еще одно. Это зависит от величины разности $a-b$. Поступающее требование снова направляется на обслуживание, как только система освободится, и все остальные вновь поступающие требования будут обслуживаться в таком же порядке. Если же новые требования не поступают, то обслуживается очередное из i начальных требований. Вновь поступающие требования обслуживаются сразу же, как только система освобождается. Так будет до тех пор, пока не будут обслужены все i требований. Как только на обслуживание поступает последнее из i начальных требований, очередь исчезает. Если же во время обслуживания этого требования поступит новое, то оно ожидает, а затем поступает на обслуживание. Так происходит до тех пор, пока поступающим требованиям уже не нужно будет ожидать.

Мы рассмотрели процесс обслуживания в таком виде для того, чтобы лучше уяснить последующий анализ. Рассмотрим увеличение времени обслуживания, которое можно получить за счет того, что $a>b$. Так, для каждого поступающего требования имеется резерв $a-b$ временных единиц. Так как время обслуживания равно b, то теоретически к каждому из $i-1$ начальных требований нужно добавить $\frac{b}{a-b}$ требований, чтобы полностью компенсировать время незанятости системы. Заметим, что первое требование уже поступило в систему. Таким образом, всего должно поступить $(i-1)\frac{b}{a-b}$ требований.

Следовательно, общее число требований, ожидающих в очереди после начала процесса обслуживания, равно

$$i-1+\frac{(i-1)b}{a-b}=\frac{(i-1)a}{a-b}, \tag{2.1}$$

а если сюда включить еще и первое начальное требование, то получим

$$\frac{(i-1)a}{a-b}+1=\frac{ia-b}{a-b}. \tag{2.2}$$

Отсюда найдем общее время, прошедшее до момента первого освобождения системы обслуживания:

$$T\equiv\left(\frac{ia-b}{a-b}\right)b. \tag{2.3}$$

Так, если требование поступает в момент времени $t<T$, то до него уже поступило $\frac{t}{a}$ требований, и общее число тех требований, которые ожидали в очереди, равно $\frac{t}{a}+i-1$. За время t будет обслужено $\frac{t}{b}$ требований. (Заметим, что, не теряя общности, можно употреблять обозначение t как для непрерывного времени, так и в том случае, если время поступления требования t кратно a.) Следовательно, к моменту времени t впереди данного требования в очереди будет находиться

$$\frac{t}{a}+(i-1)-\frac{t}{b}=t\left(\frac{1}{a}-\frac{1}{b}\right)+(i-1) \tag{2.4}$$

требований. (Заметим, что если $b>a$, то число требований будет увеличиваться при возрастании t.) В системе будет $t\left(\frac{1}{a}-\frac{1}{b}\right)+i$ требований впереди данного.

Для требования, поступающего в момент времени $t+a$, время ожидания равно

$$W(t)=\begin{cases}0, & T\leqslant t,\\ \left(t\,\frac{b-a}{ab}+i\right)b, & 0<t\leqslant T,\\ (k-1)\,b, & t=0,\end{cases} \tag{2.5}$$

где $W(0)$ — время пребывания в очереди k-го из i начальных требований.

Это решение для переходного состояния, оно зависит от времени и начального числа требований. Стационарное состояние достигается после момента времени T, так как в дальнейшем состояние системы уже не зависит от времени и начального числа требований i.

Для требования, поступающего в момент времени $t+a$, общее время пребывания в системе равно

$$\begin{array}{ll} b, & T \leqslant t, \\ \left(\frac{t}{a}+i-\frac{t}{b}\right)b, & 0<t\leqslant T, \\ kb, & t=0. \end{array} \qquad (2.6)$$

Сюда же входит и время, затраченное на обслуживание.

Допустим теперь, что в течение временно́го интервала T произошло изменение длительности обслуживания и длительности промежутков времени между моментами поступления требований и их новые значения стали равны соответственно a_1 и b_1. Пусть это изменение происходит в момент времени $t_0<T$. Обслуживание требования, уже находящегося в обслуживающем устройстве, будет произведено за время b. Затем требования, находящиеся в очереди, и обслуживаемое требование образуют начальное состояние новой очереди. Это приводит к новому значению T, которое обозначим через T_1, и к изменению выражения для времени ожидания. Значение T_1 будет зависеть от выбора a_1 и b_1, следовательно, можно получить $T_1<T$. Так, $T_1=T$ в том и только в том случае, если $a_1=a$ и $b_1=b$, что можно показать, проанализировав формулу (2.3).

Если имеется c каналов с одинаковым временем обслуживания и начальное число требований $i>c$, то, приняв $\frac{b}{ca}<1$, найдем, что потребуется приращение, равное $\frac{a-b}{c}$, и т. д. За клиентами, прибывающими в различное время, можно закрепить определенные приоритеты и исследовать эту задачу как для системы с прерыванием обслуживания (т. е. когда требование с низшим приоритетом возвращается в очередь, если поступает требование с высшим приоритетом), так и для системы без прерывания обслуживания. Хотя рассмотренная задача массового обслуживания довольно проста, с ее помощью можно объяснить многие понятия. Однако и эта простая задача может оказаться весьма сложной, в чем легко убедиться, исследовав систему с двумя приоритетами, имеющими различные интервалы между требованиями и различное время обслуживания.

Упражнение 1. Какое влияние окажут вновь поступающие требования на длину очереди и время ожидания, если первое требование поступает в момент времени $a-\varepsilon$, второе — в момент времени $2a+\varepsilon$, третье — в момент времени $3a-\varepsilon$ и т. д., при произвольном чередовании запаздываний и опережений, где $0<\varepsilon<a$?

Упражнение 2. Пусть $i=7$, $b=2$ и $a=4$. Найдите время пребывания в системе $W(t)$ и время ожидания требования, выбираемого из очереди случайным образом.

Упражнение 3. Найдите выражения для времени ожидания в очереди и времени пребывания в системе с двумя приоритетами без прерывания обслуживания с промежутками времени между моментами поступления требований, равными a_1 и a_2, соответственно для требований с высшим и низшим приоритетами, и одинаковым временем обслуживания, равным b, при $\frac{b}{a_1}+\frac{b}{a_2}<1$ и $a_1<a_2$; в момент времени $t=0$ в системе находится i_1 требований с высшим приоритетом и i_2 требований с низшим приоритетом.

2.3. Вероятностная модель

Обычно предварительная информация относительно моментов поступления требований и длительности их обслуживания отсутствует. Чтобы провести анализ состояния системы массового обслуживания, делаются допущения, основанные на изучении особенностей и потребностей клиентов. Такую задачу можно решить, используя аппарат теории вероятностей.

В случае детерминированной системы было заранее известно, что требования будут поступать в точно заданные моменты времени и будут обслуживаться в течение определенного промежутка времени. Рассмотрим время обслуживания требований. В общем случае неизвестно, сколько времени будет продолжаться обслуживание поступившего требования. Однако на основании его свойств, наблюдавшихся ранее, можно определить, как часто потребуется обслуживание данной длительности, т. е. определить отношение промежутка времени, в течение которого обслуживание имеет данную длительность, к общему времени обслуживания. Иначе говоря, существует вероятность того, что обслуживание будет продолжаться до заданного момента времени. Эта вероятность равна сумме всех относительных частот обслуживания с длительностью меньшей заданной. Например, если время обслуживания является дискретным и принимает значения, равные 1, 2, 3,... временны́м единицам, с вероятностями, равными соответственно $p_1, p_2, \ldots$, то, обозначив через P_n вероятность того, что на обслуживание будет затрачено время не большее n, получим

$$P_n=p_1+\ldots+p_n, \tag{2.7}$$

и, следовательно, $p_n=P_n-P_{n-1}$.

Важными характеристиками, вычисляемыми с помощью этих вероятностей, являются математическое ожидание и дисперсия. Математическое ожидание находим, умножая каждое значение n на вероятность его появления и суммируя по всем n. Так, например, при бросании правильно выполненной игральной кости вероятность появления каждой грани равна $\frac{1}{6}$, а математическое ожидание числа выпавших очков равно

$$1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=3{,}5.$$

Следовательно, при большом числе бросаний среднее значение числа очков будет стремиться к 3,5. Для системы массового обслуживания при аналогичных вычислениях получим ожидаемую длину очереди. Кроме того, можно определить дисперсию, как показатель изменчивости длины очереди. В случае игральной кости при определении рассеяния вначале вычисляется дисперсия. Квадратный корень из дисперсии есть среднее квадратическое отклонение от среднего значения, которое является мерой рассеяния. Можно вычислить вероятность того, что при большом числе бросаний полученное среднее значение находится в пределах среднего квадратического отклонения относительно 3,5. Величина дисперсии находится из выражения

$$(1-3{,}5)\frac{1}{6}+(2-3{,}5)\frac{1}{6}+(3-3{,}5)\frac{1}{6}+(4-3{,}5)\frac{1}{6}+$$
$$+(5-3{,}5)\frac{1}{6}+(6-3{,}5)\frac{1}{6}\approx 2{,}92.$$

Среднее квадратическое отклонение составляет $\sqrt{2{,}92}$. Приращения возводятся в квадрат, поскольку представляет интерес абсолютная величина отклонения относительно среднего значения независимо от того, в какую сторону происходит отклонение. Вес каждого из этих отклонений определяется вероятностью появления рассматриваемого значения.

Математическое ожидание представляет собой первый начальный момент, а дисперсия — второй центральный момент. Можно вычислить и другие начальные и центральные моменты. Например, третий начальный момент равен сумме кубов частных значений, умноженных на вероятности их появления. Возвращаясь к p_n, укажем, что всегда должно выполняться условие $\sum_{n=0}^{\infty} p_n = 1$, так как это сумма вероятностей появления любой длительности обслуживания.

При непрерывном времени обслуживания вероятность того, что последнее не превзойдет x, обозначается через $F(x)$, т. е. $F(x) = P(X \leqslant x)$, где X — случайная величина, которая может принимать любое возможное значение длительности обслуживания. Если $f(x)$ — соответствующая плотность распределения (т. е. $f(x)dx$ — вероятность появления любого значения X, лежащего в пределах от x до $x+dx$), а x — непрерывная величина, то по аналогии со случаем дискретного времени обслуживания функция распределения имеет вид

$$F(x) = \int_0^x f(y)\,dy. \tag{2.8}$$

Если $F(x)$ — дифференцируемая функция, то можно записать соотношение

$$f(x) = \frac{dF(x)}{dx}. \tag{2.9}$$

Если имеется n неоднородных требований, то будем иметь несколько случайных величин $X_1, \ldots, X_n$ с соответствующими функциями распределения $F_1(x_1), \ldots, F_n(x_n)$.

Если эти случайные величины независимы (более точный смысл этого понятия будет указан позже), то совместную функцию распределения можно записать в следующем виде:

$$F(x_1, \ldots, x_n) = F_1(x_1) F_2(x_2) \ldots F_n(x_n) = \prod_{i=1}^{n} F_i(x_i). \qquad (2.10)$$

Коль скоро совместная функция распределения известна, можно определить и все остальные характеристики, которые описываются этим распределением.

В различные моменты времени требования могут иметь различные свойства. Так, можно наблюдать тенденцию увеличения длительности обслуживания в вечернее время (например, в ресторане) по сравнению с длительностью обслуживания в дневные часы. В любой момент времени вероятность того, что обслуживание имеет данную длительность, описывается законом распределения, который в данном случае является функцией времени. Длительность обслуживания описывается случайными величинами X_t, также зависящими от времени. Для момента времени t функция распределения имеет вид

$$P(X_t \leqslant x) = F(x;\ t).$$

Для каждого значения t существует определенная функция распределения X_t. Для значений t из некоторого интервала существует семейство случайных величин X_t. Как уже указывалось ранее, семейство случайных величин, зависящих от параметра, определяет вероятностный процесс. Если нет оснований выделять различные требования, поступающие в одно и то же время, по их свойствам, то моменты поступления требований могут описываться одним и тем же вероятностным процессом. В противном случае для каждого требования существует свой вероятностный процесс $X(k, t)$, который описывает признак требования (k-я заявка на обслуживание) и время его поступления в систему.

Если X — случайная величина, которая описывает различные метеорологические условия через вероятности появления каждого из них для различных моментов времени и в различных местах, то данная случайная величина зависит от четырех параметров: времени и трех пространственных координат. В теории массового обслуживания обычно рассматривается один параметр — время. Примером вероятностного процесса является пуассоновский процесс, который будет рассмотрен в следующем параграфе. Он описывается следующим выражением:

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \qquad (2.11)$$

где для случая, рассматриваемого в § 2.4, $P_n(t)$ — вероятность того, что за промежуток времени t поступит n требований.

Мы рассмотрим некоторые свойства этого полезного процесса, введя в последующих частях книги различные стоимостные подходы.

2.4. Пуассоновский процесс

Рассматриваемая здесь модель будет связана с распределением входящего потока, на который накладываются определенные условия. При рассмотрении модели исследуются свойства пуассоновского процесса. В соответствии с этими свойствами делаются определенные допущения, после чего пуассоновский процесс представляется как распределение входящего потока.

Из выражения $\frac{(\lambda t)^n e^{-\lambda t}}{n!}$ следует, что вероятность отсутствия требования в течение времени t равна $e^{-\lambda t}$, а вероятность поступления одного требования за время t равна $\lambda t e^{-\lambda t}$; вероятность поступления за время t более одного требования определяется выражением

$$1-(e^{-\lambda t}+\lambda t e^{-\lambda t})=1-\left\{\left[1-\lambda t+\frac{(\lambda t)^2}{2!}-\ldots\right]+\right.$$
$$\left.+\lambda t\left[1-\lambda t+\frac{(\lambda t)^2}{2!}-\ldots\right]\right\}=\frac{\lambda t^2}{2}+\ldots=O(t^2), \qquad (2.12)$$

т. е. функцией, которая ведет себя как t^2.

Если t мало, то члены с t^2 пренебрежимо малы по сравнению с членами без t или содержащими t в первой степени. Отсюда следует, что при малых t вероятность поступления более одного требования пренебрежимо мала. Это свойство весьма удовлетворяет многие практические приложения; по этой причине процесс Пуассона получил широкое распространение. Вероятность поступления за время t по крайней мере одного требования определяется как

$$1-e^{-\lambda t}=\lambda t+O(t^2).$$

Вероятность отсутствия требований равна $e^{-\lambda t}=1-\lambda t+O(t^2)$.

Если рассматриваются только малые значения t, то в обоих выражениях можно пренебречь последней величиной в правой части. В самом деле перечисленные свойства не являются независимыми; однако в соответствии с установившейся методикой, мы допустим существование всех этих свойств, т. е. предположим, что вероятность поступления требования в течение малого временного интервала Δt равна $\lambda\Delta t$, а вероятность поступления более одного требования за время Δt ничтожно мала. После этого можно вывести пуассоновское распределение, которому присущи эти свойства.

Пусть $P_n(t)$ — вероятность того, что за время t поступит n требований. Заметим, что $0 \leqslant P_n(t) \leqslant 1$. Кроме того, $\sum_{n=0}^{\infty} P_n(t) = 1$, поскольку за время t или не поступит ни одного требования или же поступит какое-то количество их. Чтобы показать, что происходит в течение следующего малого промежутка времени Δt, запишем

$$\begin{aligned} &P_0(t+\Delta t) = P_0(t)(1-\lambda\Delta t), \\ &P_n(t+\Delta t) = P_n(t)(1-\lambda\Delta t) + P_{n-1}(t)\lambda\Delta t, \quad n \geqslant 1. \end{aligned} \tag{2.13}$$

Первое уравнение позволяет определить вероятность того, что за время $t+\Delta t$ не поступит ни одного требования. Эту вероятность можно связать с состоянием системы в момент времени t. По закону умножения вероятностей двух независимых событий значение $P_0(t+\Delta t)$ равно вероятности отсутствия требований за время t, умноженной на вероятность отсутствия требований за время Δt. Для случаев $n \geqslant 1$ это свойство (т. е. наличие этого же числа требований к моменту времени t и отсутствие требований за время Δt) также справедливо, но, кроме этого, за время t может поступить $n-1$ требований, а за время Δt дополнительно поступить еще одно требование. Произведение этих величин равно второму члену правой части. Возможность поступления более одного требования в течение малого временно́го интервала Δt не рассматривается, поскольку она пренебрежимо мало вероятна; как легко видеть, в последующих выкладках соответствующие слагаемые можно отбросить.

Произведем умножение, перенесем $P_n(t)$ в левую часть и, разделив на Δt, получим систему уравнений

$$\frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n \geqslant 1.$$

При $\Delta t \to 0$ правая часть, по определению, есть производная $P'_n(t) = \frac{dP(t)}{dt}$, и система уравнений принимает вид

$$\begin{aligned} &P'_0(t) = -\lambda P_0(t), \\ &P'_n(t) = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n \geqslant 1. \end{aligned} \tag{2.14}$$

Это линейные дифференциальные уравнения относительно t и линейные разностные уравнения первого порядка относительно n; обычно их называют дифференциально-разностными уравнениями.

Решение этих уравнений удобно находить, используя производящую функцию. Последняя определяется как

$$P(z, t) \equiv \sum_{n=0}^{\infty} P_n(t) z^n = P_0(t) + P_1(t) z + P_2(t) z^2 + \ldots \tag{2.15}$$

Можно показать, что $P_n(t)$ получим, продифференцировав $P(z, t)$ n раз по z, разделив результат на $n!$ и положив $z=0$. Поэтому если $P(z, t)$ известно, то нетрудно найти и $P_n(t)$.

При решении уравнений начало отсчета времени можно выбрать произвольно, даже после того, как поступит некоторое число требований. Так, возможно, что к моменту времени $t=0$ уже поступило i требований. В этом случае $P_n(0)=0$, если $n \neq i$ и $P_n(0)=1$, если $n=1$. Таким образом,

$$P(z, 0)=\sum_{n=0}^{\infty} P_n(0) z^n = P_i(0) z^i = z^i. \tag{2.16}$$

Заметим также, что $P(1, t)=1$ и

$$\frac{\partial P(z, t)}{\partial t}=\frac{\partial}{\partial t}\sum_{n=0}^{\infty} P_n(t) z^n=\sum_{n=0}^{\infty} P'_n(t) z^n. \tag{2.17}$$

Если умножить систему дифференциально-разностных уравнений для $n \geqslant 1$ на z^n, а первое уравнение на z^0 и просуммировать по n, то получим, что сумма левых частей равна $\frac{\partial P(z, t)}{\partial t}$, а сумма первых выражений правых частей равна $-\lambda P(z, t)$. (Для читателя полезно подставить несколько значений n и z, просуммировать и посмотреть полученный результат.)

Суммируя вторые члены правой части по n, получим выражение

$$\sum_{n=1}^{\infty} \lambda P_{n-1}(t) z^n = \lambda P_0(t) z + \lambda P_1(t) z^2 + \ldots . \tag{2.18}$$

Если из членов правой части вынести множитель λz, то их все можно записать в виде $\lambda z P(z, t)$. Таким образом, система приводится к линейному дифференциальному уравнению для производящей функции, которое имеет вид

$$\frac{\partial P(z, t)}{\partial t} - \lambda(z-1) P(z, t) = 0. \tag{2.19}$$

Это одно из простейших обыкновенных дифференциальных уравнений, рассматриваемых в общих курсах. Решение его при значении z, принятом постоянным (поскольку оно не зависит от t), имеет вид

$$P(z, t) = C e^{\lambda(z-1)t}. \tag{2.20}$$

Результат можно проверить путем подстановки, причем C может зависеть от z.

Допустим, что к моменту времени $t=0$ не поступило ни одного требования, тогда $P(z, 0)=1$, так как $i=0$. Таким образом, $C=1$ и

$$P(z, t) = e^{\lambda(z-1)t}. \tag{2.21}$$

Как уже указывалось выше, $P_n(t)$ можно получить путем дифференцирования производящей функции, следовательно,

$$P_n(t) = \frac{1}{n!} \frac{\partial^n P(z, t)}{\partial z^n} \bigg|_{z=0}. \tag{2.22}$$

Таким образом,

$$P_0(t) = e^{-\lambda t}, \tag{2.23а}$$

$$P_1(t) = \lambda e^{-\lambda t} \tag{2.23б}$$

и

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \tag{2.23в}$$

что является искомым выражением для пуассоновского процесса. Теперь уже должно быть ясно, что t можно рассматривать не как абсолютное время, а как длительность временнóго интервала, на протяжении которого происходят события.

Для многих задач массового обслуживания такой способ получения решения является типичным. Уравнения могут быть более сложными, но методика в основном нередко остается такой же. Читатель может легко показать, что если в начальный момент времени имеется i требований $(i > 0)$ и по-прежнему понимать под $P_n(t)$ число требований, поступивших до момента t, то

$$P_n(t) = \frac{(\lambda t)^{n-i} e^{-\lambda t}}{(n-i)!}, \quad n \geqslant 1, \tag{2.24}$$

так как в этом случае $P(z, 0) = z^i$, то $C = z^i$ и

$$P(z, t) = z^i \sum_{n=i}^{\infty} P_n(t) z^{n-i}.$$

В других случаях $P_n(t)$, кроме вероятности поступления n требований за время t, может обозначать вероятность того, что в момент времени t в системе (в очереди и на обслуживании) находится n требований, или вероятность того, что занято n каналов, если число каналов равно c, и т. д. Хотя в таких случаях для получения различных показателей эффективности функционирования систем массового обслуживания важно найти величину $P_n(t)$, все же вычисление $P_n(t)$ не всегда необходимо, а иногда и невозможно. Как увидим далее, выражения для времени ожидания можно найти и другими способами.

Найдя $P_n(t)$, можно подсчитать среднее число требований, поступивших к моменту времени t,

$$L = \sum_{n=0}^{\infty} n P_n(t), \tag{2.25}$$

которое для пуассоновского процесса равно λt.

Дисперсия

$$\sum_{n=0}^{\infty} (n-L)^2 P_n(t)$$

также равна λt.

Упражнение 4. Проверьте значения математического ожидания и дисперсии для пуассоновского процесса.

Промежутки времени между требованиями для пуассоновского процесса с параметром λ имеют экспоненциальное распределение с плотностью $\lambda e^{-\lambda t}$. Чтобы доказать это, заметим: если только что поступило одно требование, то время, через которое поступит следующее требование, будет меньше t в том и только в том случае, если за этот промежуток времени поступит одно или большее число требований. Вероятность поступления одного или большего числа требований, а следовательно, и вероятность того, что промежуток времени между требованиями меньше или равен t, равна

$$\sum_{n=1}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} = 1 - e^{-\lambda t}. \qquad (2.26)$$

Правая часть данного выражения есть функция распределения, из которой путем дифференцирования находим экспоненциальную плотность $\lambda e^{-\lambda t}$ для распределения промежутков времени между требованиями. *Следовательно, если моменты поступления требований в систему распределены по закону Пуассона, то промежутки времени между требованиями имеют связанное с ним экспоненциальное распределение.* И наоборот, если, например, время обслуживания имеет экспоненциальное распределение, то поступление требований на обслуживание описывается как пуассоновский процесс.

Если $\mu \Delta t$ — вероятность того, что обслуживание закончится за время Δt (следовательно, $1 - \mu \Delta t$ — вероятность того, что за время Δt обслуживание не закончится), и если $P(t)$ — вероятность того, что обслуживание не закончится за время t, то

$$P(t+\Delta t) = (1-\mu \Delta t) P(t). \qquad (2.27)$$

Отсюда

$$\frac{dP(t)}{dt} = \lim_{\Delta t \to 0} \frac{P(t+\Delta t) - P(t)}{\Delta t} = -\mu P(t).$$

Это дифференциальное уравнение имеет решение

$$P(t) = c e^{-\mu t}. \qquad (2.28)$$

Так как рано или поздно обслуживание должно окончиться, то

$$\int_0^{\infty} c e^{-\mu t} dt = 1.$$

Следовательно, $c=\mu$. Таким образом, время обслуживания описывается с помощью экспоненциального распределения с плотностью вероятности, равной $\mu e^{-\mu t}$. Заметим, что экспоненциальное распределение обладает свойством «отсутствия памяти». Например, если время обслуживания распределено по экспоненциальному закону, то остаток незаконченного обслуживания имеет то же распределение, что и длительность только что начавшегося обслуживания.

Для многих задач переходный режим не является существенным. По истечении определенного промежутка времени система может войти в стационарный режим. Для многих систем можно ожидать, что это состояние появится при $t\to\infty$. Установление условий существования, подобных $p_n=\lim\limits_{t\to\infty}P_n(t)$, является основным предметом исследования эргодической теории; подобный анализ применяется в том случае, когда хотят найти характеристики установившегося режима. Заметим также, что условие отсутствия изменения вероятностей $P_n(t)$ с течением времени определяет стационарное распределение.

Изменение $P_n(t)$ в зависимости от t описывается с помощью производной $P'_n(t)$, а при стационарном состоянии $P'_n(t)=0$. Таким образом, по существу, имеется два способа получения вероятностей стационарного состояния: 1) из условия $P'_n(t)=0$, которое обеспечивает независимость p_n от t, и 2) из рассмотрения $\lim\limits_{t\to\infty}P_n(t)$, что также приводит к вероятности p_n, не зависящей от t. Используем теперь понятие стационарного состояния при решении двух задач массового обслуживания. Первая из них связана с моделью Эрланга, а вторая содержит вывод формулы Полячека—Хинчина.

2.5. Модель Эрланга. Формула Полячека—Хинчина

Модель Эрланга. Если допустить, что процесс начинается при отсутствии требований в очереди, то одноканальная система массового обслуживания с пуассоновским входящим потоком с параметром λ и экспоненциальным временем обслуживания с параметром μ при обслуживании по принципу «первым пришел — первым обслужен» описывается следующими уравнениями:

$$P_n(t+\Delta t)=P_n(t)\,[1-(\lambda+\mu)+P_{n-1}(t)\lambda\Delta t+P_{n+1}(t)\mu\Delta t,\quad n\geqslant 1,$$
$$P_0(t+\Delta t)=P_0(t)(1-\lambda\Delta t)+P_1(t)\mu\Delta t,\quad n=0.\qquad(2.29)$$

Из этих уравнений следует, что вероятность того, что в момент времени $t+\Delta t$ в системе находится n требований, равна вероятности того, что в момент времени t в системе находится n требований, умноженной на вероятность того, что за время Δt в систему не поступит ни одного требования и ни одно требование не будет обслужено; плюс вероятность того, что в момент времени t в системе находится $n-1$ требований, умноженная на вероятность того, что за время Δt поступит одно требование и ни одно требование не будет обслужено; плюс вероятность того, что в момент времени t в системе находится $n+1$ требований, умноженная на вероятность того, что за время Δt одно требование покинет систему и не поступит ни одного требования.

Заметим, что вероятность того, что за время Δt не поступит ни одного требования и ни одно требование не покинет систему, равна $(1-\lambda\Delta t)(1-\mu\Delta t)$. Член, содержащий $(\Delta t)^2$, при составлении дифференциального уравнения опускается. Следовательно, можно записать $1-(\lambda+\mu)\Delta t$. Относительно остальных двух членов первого уравнения заметим, что $\lambda\Delta t(1-\mu\Delta t)\approx\lambda\Delta t$ и $\mu\Delta t(1-\lambda\Delta t)\approx\mu\Delta t$.

Перенеся $P_n(t)$ влево и устремив Δt к нулю, получим

$$\frac{dP_n(t)}{dt}=-(\lambda+\mu)P_n(t)+\lambda P_{n-1}(t)+\mu P_{n+1}(t), \quad n\geqslant 1, \tag{2.30}$$

$$\frac{dP_0(t)}{dt}=-\lambda P_0(t)+\mu P_1(t), \quad n=0.$$

Как уже указывалось ранее, независимое от времени (или стационарное) решение получим, либо решая зависимые от времени уравнения для переходного распределения, а затем полагая $t\to\infty$ (пример будет приведен в следующей главе), либо приравняв нулю производные по времени и решая полученные уравнения для стационарного состояния. Переходные решения особенно важны в тех случаях, когда загрузка системы $\rho=\frac{\lambda}{\mu}\geqslant 1$, так как в этом случае стационарный режим не устанавливается. Найдем выражение для математического ожидания числа требований, находящихся в очереди, и среднего времени ожидания для стационарного состояния при $\frac{\lambda}{\mu}<1$. Если $\frac{\lambda}{\mu}\geqslant 1$, то число ожидающих требований будет возрастать бесконечно. Сравните эти условия с аналогичными условиями, рассмотренными ранее для случая детерминированной системы.

Приравняв нулю производные по времени и исключив, таким образом, t из уравнений (2.30), получим

$$(\lambda+\mu)p_n=\lambda p_{n-1}+\mu p_{n+1}, \quad n\geqslant 1,$$

$$\lambda p_0=\mu p_1, \quad n=0. \tag{2.31}$$

Введем $\rho=\frac{\lambda}{\mu}$; тогда эти уравнения примут вид

$$(1+\rho)p_n=p_{n+1}+\rho p_{n-1}, \quad n\geqslant 1,$$

$$p_1=\rho p_0, \quad n=0.$$

Пусть в первом уравнении $n=1$. Тогда $(1+\rho)p_1=p_2+\rho p_0$. Подставив значение p_1 из второго уравнения, найдем, что $p_2=\rho^2 p_0$. Повторяя процесс, получим $p_n=\rho^n p_0$. Здесь $\sum_{n=0}^{\infty} p_n=1$, так как это сумма вероятностей того, что в системе нет ни одного требования, находится одно требование, два требования и т. д. Сумма всех вероятностей должна быть равна единице, так как рассматриваются все возможные состояния системы. Поэтому

$$\sum_{n=0}^{\infty} \rho^n p_0 = 1$$

или

$$\sum_{n=0}^{\infty} p_0 \rho^n = p_0 \sum_{n=0}^{\infty} \rho^n = \frac{p_0}{1-\rho} = 1,$$

откуда

$$p_0 = 1 - \rho.$$

Следовательно,

$$p_n = \rho^n (1-\rho), \tag{2.32}$$

т. е. получим выражение, которое представляет собой распределение вероятностей, известное как геометрический закон.

Математическое ожидание числа требований, находящихся в системе, равно

$$L = \sum_{n=0}^{\infty} n p_n = (1-\rho) \sum_{n=0}^{\infty} n\rho^n = \frac{\rho}{1-\rho}, \tag{2.33}$$

что можно легко проверить. Заметим, что L — среднее значение, в действительности же возможно колебание числа требований, ожидающих обслуживания. Это лучше всего показать, вычислив дисперсию

$$\sum_{n=0}^{\infty} (n-L)^2 p_n = \sum_{n=0}^{\infty} n^2 p_n - \left(\frac{\rho}{1-\rho}\right)^2.$$

Но

$$\sum_{n=0}^{\infty} n^2 p_n = (1-\rho) \sum_{n=0}^{\infty} n^2 \rho^n = \frac{\rho}{1-\rho} + \frac{2\rho^2}{(1-\rho)^2}.$$

Следовательно, дисперсия равна

$$\frac{\rho}{1-\rho} + \frac{\rho^2}{(1-\rho)^2} = L + L^2. \tag{2.34}$$

Математическое ожидание числа требований, находящихся в очереди, равно

$$L_q = \sum_{n=1}^{\infty} (n-1) p_n = L - \rho = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}. \qquad (2.35)$$

Заметим, что существуют такие промежутки времени, когда канал свободен. Среднее время ожидания в очереди W_q равно $\frac{L_q}{\lambda}$ или $\frac{L}{\mu}$.

Удобным способом быстрой проверки возможной неправильности формулы является анализ размерности. Так, например, в последнем выражении правая и левая части — безразмерные величины.

Упражнение 5. Решите систему (2.31), используя производящую функцию, аналогичную (2.15) при опущенном t.

Формула Полячека—Хинчина. Предположим, что прибытие клиентов в одноканальную систему, находящуюся в стационарном состоянии, описывается как пуассоновский случайный процесс с параметром λ. Предположим также, что время обслуживания имеет произвольное распределение с интенсивностью обслуживания, равной μ клиентов в единицу времени. Клиенты обслуживаются в порядке поступления. Как и в случае детерминированной системы, будем считать, что $\frac{\lambda}{\mu} < 1$. Допустим, что в момент времени, когда обслуженный клиент покидает систему, в ней остается q клиентов, считая и того, который находится на обслуживании. Время обслуживания этого клиента равно t. Пусть за время t поступит еще r клиентов. Если после ухода следующего клиента в системе остается еще q' клиентов, то величины q и q' можно связать следующим образом:

$$q' = \max(q-1,\ 0) + r = q - 1 + \delta + r, \qquad (2.36)$$

где

$$\delta(q) = \begin{cases} 0 & \text{при} \quad q > 0, \\ 1 & \text{при} \quad q = 1. \end{cases} \qquad (2.37)$$

Заметим, что введя обозначение δ, мы избавились от необходимости употреблять символ максимума.

Допустим, что при равновесном состоянии системы существуют моменты первого и второго порядков $E[q]$ и $E[q^2]$, а q рассматривается как случайная величина. По определению, $\delta^2 = \delta$ и $q(1-\delta) = q$. Кроме того, $E[q] = E[q']$ и $E[q^2] = E[q'^2]$, так как при равновесном состоянии q и q' имеют одинаковое распределение. Поскольку система находится в равновесии, характер очереди остается неизменным для любого момента времени, в который требования покидают систему, т. е. существует одна и та же функция распределения, независимая от времени.

Подставляя в равенство (2.36) средние значения этих величин, получим

$$E[q'] = E[q] - E[1] + E[\delta] + E[r], \tag{2.38}$$

откуда

$$E[\delta] = 1 - E[r]. \tag{2.39}$$

Если длительность обслуживания равна t, то

$$E[r] = \sum_{r=0}^{\infty} r \frac{(\lambda t)^r}{r!} e^{-\lambda t} = \lambda t \tag{2.40}$$

и

$$E[r^2] = \sum_{r=0}^{\infty} r^2 \frac{(\lambda t)^r}{r!} e^{-\lambda t} = (\lambda t)^2 + \lambda t. \tag{2.41}$$

Рассматривая средние значения за время обслуживания t, получим $E[r] = \frac{\lambda}{\mu} \equiv \rho$. Если время обслуживания имеет, к примеру, экспоненциальное распределение, то

$$E[r] = \mu \int_0^{\infty} (\lambda t) e^{-\mu t} dt = \frac{\lambda}{\mu} \equiv \rho, \tag{2.42}$$

так как среднее значение времени обслуживания равно $\frac{1}{\mu}$.

Заметим, что $E[r]$ — число, которое не изменяется при усреднении. Отсюда находим, что $E[\delta] = 1 - \rho$. Вероятность поступления r требований не зависит от длины очереди q и от δ (предполагается, что δ зависит только от q — величины, не зависящей от r). Следовательно, если берутся средние значения величин r, q и δ, то можно получить отдельно средние значения r и q из имеющегося произведения этих величин. Это верно также для r и δ. Усредняя r^2 по времени, получим

$$E[r^2] = \lambda^2\sigma^2(t) + \rho^2 + \rho. \tag{2.43}$$

Возводя в квадрат правую и левую части уравнения (2.36), связывающего q и q', и принимая во внимание, что $\delta^2 = \delta$ и $\delta q \equiv 0$, находим

$$q'^2 = q - 2q(1 - r) + (r - 1)^2 + \delta(2r - 1). \tag{2.44}$$

Так как рассматривается равновесное состояние, то

$$0 = E[q'^2] - E[q^2] = 2E[q]E[r-1] +$$
$$+ E[(r-1)^2] + E[\delta]E[2r-1].$$

Упрощая выражение и используя ранее полученные зависимости, получим формулу Полячека—Хинчина:

$$E[q]=\frac{E[(r-1)^2]+E[\delta]\,E[2r-1]}{2E[1-r]}=$$

$$=\frac{E[r^2]-2E[r]+1+E[\delta]\,(2E[r]-1)}{2(1-E[r])}=$$

$$=\frac{\lambda^2\sigma^2(t)+\rho^2+\rho-2\rho+1+(1-\rho)(2\rho-1)}{2(1-\rho)}=$$

$$=\rho+\frac{\rho^2+\lambda^2\sigma^2(t)}{2(1-\rho)}. \qquad (2.45)$$

Поскольку для заданного закона распределения известна дисперсия времени обслуживания t, то можно определить среднее число требований, находящихся в системе. Необходимо заметить, что в данном случае среднее значение берется для моментов, непосредственно *следующих* за окончанием обслуживания требования, и не является средним по времени числом требований, находящихся в системе. Если $E_t[q]$ — среднее по времени число требований, то, как легко видеть (см. гл. 7), $E[q]<E_t[q]<E[q]+1$.

Заметим, что в любом случае (даже для системы с несколькими параллельными каналами) среднее число требований, находящихся в системе, равно сумме среднего числа занятых каналов (в данном случае это ρ — загрузка системы) и среднего числа ожидающих требований.

Для нахождения среднего времени ожидания рассуждаем следующим образом. Если среднее время ожидания в очереди (исключая время обслуживания) равно $E[w]$, то $\lambda\left(E[w]+\frac{1}{\mu}\right)$ — среднее число требований, которые поступят в систему за время ожидания и время обслуживания одного требования, т. е. за время его пребывания в системе. Но это должно быть то же самое число требований, которое будет в системе сразу же после того, как данное требование покинет систему, т. е. $E[q]$. Таким образом,

$$W_q\equiv E[w]=\frac{\rho^2+\lambda^2\sigma^2(t)}{2\lambda(1-\rho)}=\frac{L_q}{\lambda}. \qquad (2.46)$$

Важным показателем является отношение среднего времени ожидания требования к среднему времени обслуживания. Естественно, что не хочется ожидать слишком долго, если обслуживание будет произведено быстро. Выражение для этого показателя имеет вид

$$\mu E[w]=\frac{\rho}{2(1-\rho)}(1+C_t^2), \qquad (2.47)$$

где C_t — коэффициент вариации времени обслуживания, который определяется из следующего выражения:

$$C_t^2=\frac{\sigma^2(t)}{E^2[t]}=\frac{\sigma^2(t)}{\frac{1}{\mu^2}}. \qquad (2.48)$$

Ясно, что показатель $\mu E[w]$ зависит от ρ и от C_t^2. Любое изменение средней интенсивности входящего потока λ или интенсивности обслуживания μ влияет и на их отношение $\frac{\lambda}{\mu}=\rho$. Поэтому, чтобы определить чувствительность к изменению значений ρ относительно данного среднего значения $\bar{\rho}$, например, существующего в настоящий момент, разложим правую часть в ряд Тейлора для малых отклонений относительно $\bar{\rho}$. Имеем

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\ldots+$$
$$+\frac{f^n(a)}{n!}(x-a)^n+\ldots.$$

Здесь число штрихов определяет порядок производной.

Применяя это разложение к правой части формулы (2.47), получим

$$\mu E[w]=\frac{\bar{\rho}}{2(1-\bar{\rho})}\left(1+C_t^2\right)+\frac{1+C_t^2}{2!(1-\bar{\rho})^3}(\rho-\bar{\rho})^2+\ldots. \tag{2.49}$$

Если взять среднее значение относительно ρ, то второй член обращается в нуль, так как $E[\rho-\bar{\rho}]=\bar{\rho}-\bar{\rho}$. Взяв члены до второй степени включительно, получим

$$\frac{\bar{\rho}}{2(1-\bar{\rho})}(1+C_t^2)\left[1+\frac{\sigma_\rho^2}{\bar{\rho}(1-\bar{\rho})^2}\right], \tag{2.50}$$

где

$$\sigma_\rho^2=E[(\rho-\bar{\rho})]^2.$$

Для значений $\bar{\rho}=0{,}6$ и $\sigma_\rho=0{,}1$ Кокс [157] нашел, что $1+C_t^2=1{,}1$, следовательно, время ожидания увеличивается на 10%. Этот результат является важным, когда процесс обслуживания можно рассматривать как последовательность процессов, каждый из которых является стационарным и имеет свои параметры. Затем можно определить приемлемую ширину диапазона колебаний значений параметров времени обслуживания (перед переходом к новому стационарному состоянию) при допустимом изменении времени ожидания.

2.6. Проверка гипотезы

При накоплении данных для оценки распределения входящего потока или распределения времени обслуживания число поступающих требований и длительность обслуживания записываются как функции времени. Затем эти данные могут группироваться по временны́м интервалам. После этого можно проверить гипотезу о взаимной независимости средних значений в последовательных временны́х интервалах; для этого можно, например, использовать такой показатель, как отношение дисперсий, или другие показатели, которые приводятся в общих курсах математической статистики. Если же окажется, что существует зависимость от времени, то экспериментальные данные нужно распределить по таким временны́м интервалам, внутри которых стационарное состояние является приемлемой аппроксимацией.

После того как группировка данных или той их части, которая окажется однородной, будет выполнена, берется 10—20 подгрупп и строится гистограмма частот. По внешнему виду такой гистограммы можно судить, является ли данное распределение нормальным, экспоненциальным или же это распределение суммы взаимно независимых экспоненциально распределенных случайных величин.

Оценка параметров предполагаемого распределения производится на основании экспериментальных данных. Наблюденное распределение сравнивается со стандартным с помощью критерия хи-квадрат. Если обнаруживаются статистически значимые отклонения, то решение относительно того, нужно ли менять модель распределения, можно принять лишь исходя из реальных условий определенной задачи.

Иные соображения необходимо иметь в виду, когда экспериментальных данных недостаточно. Параметры, оцененные на основании этих данных, могут весьма заметно отличаться от параметров, соответствующих истинному распределению даже в том случае, если критерий хи-квадрат не показывает существенного отклонения от модели.

Последние два положения оказываются менее важными, если модель можно испытать на практике. Иногда необходима такая экспериментальная проверка модели. Закрыв, например, один из проездов на шоссе с многорядным движением для проверки модели, можно предсказать, какое влияние окажет открытие дополнительного проезда.

В исключительных случаях, когда экспериментальная проверка невозможна, необходима особая тщательность при определении соответствия имеющихся данных с моделью. Определить влияние изменения значений параметров на модель, вероятно, можно аналитически, изменяя параметры распределения входящего потока и параметры распределения времени обслуживания в диапазоне, соответствующем ожидаемому разбросу, и наблюдая влияние этих изменений на вычисленные значения, как было показано ранее при анализе чувствительности.

Допустим, что необходимо проверить гипотезу о том, что входящий поток имеет пуассоновское распределение со средней интенсивностью, равной 0,4 требований в минуту. Интервалы времени для подсчета числа поступающих требований приняты равными 2 мин. Подсчитывается общее число требований в каждом таком промежутке. Затем вычисляется частота появления интервалов, в которых имеет место данное число требований. Так находим два первых столбца табл. 2.1.

Таблица 2.1

Сравнение экспериментальных данных для входящего потока с пуассоновским распределением

Число требований n	Число интервалов, содержащих данное число требований n	Значение вероятности при пуассоновском распределении	Математическое ожидание числа интервалов, содержащих данное число требований n
1	2	3	4
0	21	0,449	26,9
1	23	0,359	21,5
2	10	0,144	8,6
3	4		
4	1		
5	0 } 6	0,048	2,9
6	0		
7	1		
	60		

Когда число во втором столбце мало, например меньше 5, то оно объединяется с другими, чтобы увеличить объем выборки. Так, интервалы, в которых появляется от 3 до 7 требований, рассматриваются совместно; число их равно шести. Используя выражение для пуассоновского распределения

$$P_n(t) = \frac{(0,4t)^n e^{-0,4t}}{n!}, \tag{2.51}$$

при $t=2$ вычисляем третий столбец для значений n, взятых из первого столбца. Получаем вероятность поступления заданного числа требований. Умножив каждую вероятность на 60, получим число интервалов, в которых ожидается поступление заданного числа требований. И, наконец, чтобы сделать вывод относительно того, будет ли принятый пуассоновский процесс с достаточной точностью описывать статистическое распределение, находим выражение для хи-квадрат:

$$\chi^2 = \sum \frac{(\text{наблюденное число} - \text{математическое ожидание})^2}{\text{математическое ожидание}}. \tag{2.52}$$

Здесь разность между числами из второго и последнего столбцов возводится в квадрат и делится на число из последнего столбца. Затем результат суммируется для всех соответствующих пар чисел. Получаем

$$\frac{(26,9-21)^2}{26,9} + \frac{(21,5-23)^2}{21,5} + \frac{(8,6-10)^2}{8,6} + \frac{(2,9-6)^2}{2,9} = 4,94.$$

Из таблицы для хи-квадрат с $k-1=3$ степенями свободы (рассматривается четыре группы чисел, а число степеней свободы равно числу групп, уменьшенному на единицу) находим, что это число меньше того, которое указано для наибольшего уровня значимости. Следовательно, можно сделать вывод о том, что если математическое ожидание числа требований, равное 0,4, является приемлемым, то экспериментальные данные незначительно отличаются от значений, полученных для «настоящего» пуассоновского распределения. Если же вычисленное значение превосходит табличное, то при данном уровне значимости кривая признается несоответствующей закону распределения.

2.7. Оценка параметров систем массового обслуживания

Для оценки параметров λ и μ, характеризующих входящий поток и время обслуживания в случае стационарного состояния одноканальной системы с пуассоновским входящим потоком и экспоненциальным временем обслуживания при обслуживании по принципу «первым пришел — первым обслужен», Кларк [125] использовал метод наибольшего правдоподобия. Обычно для оценки λ процесс наблюдается в течение времени s, и если за это время поступает n требований, то полагается $\hat{\lambda}=\frac{n}{s}$.

Для оценки μ, среднего числа требований, которое канал может обслужить за единицу времени, нужно брать в расчет только то время, когда имеются требования, ожидающие обслуживания или находящиеся на обслуживании. Это время называется периодом занятости канала. Если обслуживается m требований и общее время занятости канала равно τ, то $\hat{\mu}=\frac{m}{\tau}$. Оценка μ будет наилучшей при $\rho=\frac{\lambda}{\mu}\geqslant 1$. Если $\rho<1$, то при стационарном состоянии системы для улучшения этих оценок можно использовать наличие начальной очереди, содержащей i требований.

Если имеются все необходимые распределения рассматриваемых величин, то распределение выборок можно найти как произведение этих распределений (допуская, что выборки взаимно независимы); функция правдоподобия образуется как произведение распределений выборок. Затем для нахождения максимумов этой функции берется производная по ее параметрам.

Пусть $m>i$, где i — начальное число требований, находящихся в очереди. Допустим, что рассматривается время занятости канала, пока оно не достигнет заранее заданного значения τ. Пусть T — момент времени, в который поступает m-е требование. Случайные величины X_j, Y_j, Z_j описывают соответственно момент поступления j-го требования, момент ухода j-го требования из системы и время занятости канала до момента ухода j-го требования (все величины равны нулю при $j \leqslant 0$). Затем рекуррентным способом находим

$$Y_j = \max(Y_{j-1}, X_{j-1}) + Z_j - Z_{j-1}. \tag{2.53}$$

Отсюда следует, что время окончания обслуживания j-го требования равно времени его поступления плюс время обслуживания, если требование не ожидало; если же оно ожидало, то это время равно времени окончания обслуживания $(j-1)$-го требования плюс время обслуживания j-го требования. Таким образом, коль скоро начальное число требований i, а также X_j и Z_j известны, то процесс обслуживания описан полностью. Ясно, что X_j не зависит от Z_j.

Допустим, что начальное число требований, находящихся в очереди, задано геометрическим распределением $(1-\rho)\rho^i$ $(i = 0, 1, 2, \ldots)$, которое описывает длину очереди для стационарного состояния системы. Согласно предположению, сделанному для этой задачи, последовательность X_j $(j=1, 2, \ldots)$ представляет собой пуассоновский процесс с параметром λ, а Z_j — пуассоновский процесс с параметром μ. Оба процесса не зависят от i.

Заметим, что m — число требований, обслуженных за время τ, — зависит только от Z_j, так как m — индекс максимума, который указывает, что $Z_m \leqslant \tau$. Поэтому m не зависит от i и описывается следующим законом распределения:

$$e^{-\mu\tau}\frac{(\mu\tau)^m}{m!}, \quad m = 0,1,2,\ldots. \tag{2.54}$$

При заданном m распределение $Z_j (j=1, \ldots, m)$ не зависит от μ и λ. Оно описывает деление интервала τ на $m+1$ частей случайной длины. Итак, нам известны распределения величин X_j, Z_j, i и m. Теперь найдем распределение выборок, объем которых равен m.

Заметим, что поскольку начальное число требований, находящихся в очереди, равно i, то моменты X_j поступления требований относятся к $m-i$ членам выборки. Плотность вероятности момента поступления первого требования равна $\lambda e^{-\lambda X_1}$, для требования, поступающего в момент времени X_2, равна $\lambda e^{-\lambda(X_2-X_1)}$ и т. д. до $(m-i)$-го требования. Произведение их равно $\lambda^{m-i} e^{-\lambda X_{m-i}}$ и не зависит от Z_j.

Предположим, что к моменту времени T поступило n требований и за это же время было обслужено m требований. Чтобы общее число требований оставалось равным n, за промежуток между моментом времени X_{m-i} и моментом времени $Y_m = T$, когда заканчивается обслуживание m-го требования, должно поступить

$n-m-i$ требований. Вероятность того, что за этот промежуток времени (X_{m-i}, T) поступит данное число требований, равна

$$e^{-\lambda(T-X_{m-i})} \frac{[\lambda(T-X_{m-i})]^{n-m+i}}{(n-m+i)!}, \quad \begin{matrix} n-m+i=0,1,\ldots, \\ n+i>0. \end{matrix} \tag{2.55}$$

Перемножив между собой функцию распределения входящего потока, функцию распределения начального числа требований, функцию распределения Z_j (последняя не зависит от μ и λ) и функцию распределения m, получим функцию правдоподобия

$$K\left(1-\frac{\lambda}{\mu}\right)\mu^{m-i}\lambda^{n+i}e^{-\mu\tau-\lambda T},$$

где K — функция, которая не зависит от λ и μ и вид которой не играет роли, так как она равна нулю при приравнивании нулю производных по λ и μ соответственно.

Таким образом, критерий наибольшего правдоподобия для λ и μ выводится из двух равенств, полученных путем подстановки наблюденных значений i, m и n, при равенстве нулю соответствующих производных по λ и μ:

$$\begin{aligned} \hat{\lambda} &= (\hat{\mu}-\hat{\lambda})(n+i-\hat{\lambda}T), \\ \hat{\lambda} &= (\hat{\lambda}-\hat{\mu})(m-i-\hat{\mu}\tau). \end{aligned} \tag{2.56}$$

Кроме того, при подстановке $\hat{\lambda}=\hat{\mu}\hat{\rho}$ после исключения $\hat{\mu}$ получим квадратное уравнение, из которого найдем $\hat{\rho}$ — оценку для ρ. Данное уравнение обладает единственным решением, которое находится между нулем и единицей. Поэтому оценка $\hat{\rho}$ является однозначной.

Упражнение 6. Найдите это квадратное уравнение и однозначную оценку $\hat{\rho}$.

2.8. Процесс принятия решений. (Метод оценки частоты появления неисправностей в каналах обслуживания)

Содержание этого параграфа не имеет непосредственного отношения к теории массового обслуживания, однако излагаемый метод может оказаться полезным при определении сроков ремонта или переоборудования обслуживающего устройства. Для автоматических каналов обслуживания, например для телефонных линий, требуются периодические технические осмотры, чтобы предотвратить возможные отказы. Частота α_0, соответствующая предельному сроку действия системы, определяется с учетом стоимости эксплуатации и стоимости отыскания неисправностей, возникающих с различной частотой. Если выборки показывают, что частота отказов больше α_0, то после этого канал обслуживания исправляется, в противном случае не предпринимается никаких действий. Неверные выводы, полученные на основании выборок, могут причинить убыток.

Если n — число наблюдений, x — наблюденное число неисправностей, а α — частота появления неисправностей и если X, число неисправностей, есть случайная величина, распределенная по закону Пуассона (это допущение справедливо, если вероятность появления неисправностей при каждом наблюдении мала и постоянна, а число наблюдений велико), то

$$P(X \leqslant x) = \sum_{k=0}^{x} \frac{(n\alpha)^k}{k!} e^{-n\alpha} \equiv F_\alpha(x). \qquad (2.57)$$

Для данного n проверяется справедливость гипотезы $\alpha = \alpha_0$, конкурирующей с гипотезой $\alpha \neq \alpha_0$. Нулевой гипотезой является $\alpha = \alpha_0$, альтернативные гипотезы $\alpha < \alpha_0$ или $\alpha > \alpha_0$. Задаваясь n и α_0, находим такие значения x_1 и x_2, что

$$F_{\alpha_0}(x_1) \leqslant \varepsilon_1,$$

$$1 - F_{\alpha_0}(x_2) \leqslant \varepsilon_2,$$

где $\varepsilon_1 + \varepsilon_2$ — выбранный уровень значимости.

Если $x_1 < x \leqslant x_2$, то принимается нулевая гипотеза, в противном случае она отвергается. При этом всегда с вероятностью, меньшей или равной $\varepsilon_1 + \varepsilon_2$, рискуем принять неверное решение, т. е. отвергнуть правильную гипотезу.

С помощью последовательных выборок находят различные значения n с соответствующими значениями x_1 и x_2. Нередко должен быть установлен верхний предел числа наблюдений.

Путем исследования различных уровней риска [209] можно получить выражения для значений оптимального решения.

Упражнение 7. Возьмите все обозначения, введенные выше, и вспомните их смысл.

2.9. Моделирование и применение метода Монте-Карло

При решении задач массового обслуживания для получения численных результатов часто применяется метод Монте-Карло. Применение этого метода можно показать на примере вычисления времени ожидания в одноканальной системе с обслуживанием по принципу «первым пришел — первым обслужен» (табл. 2.2).

Таблица 2.2

Иллюстрация метода Монте-Карло

Промежутки времени между последовательными моментами поступления	Случайные числа, взятые в виде десятичных дробей $\left(1 - e^{0,5t}\right)$	Соответствующее время обслуживания, полученное из графика	Время ожидания
5	0,10	0,20	0
5	0,22	0,50	0
5	0,24	0,55	0
5	0,42	1,05	0
5	0,37	0,90	0
5	0,77	2,90	0
5	0,99	9,00	0
5	0,96	6,30	4,00
5	0,89	4,50	5,30
5	0,85	3,80	4,80
5	0,28	0,65	3,60
5	0,63	2,00	0
5	0,09	0,20	0
...	...	...	...

Примечание. Десятичные дроби округлены.

Графики функции распределения входящего потока и функции распределения времени обслуживания определяются на основании экспериментальных данных или с помощью формул. После этого составляется последовательность случайных двоичных чисел. Представленные в виде десятичных дробей они являются ординатами функции распределения входящего потока. Затем находятся соответствующие им абсциссы. Последние образуют последовательность временны́х интервалов между требованиями. Аналогично образуются две другие последовательности чисел для описания времени обслуживания. В четвертом столбце для каждого клиента вычисляется время ожидания как сумма времени ожидания и времени обслуживания предыдущего клиента минус задержка в прибытии нового клиента относительно момента прибытия предыдущего клиента. Если эта разность отрицательна, то время ожидания полагается равным нулю.

Среднее общее время ожидания вычисляется непосредственно. Для того чтобы существовало стационарное распределение, требуется условие $\frac{\lambda}{\mu} < 1$. Прежде, чем применять метод Монте-Карло, следует проверить данное неравенство. Приближенно можно считать, что точность метода пропорциональна $\sqrt{N}$, где N — число проб. Теперь проиллюстрируем это на численном примере.

Было подсчитано, что ускорение операций на конвейере дает возможность каждые 5 мин. доставлять на линию технического осмотра одно изделие. Когда измерения времени, затрачиваемого на осмотр, были соответствующим образом сгруппированы и по точкам была построена кривая, то оказалось, что она соответствует экспоненциальному распределению

$$b(t) = 0{,}5e^{-0{,}5t}.$$

В этой задаче при вычислении среднего времени ожидания, которое производится с целью увеличения производительности процесса обслуживания вследствие возросших темпов выпуска продукции, использовался метод Монте-Карло.

В данном случае входящий поток характеризуется постоянными интервалами между требованиями. В любом эксперименте должна использоваться информация о входящем потоке, если она имеется. С другой стороны, время обслуживания каждого требования случайно; в связи с этим необходимо соблюдать условие, чтобы случайные числа, используемые для моделирования времен обслуживания различных требований, были независимы. Последовательности случайных чисел были образованы так, как описано выше, и была вычерчена кривая распределения времени обслуживания. Было рассмотрено 100 изделий, поступивших на проверку. Вычисленное среднее время ожидания оказалось равным 0,25 мин., следовательно, дополнительный технический осмотр не обеспечивал плавного протекания процесса производства.

Упражнение 8. Повторите этот эксперимент, взаимно поменяв местами распределение входящего потока и распределение времени обслуживания и их параметры, т. е. рассмотрите экспоненциальное распределение промежутков времени между требованиями, равное $0{,}2\,e^{-0{,}2t}$, и обслуживание с постоянной длительностью, равной двум временны́м единицам. Вычислите среднее время ожидания. Сравните этот результат со значением, полученным по формуле Полячека—Хинчина.

2.10. Разложение в степенной ряд

Изложим метод, иллюстрирующий выполнение вычислений при получении численных решений задач массового обслуживания. В данном случае понимание реального смысла применяемых уравнений не обязательно. В качестве примера рассмотрим типичную задачу массового обслуживания и покажем, как получить решение с помощью разложения в ряд вероятностей $P_n(t)$ при начальном числе i требований, находящихся в очереди, и заданных параметрах λ и μ.

Рассмотрим задачу (с двумя поглощающими барьерами — в начале координат и в точке $n=N$; процесс прекращается, как только он достигает какой-либо из этих точек), описываемую уравнениями:

$$
\begin{aligned}
P_0'(t) &= \mu P_1(t),\\
P_1'(t) &= -(\lambda+\mu) P_1(t) + 2\mu P_2(t),\\
P_n'(t) &= (n-1)\lambda P_{n-1}(t) - n(\lambda+\mu) P_n(t) + (n+1)\mu P_{n+1}(t),\\
P_{N-1}'(t) &= (N-2)\lambda P_{N-2}(t) - (N-1)(\lambda+\mu) P_{N-1}(t),\\
P_N(t) &= (N-1)\lambda P_{N-1}(t),
\end{aligned}
$$

при начальном условии $P_n(0)=\delta_{in}$.

Пусть $i=2$ и $N=5$; допустим, что

$$P_n(t) = P_n(0) + P_n'(0)\,t + P_n''(0)\frac{t^2}{2!} + \ldots.$$

Теперь вычислим коэффициенты $P_n^{(k)}(0)$, выразив их через параметры λ и μ. Имеем $P_2(0)=1$ и $P_n(0)=0$ при $n\neq 2$. Начиная с третьего уравнения, получим

$$P_2'(0) = \lambda P_1(0) - 2(\lambda+\mu) P_2(0) + 3\mu P_3(0),$$

и, следовательно, $P_2'(0) = -2(\lambda+\mu)$, так как $P_1(0)=P_3(0)=0$. Кроме того, из уравнения, определяющего $P_2(t)$, находим

$$P_2''(0) = \lambda P_1'(0) - 2(\lambda+\mu) P_2'(0) + 3\mu P_3'(0).$$

Но

$$P_1'(0) = -(\lambda+\mu) P_1(0) + 2\mu P_2(0).$$

После подстановки начального условия $P_1'(0)=2\mu$ и

$$P_3'(0) = 2\lambda P_2(0) - 3(\lambda+\mu) P_3(0) + 4\mu P_4(0) = 2\lambda,$$

поэтому

$$P_2''(0) = 2\mu\lambda + 4(\lambda+\mu)^2 + 6\mu\lambda.$$

В свою очередь,

$$P_2'''(0) = \lambda P_1''(0) - 2(\lambda+\mu) P_2''(0) + 3\mu P_3''(0),$$

где

$$P_1''(0) = -(\lambda+\mu) P_1'(0) + 2\mu P_2'(0) = -6\mu(\lambda+\mu)$$

и

$$P_3''(0) = 2\lambda P_2'(0) - 3(\lambda+\mu) P_3'(0) + 4\mu P_4'(0).$$

Но

$$P_4'(0) = 3\lambda P_3(0) - 4(\lambda+\mu) P_4(0) = 0.$$

Следовательно, $P_3''(0) = -10\lambda(\lambda+\mu)$.

Наконец, получаем

$$
\begin{aligned}
P_2'''(0) = \lambda[-6\mu(\lambda+\mu)] - 2(\lambda+\mu)[8\mu\lambda + 4(\lambda+\mu)^2] +\\
+ 3\mu[-10\lambda(\lambda+\mu)].
\end{aligned}
$$

Аналогично можно вычислить $P_2^{IV}(0)$, и т. д. Таким образом,

$$P_2(t) = 1 - 2(\lambda + \mu)t + [8\mu\lambda + 4(\lambda + \mu)^2]\frac{t^2}{2!} + \ldots.$$

Кроме того, нам известно, что

$$P_1(t) = P_1(0) + P_1'(0)t + P_1''(0)\frac{t^2}{2!} + \ldots,$$

откуда после подстановки результатов выполненных ранее вычислений получим

$$P_1(t) = 2\mu t + [-6\mu(\lambda + \mu)]\frac{t^2}{2!} + \ldots.$$

Так так $P_0'(t) = \mu P_1(t)$, то

$$P_0(t) = 2\mu^2\frac{t^2}{2!} - 6\mu^2(\lambda + \mu)\frac{t^3}{3!} + \ldots.$$

В свою очередь,

$$P_3(t) = 2\lambda t - 10\lambda(\lambda + \mu)\frac{t^2}{2!} + \ldots.$$

Аналогично находим $P_4(t)$ и $P_5(t)$.

В качестве примера читатель может вычислить указанным способом $P_4(t)$. Вследствие необходимости выполнения условия сходимости эти разложения справедливы только в окрестности точки $t=0$. Используя метод приращений, можно вычислить значения $P_n(t)$, удаленные от начала отсчета. Для этого разделим временную ось на интервалы длительностью, скажем, в одну единицу каждый. Вычисление вероятности для промежутка времени заданной длины начнем с $t=0$, при этом получим значение $P_n(0)$. Для вычисления $P_n(1)$ используем приближенное значение приращения, которое дает значение P_n в момент времени $t+\Delta t$. В исходной системе, n-е уравнение которой имеет вид

$$P_n(t + \Delta t) = P_n(t - \Delta t) + 2\Delta t\,[(n-1)\lambda P_{n-1}(t) - n(\lambda + \mu)P_n(t) + \\ + (n+1)\mu P_{n+1}(t)], \tag{2.58}$$

имеем

$$P_n'(t) \approx \frac{P_n(t + \Delta t) - P_n(t - \Delta t)}{2\Delta t}. \tag{2.59}$$

Таким образом, начинаем с $t=0$ и $\Delta t=1$ и т. д. Заметим, что для отрицательных значений аргумента вероятности равны нулю и только $P_2(-1)=1$.

При таких вычислениях встречаются ошибки двух типов: 1) ошибки округления вследствие пренебрежения остаточным членом ряда при применении аппроксимаций; 2) ошибки вследствие отбрасывания членов при замене производных первыми разностями. Мы не станем углубляться дальше в этот вопрос. Желающие могут изучить методы численного анализа, которые достаточно развиты и повсеместно используются.

Однако, представляется очевидным, что можно найти какой-то прием, аналогичный описанному выше.

Замечание. Кэмп [104] исследовал различные методы получения верхних и нижних оценок решений задач массового обслуживания. Это еще одна интересная идея; она полезна в том случае, когда ясно, какой ответ следует ожидать.

Знание характера решения при указании верхней и нижней границ его значений представляет в этом случае весьма ценную информацию. Понятия о верхнем и нижнем пределах будут встречаться в тексте.

Задачи

1. Используя формулу Полячека—Хинчина, определите среднее число требований, находящихся в очереди и в системе, и соответствующие им время ожидания и время пребывания в системе при экспоненциальном времени обслуживания.

2. Решите систему уравнений

$$qp_1 + qp_2 = p_1,$$
$$pp_{j-1} + qp_{j+1} = p_j, \quad j = 1, \ldots, c-1,$$
$$pp_{c-1} + pp_c = p_c,$$

где

$$\sum_{j=1}^{c} p_j = 1$$

и

$$p + q = 1.$$

Покажите, что

$$p_j = \frac{1 - \frac{p}{q}}{1 - \left(\frac{p}{q}\right)^c} \left(\frac{p}{q}\right)^{j-1} \quad \text{при } p \neq q$$

и

$$p_j = \frac{1}{c} \quad \text{при } p = q.$$

3. Используя таблицу случайных чисел, составьте модель для 100 промежутков времени и определите среднее время ожидания и среднее время простоя канала в одноканальной системе с пуассоновским входящим потоком ($\lambda = 1$) и постоянной интенсивностью обслуживания, равной двум требованиям в единицу времени, при обслуживании в порядке поступления. Сравните ответ с результатом, полученным по формуле Полячека—Хинчина.

4. Для вывода экспоненциального распределения рассмотрим урну с большим числом шаров диаметром d. В урне B черных и A белых шаров. Задано, что вынут черный шар. Вероятность того, что расстояние x между его центром и центром черного шара, который будет вынут следующим, не больше rd, задана в виде

$$P(x \leqslant rd) = 1 - (1 - b)^r, \quad b = \frac{B}{A+B}.$$

Пусть $F_s(x)$ — плотность распределения расстояния x, $d = \frac{1}{s}$ и $sb = \lambda$. Если $b \to 0$ и $s \to \infty$ так, что λ и x остаются неизменными, то $F_s(x) = 1 - \left(1 - \frac{\lambda}{s}\right)^{sx} \to 1 - e^{-\lambda x}$. Интерпретируйте λ, как математическое ожидание частоты событий на единицу длины.

Глава 3

ТЕОРИЯ ВЕРОЯТНОСТЕЙ, ЦЕПИ И ПРОЦЕССЫ МАРКОВА, ЭРГОДИЧЕСКИЕ СВОЙСТВА СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

3.1. Введение

Для удобства читателя в этой главе сообщаются некоторые понятия из теории вероятностей и теории случайных процессов. Бóльшая часть материала книги не связана с излагаемыми здесь вопросами, за исключением теорем об эргодичности систем массового обслуживания и примеров на вычисление времени ожидания. В дополнение к этому материалу читателю полезно изучить вводный курс теории вероятностей. Здесь можно найти некоторые указания, которыми можно воспользоваться для дальнейшего углубления знания материала. Однако при выводе большинства уравнений в различных частях книги не потребуется глубокого знания теории вероятностей.

Рассмотрение некоторых частных случаев процесса рождения и гибели, подобных описанным в предыдущей главе (модель Эрланга), в сочетании с хорошим пониманием рассматриваемой системы обслуживания, как нам кажется, будет очень полезно для формирования интуиции читателя.

Понятие цепей Маркова и марковских процессов является фундаментальным и весьма полезным. Хотя марковские процессы и встречаются в теории массового обслуживания, однако нельзя делать вывод о том, что все процессы массового обслуживания являются марковскими.

Марковские процессы характеризуются тем, что при фиксированном состоянии системы ее будущее поведение не зависит от поведения в прошлом. Это свойство упрощает анализ систем массового обслуживания, так как приводит к удобным функциональным уравнениям. Впрочем, исследуются также и немарковские процессы.

В этой главе кратко рассматриваются такие понятия, как событие, булева алгебра, сигма-алгебра, пространство с мерой и вероятность. Кроме того, вводятся такие понятия, как условная вероятность и независимые события. Затем следует определение случайной величины. Кратко сообщаются наиболее важные положения из теории вероятностей. Из-за недостатка места вывод формул, доказательства и примеры опущены. Это предлагается выполнить читателю. Для некоторых важных распределений дается сводка формул. Затем вводится понятие вероятностного процесса, после чего рассматриваются цепи и процессы Маркова, приводится классификация состояний по Феллеру и даются теоремы, связанные с этим вопросом. Все эти положения используются затем в теории массового обслуживания. Наконец, приводятся примеры некоторых вероятностных процессов. Наиболее важным из них вследствие его широкого применения в теории массового обслуживания является процесс рождения и гибели, который будет рассмотрен в следующей главе.

3.2. Основные понятия теории вероятностей

Теория вероятностей занимается количественным описанием случайных событий. Изучение ее основных понятий необходимо начинать с события. В абстрактной форме события можно представить как множества точек. Комбинируя их, можно получать другие события. Для двух событий a и b запись в виде $a \cup b$ употребляется для обозначения события, которое происходит в том и только в том случае, если происходит хотя бы одно из них, а $a \cap b$ означает событие, которое состоит в совместном появлении событий a и b. Событие a' является противоположным событию a, оно происходит в том и только в том случае, если событие a не происходит. Достоверное событие обозначается e, а невозможное O.

Единственным основанием для введения алгебраических операций над событиями является следующее: 1) событие может являться комбинацией других событий и 2) необходимы правила для обращения с комбинациями событий, определяемыми через ИЛИ и И.

Система элементов (в данном случае совокупность событий) $O, a, b, \ldots, e$, в которой определены операции a', $a \cup b$, $a \cap b$, называется булевой алгеброй, если выполняются следующие соотношения:

$$
\begin{aligned}
O \cap a &= O, & O \cup a &= a,\\
e \cap a &= a, & e \cup a &= e,\\
a \cap b &= b \cap a, & a \cup b &= b \cup a,\\
(a \cap b) \cap c &= a \cap (b \cap c), & (a \cup b) \cup c &= a \cup (b \cup c),\\
a \cap (b \cup c) &= (a \cap b) \cup (a \cap c), & a \cup (b \cap c) &= (a \cup b) \cap (a \cup c),\\
O' &= e, & e' &= O, \ (a')' = a,\\
(a \cap b)' &= a' \cup b', & (a \cup b)' &= a' \cap b',\\
a \cap a' &= O, & a \cup a' &= e.
\end{aligned}
$$

Это определение справедливо для конечного числа событий. При бесконечном множестве событий необходимо ввести понятие сигма-алгебры. Вначале рассмотрим обозначение $a\subset b$. Оно означает, что a содержится в b, или b включает в себя a (например, совокупность элементов $\{1, 2, 3\}$ содержится в $\{1, 2, 3, 4\}$). В дополнение к приведенным соотношениям в сигма-алгебре требуется, чтобы вместе с любой бесконечной последовательностью событий $\{a_n\}$, принадлежащих сигма-алгебре, она содержала также событие, задаваемое условием $a\equiv a_1\cup a_2\cup\ldots$. Таким образом, если в определение булевой алгебры входит свойство «счетной аддитивности», то имеем сигма-алгебру.

Кроме того, для булевой алгебры справедливы следующие соотношения: $a\cap a=a$; $a\cup a=a$; $a\subset a$; из $a\subset b$ и $b\subset a$ вытекает, что $a=b$; из $a\subset b$ и $b\subset c$ вытекает, что $a\subset c$; три условия: $a\subset b$, $a\cap b=a$ и $a\cup b=b$ — эквивалентны.

Вероятность есть числовая функция $P(a)$ элементов булевой сигма-алгебры, для которой

$$0\leqslant P(a)\leqslant 1,$$

а также $P(a)=0$ при $a=0$ и $P(a)=1$ при $a=e$. Кроме того,

$$P(a\cup b)=P(a)+P(b)$$

при $a\cap b=O$. По индукции находим

$$P(a_1\cup\ldots\cup a_n)P(a_1)+\ldots+P(a_n),$$

где $a_i\cap a_j=O$ при $i\neq j$.

Если $\{a_n\}$ — бесконечное множество, то для вероятности выполняется условие счетной аддитивности, т. е.

$$P(a_1\cup a_2\cup\ldots)=P(a_1)+P(a_2)+\ldots,$$

где

$$a_i\cap a_j=O;\quad i\neq j.$$

Всеми этими свойствами с незначительными изменениями обладает мера общего вида (которая, однако, не обязательно должна находиться между нулем и единицей). Таким образом, вероятность есть мера, т. е. это конечная, неотрицательная счетно-аддитивная функция элементов булевой сигма-алгебры. Кроме того, $P(a)=1$ означает, что $a=e$.[1]

Как показал Халмош, если B есть любая булева сигма-алгебра, а P есть вероятностная мера в B, то существует такое пространство Ω с мерой, что система B абстрактно тождественна алгебре подмножеств из Ω (в случае необходимости множества, разность которых имеет меру нуль, отождествляются), а значение P для

---

[1] Читатель должен иметь в виду, что изложение вопросов, связанных с мерой, у автора весьма поверхностно и нестрого. Поскольку соответствующие факты имеются в строгом изложении во многих книгах, мы не станем комментировать подобные рассуждения. — *Прим. ред.*

любого события тождественно со значением меры для соответствующего подмножества из пространства Ω. Таким образом, понятия меры и вероятности тождественны. В теории вероятностей мера всего пространства конечна и равна 1.

Теперь дадим определение условной вероятности и независимости событий. Введем обозначение

$$P_b(a) \equiv \frac{P(a \cap b)}{P(b)}, \tag{3.1}$$

где $P_b(a)$ — вероятность появления события a при условии, что произошло событие b.

Если же появление события b не влияет на появление события a, то

$$P_b(a) = P(a).$$

Следовательно, события a и b независимы в том и только в том случае, если

$$P(a \cap b) = P(a) P(b). \tag{3.2}$$

Широкое применение находят следующие формулы:

1. Формула сложения вероятностей

$$P(a_1 \cup \ldots \cup a_n) = \sum_{i=1}^{n} P(a_i) - \sum_i \sum_j P(a_i \cap a_j) + \\ + \sum_i \sum_j \sum_k P(a_i \cap a_j \cap a_k) - \ldots \pm P(a_1 \cap a_2 \cap \ldots \cap a_n).$$

2. Формула полной вероятности

$$P(a \cap b) = \sum_i P(b_i) P(a \mid b_i),$$

где $P(a \mid b_i)$ — вероятность появления события a при условии, что произошло событие b_i, вероятность появления которого равна $P(b_i)$, и $b = \bigcup_i b_i$, где b_i — несовместные события.

Случайная величина есть измеримая функция, определенная в пространстве с мерой, при суммарной мере, равной единице[1]. Говорят, что функция $x(\omega)$, определенная в пространстве Ω, измерима в том и только в том случае, если при любых α и β множества тех ω, для которых $\alpha \leqslant x(\omega) \leqslant \beta$, принадлежат основной сигма-алгебре B из пространства Ω.

Функция распределения $F(x)$ случайной величины X есть функция действительного переменного, определенная для каждого действительного значения x как вероятность того, что $X \leqslant x$, т. е. $F(x) = P(X \leqslant x)$.

---

[1] Halmos P. R. The Foundations of Probability, *Am. Math. Monthly*, 1944, vol. 51, p. 493—510.

Множество значений x может быть дискретным или непрерывным. Мы будем рассматривать только непрерывные величины, так как краткое описание случая дискретных величин уже давалось. При вычислении функции $F(x)$ читатель может воспользоваться понятиями, которые приводились для случая дискретных значений, как, например, при бросании двух игральных костей.

Функция $F(x)$ (которая непрерывна справа) — неубывающая функция, она стремится к нулю при $x \to -\infty$ и к единице при $x \to \infty$. Плотность вероятности случайной величины X определяется выражением

$$f(x) = \frac{dF(x)}{dx}; \tag{3.3}$$

плотность существует, если функция $F(x)$ абсолютно непрерывна. В этом случае

$$F(x) = \int_{-\infty}^{x} f(y)\,dy.$$

Определим характеристическую функцию распределения $F(x)$ с помощью интеграла Римана—Стилтьеса:

$$\varphi(t) \equiv E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} dF(x), \quad i = \sqrt{-1}, \tag{3.4}$$

который существует для действительных t и необязательно существует для мнимых t. Если моментная производящая функция существует, то она имеет вид

$$M(t) = \int_{-\infty}^{\infty} e^{-tx} dF(x), \tag{3.5}$$

т. е. она является преобразованием Лапласа—Стилтьеса функции $F(x)$. Если $M(t)$ существует, то $M(t) = \varphi(it)$.

Характеристическая функция и моментная производящая функция являются важными инструментами при вычислении моментов распределений, при отыскании распределения суммы случайных величин и при исследовании пределов последовательностей распределений [см. (3.8) и задачу 3], что будет проиллюстрировано на примерах.

Характеристическая функция (моментная производящая функция) суммы независимых случайных величин равна произведению их характеристических функций (моментных производящих функций). Если $Z = X + Y$, где X и Y — независимые случайные величины, характеристические функции (моментные производящие функции) которых равны соответственно $F(x)$ и $G(x)$, то характеристическая функция (моментная производящая функция) для Z имеет вид

$$H(z)=\int_{-\infty}^{\infty}\int_{-\infty}^{z-y} dF(x)\,dG(y)=\int_{-\infty}^{\infty} F(x)\Big|_{-\infty}^{z-y} dG(y)=$$

$$=\int_{-\infty}^{\infty} F(z-y)\,dG(y). \tag{3.6}$$

Если соответствующая ей плотность распределения существует, то она равна

$$h(z)=\int_{-\infty}^{\infty} f(z-y)\,g(y)\,dy. \tag{3.7}$$

Если распределения сосредоточены только на положительных значениях переменных, а для остальных значений равны нулю, то

$$H(z)=\int_{0}^{\infty} F(z-y)\,dG(y)=\int_{0}^{z} F(z-y)\,dG(y). \tag{3.8}$$

Плотность вероятности, если она существует, может быть получена также путем дифференцирования по z. (Заметим, что интеграл

$$\int_{0}^{x} f(x-y)\,g(y)\,dy \tag{3.9}$$

называется сверткой функций f и g.) Как будет показано впоследствии, этот вывод можно распространить на сумму нескольких случайных величин. Вводя производящие функции для дискретных случайных величин, можно получить аналогичные результаты.

В книге Дуба [187] приводится доказательство того, что с помощью характеристической функции можно получить функцию распределения $F(x)$:

$$\frac{1}{2}[F(x+0)-F(x-0)]-\frac{1}{2}[F(0+0)-F(0-0)]=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)\frac{1-e^{-itx}}{it}\,dt, \tag{3.10}$$

где, например, $x+0$ означает, что аргумент стремится к x со стороны значений, бо́льших x. Начальные моменты (первый из которых есть математическое ожидание μ_1) можно получить, взяв n-ю производную от φ при нулевом значении аргумента:

$$\mu_n=E(X^n)=(-i)^n\varphi^{(n)}(0),\qquad n=0,\ 1,\ \ldots.$$

В дополнение к моментам семиинварианты k_r образуют другой ряд важных характеристик. Они определяются из тождества

$$\exp\left(k_1 t + \frac{k_2 t^2}{2!} + \ldots + \frac{k_r t^r}{r!} + \ldots\right) =$$

$$= 1 + \mu_1' t + \frac{\mu_2' t^2}{2!} + \ldots + \frac{\mu_r' t^r}{r!} + \ldots,$$

где $\mu_r' \equiv E[(X-a)^r]$ — r-й момент относительно произвольной точки a.

Если заменить t на it, то правая часть выражения будет представлять разложение функции $\varphi(t)$ в степенной ряд. Таким образом, μ_r' — коэффициент при $\frac{(it)^r}{r!}$ в разложении функции $\varphi(t)$, а k_r — коэффициент при $\frac{(it)^r}{r!}$ в разложении функции $\log\varphi(t)$, т. е.

$$k_r = (-i)^r \frac{d^r}{dt^r} \log\varphi(t)\,|_{t=0}.$$

Значит,

$$\mu_1' = k_1,\ \mu_2' = k_2 + k_1^2.$$

Для моментов относительно математического ожидания μ_1 (центральных моментов) имеем

$$k_1 = 0,\ \mu_2 = k_2.$$

Для пуассоновского распределения, характеристическая функция которого равна $\exp[\lambda(e^{it} - 1)]$, все семиинварианты равны λ.

Поясним некоторые из этих положений на примере, а затем используем полученные результаты для вычисления времени ожидания в марковской модели массового обслуживания.

Пример 1. Покажем, что закон Пуассона можно получить, рассматривая распределение суммы одинаково распределенных независимых случайных величин $t_i (i - 1, \ldots, n)$ с соответствующими экспоненциальными распределениями $\mu e^{-\mu t_i}$.

Моментная производящая функция экспоненциальной плотности равна

$$M(\theta;\ t_i) = \int_0^\infty e^{\theta t_i} \mu e^{-\mu t_i} dt_i = \frac{\mu}{\mu - \theta}.$$

Если рассматривать каждое значение t_i как длительность обслуживания i-го требования в одноканальной системе при обслуживании в порядке поступления требований, то нам потребуется определить время ожидания n-го требования при условии, что в очереди находится n требований и одно требование только что поступило на обслуживание.

Чтобы получить распределение суммы случайных величин, возьмем произведение их моментных производящих функций и найдем плотность вероятности, соответствующую полученному произведению. Последнее равно $(1-\theta/\mu)^{-n}$, следовательно, плотность вероятности будет равна

$$\mu \frac{(\mu t)^{n-1} e^{-\mu t}}{(n-1)!}, \quad t \geqslant 0.$$

Упражнение. Покажите, что математическое ожидание этого распределения равно $\frac{n}{\mu}$, а дисперсия $-\frac{n}{\mu^2}$.

Функция распределения, соответствующая этой плотности, есть вероятность того, что длительность ожидания n-го требования меньше t. Требуется вычислить $P_n(\tau)$ — вероятность того, что за время τ будет обслужено ровно n требований. Пусть $\tau = t + (\tau - t)$. Вероятность того, что за время t будет обслужено n требований, получена выше, а вероятность того, что за время $\tau - t$ ни одно требование не будет обслужено, равна $e^{-\mu(\tau-t)}$, так как вероятность того, что обслуживание одного требования закончится за время x, равна $1-e^{\mu x}$, а вероятность того, что за время x требование не будет обслужено, равна $1-(1-e^{-\mu x}) = e^{-\mu x}$. Таким образом,

$$P_n(\tau) = \int_0^{\tau} \mu \frac{(\mu t)^{n-1}}{(n-1)!} e^{-\mu t} e^{-\mu(\tau-t)} dt = \frac{(\mu\tau)^n}{n!} e^{-\mu\tau},$$

т. е. получили закон Пуассона.

Пример 2. Вычислим время ожидания и время пребывания в системе, описанной с помощью модели Эрланга. Воспользовавшись выражениями, полученными в предыдущем примере, произведем непосредственное вычисление распределения времени ожидания следующим образом.

Поступающее требование будет в системе $(n+1)$-м, если впереди него находится n требований. Как уже было показано, вероятность этого события равна $\rho^n(1-\rho)$. Время ожидания поступающего требования — величина случайная. Оно равно сумме случайных длительностей обслуживания n предыдущих требований. Нам известно, что время обслуживания имеет экспоненциальное распределение с параметром μ. Кроме того, мы знаем, что вследствие свойства «отсутствия памяти» экспоненциального распределения время, оставшееся до конца обслуживания требования, имеет то же распределение, что и время обслуживания требования, только что поступившего на обслуживание.

В предыдущем параграфе мы получили выражение для плотности времени ожидания $(n+1)$-го требования

$$w_n(t) \equiv \frac{\mu (\mu t)^{n-1} e^{-\mu t}}{(n-1)!}. \tag{3.11}$$

Так как рассматривается произвольное требование, которое может занимать в очереди любое положение, то для того чтобы найти вероятность ожидания в промежутке между t и $t+\Delta t$, необходимо умножить вероятность наличия в системе n требований на плотность вероятности времени ожидания $(n+1)$-го требования и результат просуммировать по всем $n \geqslant 1$. При $t>0$ получим

$$w(t) \equiv \sum_{n=1}^{\infty} \rho^n (1-\rho) \mu \frac{(\mu t)^{n-1} e^{-\mu t}}{(n-1)!} = \mu\rho(1-\rho) e^{-(1-\rho)\mu t}. \quad (3.12)$$

Среднее время ожидания найдем, умножив правую часть выражения (3.12) на t и взяв интеграл от нуля до бесконечности. Получим

$$W_q = \frac{\rho}{\mu(1-\rho)}. \quad (3.13)$$

Функция распределения $P(\leqslant t)$ есть вероятность того, что длительность ожидания не превышает t. Она находится следующим образом:

$$P(\leqslant t) = 1 - P(>t) = 1 - \int_t^{\infty} w(x) dx = 1 - \rho e^{-(1-\rho)\mu t}. \quad (3.14)$$

Заметим, что $w(0) = (1-\rho)\delta(t)$, где $\delta(t)$ — дельта-функция, которая всюду равна нулю, кроме начальной точки, где ее значение равно бесконечности; интеграл от дельта-функции в бесконечных пределах равен единице.

Упражнение. Покажите, что функция распределения времени пребывания в системе имеет вид

$$1 - e^{-(1-\rho)\mu t}.$$

Найдите среднее время пребывания в системе.

Пример 3. Вычислим число требований, находящихся в системе, и время ожидания в очереди для модели Эрланга, если система имеет ограниченное место для ожидания. Если число требований, находящихся в системе, не может превосходить N (т. е. место для ожидания ограничено), то уравнения Эрланга применимы до $n=N$; при этом вид последнего уравнения несколько изменится. Решение для этого случая легко найдем, используя решение, полученное для $N=\infty$:

$$p_n = \frac{p_0\left(\frac{\lambda}{\mu}\right)^n}{p_0 \sum\limits_{n=0}^{N} \left(\frac{\lambda}{\mu}\right)^n} = \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{N+1}} \left(\frac{\lambda}{\mu}\right)^n, \quad n = 0, \ldots, N. \quad (3.15)$$

Применяя правило Лопиталя, можно найти решение также при $\lambda=\mu$. Если $N \to \infty$, получим значения вероятностей для системы с бесконечной очередью при $\lambda<\mu$.

Плотность вероятности времени ожидания равна

$$w(t)=\sum_{n=0}^{N}p_n w_n(t).$$

Следовательно, среднее время ожидания в очереди составляет

$$W_q=\sum_{n=0}^{N}p_n\int_0^\infty t w_n(t)\,dt=\frac{1}{\mu}\sum_{n=1}^{N}np_n=$$

$$=\frac{\rho\left[1-(N+1)\rho^N+N\rho^{N+1}\right]}{\mu\left(1-\rho^{N+1}\right)(1-\rho)},\quad \rho=\frac{\lambda}{\mu},\tag{3.16}$$

откуда можно получить известную уже формулу для среднего времени ожидания, когда не накладываются ограничения на место для ожидания, т. е. когда $N\to\infty$. Заметим, что согласно (3.16) $W_q=\frac{L}{\mu}$.

Упражнение. Проверьте это выражение для W_q и вычислите W_q при $N\to\infty$. Найдите $P(\leqslant t)$ из выражения для $w(t)$.

Пример 4. Незначительно изменим полученные ранее соотношения между временем ожидания n-го и временем ожидания $(n+1)$-го требований, включив вероятность отсутствия ожидания. Получим следующее выражение:

$$w_{n+1}=\max(w_n+s_n-t_n,\ 0).$$

Нужно вычислить распределение времени ожидания $P(\leqslant t)$ при стационарном состоянии. Допустим, что $t>0$, т. е. $(n+1)$-е требование будет ожидать:

$$P(w_{n+1}<w)=P(w_n+s_n-t_n<w)=P(w_n<w-s_n+t_n).$$

Обозначим через $P_{n+1}(w)$ и $P_n(w)$ распределение времени ожидания в очереди соответственно $(n+1)$-го и n-го требований. Члены s_n и t_n в правой части выражения — независимые случайные величины с распределениями, равными соответственно $A(x)$ и $B(y)$. Кроме того, время ожидания n-го требования не зависит от времени его обслуживания и от промежутка времени, через который поступит $(n+1)$-е требование.

Поэтому вероятность сложного события, состоящего в том, что 1) время ожидания w_n удовлетворяет приведенному выше неравенству; 2) промежуток времени между моментами поступления требований лежит в пределах $x\leqslant t_n<x+dx$ и 3) длительность обслуживания s_n заключена в пределах $y\leqslant s_n<y+dy$, равна

$$P_n(w+x-y)\,dA(x)\,dB(y).$$

Суммируя эти вероятности, получим [680]

$$P_{n+1}(w)=\int_0^\infty\int_0^\infty P_n(w+x-y)\,dA(x)\,dB(y).$$

Эту формулу можно также получить, рассматривая распределение суммы трех независимых случайных величин. Как будет показано в гл. 9, в стационарном состоянии все требования будут иметь одно и то же распределение времени ожидания $P(w)$; следовательно,

$$P(w)=\int_0^\infty\int_0^\infty P(w+x-y)\,dA(x)\,dB(y).$$

Допустим теперь, что время обслуживания имеет постоянную длительность, равную одной единице времени. Тогда

$$B(y)=\begin{cases}0, & y<1,\\ 1, & y>1,\end{cases}$$

$$B(1+0)-B(1-0)=1$$

и

$$P(w)=\int_0^\infty P(w+x-1)\,dA(x),$$

где $P(w)=0$ при $w<0$.

Так как для отрицательных аргументов вероятность равна нулю, то при $w<1$ имеем

$$P(w)=\int_{1-w}^\infty P(w+x-1)\,dA(x).$$

Интегрируя по частям, получим преобразование Лапласа—Стилтьеса функции $P(w)$

$$\gamma(s)=\int_0^\infty e^{-sw}\,dP(w)=s\int_0^\infty e^{-sw}P(w)\,du.$$

Положим $A(x)=1-e^{-\lambda x}$, подставим значение $P(w)$, учитывая последнее замечание. Получим выражение

$$\gamma(s)=s\int_0^1 e^{-sw}dw\int_{1-w}^\infty \ldots+s\int_1^\infty e^{-sw}dw\int_0^\infty \ldots,$$

которое на основании свойств преобразований приводится к более простому:

$$\gamma(s)=\frac{se^{-\lambda}\gamma(\lambda)}{s-\lambda+\lambda e^{-s}}.$$

Пусть $s\to 0$; тогда $\gamma(s)\to 1$ и $e^{-\lambda}\gamma(\lambda)\to 1-\lambda$.

Окончательно имеем

$$\gamma(s)=\frac{(1-\lambda)\,s}{s-\lambda+\lambda e^{-s}}.$$

Для существования стационарного режима требуется выполнение неравенства $\lambda<1$.

Обратное преобразование можно получить из выражения

$$P(w)=\frac{1}{2\pi i}\int\limits_{C} e^{st}\gamma(s)\frac{ds}{s},$$

где C — контур, идущий от $-i\infty$ до $i\infty$ и обходящий начало координат по малой полуокружности справа.

Подставив значение $\gamma(s)$, передвинув путь интегрирования к

$$\operatorname{Re}(s)=\lambda+\delta>\lambda,$$

что допустимо, так как $\gamma(s)$ регулярно для $\operatorname{Re}(s)\geqslant 0$, и разложив знаменатель по формуле геометрической прогрессии, получим

$$P(w)=(1-\lambda)\sum_{n=0}^{[t]}\frac{\lambda^n(n-t)^n}{n!}e^{\lambda(t-n)},$$

где $[t]$ — наибольшее целое число, меньшее t.

Вычет для членов, находящихся под знаком суммы, находится на полюсе подынтегральной функции $s=\lambda$, в то время как интеграл остальной части разложения, где $t-[t]-1<0$, равен нулю, что найдем, взяв в качестве пути интегрирования полуокружность в правой полуплоскости и устремив ее радиус к бесконечности (см. описание подобной задачи в гл. 9).

Допустим, что имеется несколько случайных величин $X_i\,(i=1,\ \ldots,\ N_t)$, которые непрерывны, независимы и имеют одинаковые распределения $F(x)$. Допустим, что, в свою очередь, N_t, т. е. само число случайных величин, есть величина случайная. Найдем распределение их суммы $S_{N_t}=X_1+\ldots+X_{N_t}$. Можно показать, что если распределение случайных величин N_t не зависит от распределения случайных величин $\{X_i\}$, то справедливо следующее соотношение:

$$P(S_{N_t}\leqslant x)=\sum_{n=0}^{\infty}P(N_t=n)\,P(S_n\leqslant x).$$

Моментная производящая функция этого распределения будет иметь вид

$$\sum_{n=0}^{\infty}P(N_t=n)\left[\int\limits_{-\infty}^{\infty}e^{-\theta x}\,dF(x)\right]^n.$$

Моментная производящая функция имеет показатель степени n, так как рассматривается свертка порядка n. В результате имеем производящую функцию $P(z)$ для вероятностей $P(N_t=n)$

значений дискретной случайной величины. Здесь вместо z (как в гл. 2) используется выражение

$$z = \int_{-\infty}^{\infty} e^{-\theta x} dF(x) \equiv M(\theta).$$

Следовательно, моментная производящая функция для $P(S_{N_t} \leqslant x)$ задана через $P[M(\theta)]$. Математическое ожидание найдем путем дифференцирования по θ, положив $\theta = 0$:

$$E(S_{N_t}) = \frac{dP}{dM}\bigg|_{M=1} (-1) \frac{dM}{d\theta}\bigg|_{\theta=0} = E(N_t) E(X),$$

где X — любая случайная величина из X_i, так как все они имеют одно и то же математическое ожидание.

Вывод формулы для дисперсии предлагаем сделать читателю. Должно получиться выражение

$$\sigma^2(S_{N_t}) = \sigma^2(N_t) E^2(X) + \sigma^2(X) E(N_t).$$

Эти идеи будут использованы в теории восстановления.

3.3. Некоторые важные распределения

В табл. 3.1 дается перечень некоторых распределений и соответствующих им характеристических функций. Для лучшего понимания некоторых из этих распределений следует обратиться к обычным руководствам по теории вероятностей.

Распределение Пирсона III типа является общим типом распределения, которое Пирсон получил при разработке классификации распределений. При частных значениях его параметров получаются известные распределения, относящиеся к этому типу: гамма-распределение, распределение хи-квадрат, эрланговское и экспоненциальное распределения.

Значение хи-квадрат равно

$$\chi^2 = X_1^2 + \ldots + X_k^2, \tag{3.17}$$

где X_i — нормально распределенные взаимно независимые случайные величины с нулевым математическим ожиданием и единичной дисперсией, а k — число степеней свободы распределения.

Эрланговское распределение есть распределение хи-квадрат с $2k$ степенями свободы.

Наличие параметра k в эрланговском распределении, которое, как было показано в начале главы, находится как распределение суммы случайных величин, каждая из которых имеет одинаковое экспоненциальное распределение с математическим ожиданием, равным $\frac{1}{k\mu}$, имеет важное практическое применение. Заметим, что при $k=1$ получаем экспоненциальное распределение, а при

Таблица 3.1

Вид распределения	Выражения для плотности распределения	Характеристическая функция
Равномерное, или прямоугольное	$\frac{1}{b-a}, \; a < x < b$	$\frac{e^{ibt} - e^{iat}}{i(b-a)t}$
Биномиальное . . .	$\frac{n!}{(n-k)!\,k!} q^{n-x} p^{x}, \; p+q=1$	$(q+pe^{it})^{n}$
Гипергеометрическое	$\frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}, \; N_1+N_2=N$	
Пуассоновское . . .	$\frac{\mu^{k} e^{-\mu}}{k!}$	$e^{\mu(e^{it}-1)}$
Нормальное	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(y-\mu)^2}{2\sigma^2}\right]}$	$e^{i\mu t - \frac{\sigma^2 t^2}{2}}$
Логарифмически нормальное	$\frac{1}{\sigma\sqrt{2\pi}} \cdot \frac{1}{y-a} \times$ $\times \exp\left\{-\frac{1}{2\sigma^2}[\log(y-a)-\mu]^2\right\},$ где $a \leqslant y$	
Пирсона III типа . .	$A(y-\mu)^{a-1} e^{-b(y-\mu)}$	
Гамма-распределение	$\frac{b^a y^{a-1} e^{-by}}{\Gamma(a)}$	$\left(1-\frac{it}{b}\right)^{-a}$
Распределение хи-квадрат	$\frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} y^{\left(\frac{n}{2}\right)-1} e^{-\frac{y}{2}}$	
Эрланговское (распределение длительности интервала времени до появления k событий процесса Пуассона)	$\frac{1}{(k-1)!}(k\mu)^{k} y^{k-1} e^{-k\mu y}$	
Экспоненциальное .	$\mu e^{-\mu y}$	
Паскаля, или отрицательное биномиальное	$\binom{n+k-1}{k} p^{k} q^{n}$	$\left(\frac{q}{1-pe^{it}}\right)^{n}$

$k \to \infty$ — вырожденное распределение, т. е. от случайной величины переходим к детерминированной. Многие важные распределения, встречающиеся при решении реальных задач, занимают промежуточное положение между этими двумя случаями и могут быть приближенно описаны как эрланговские при соответствующем выборе значения k. Следовательно, если система массового обслуживания описывается эрланговским распределением, то полученные для нее решения можно применить и в других случаях, когда выбор значения k обеспечивает хорошее совпадение.

Модифицированный эрланговский входящий поток (к одному каналу), в котором k заменено на c, а $k\mu$ — на λ, может рассматриваться как отфильтрованный пуассоновский входящий поток с параметром λ, поступающий к c параллельным каналам, каждый из которых имеет свою очередь. Требование, поступившее первым, становится в очередь к первому каналу, требование, поступившее вторым, — ко второму, c-е требование — к c-му каналу, а затем $(c+1)$-е становится в очередь к первому каналу, а $(c+2)$-е — ко второму и т. д. в таком же порядке. Входящий поток каждого канала и есть данный модифицированный эрланговский поток. Это положение будет встречаться и в дальнейшем.

В качестве упражнения вычислите первый и второй моменты этого распределения. Затем найдите дисперсию σ^2 = (второй начальный момент) — (математическое ожидание)$^2 = \mu_2 - \mu_1^2$, где σ — среднее квадратическое отклонение.

Для многих случаев полезно знать, что распределение выборочного среднего произвольного распределения приближается к нормальному закону в том смысле, который определен центральной предельной теоремой. В частности, если совокупность имеет конечные дисперсию σ^2 и математическое ожидание μ, то по мере увеличения объема выборки n распределение выборочного среднего приближается к нормальному распределению с дисперсией $\frac{\sigma^2}{n}$ и математическим ожиданием μ.

3.4. Вероятностный процесс

Процесс есть изменение состояния системы, подверженной постоянному воздействию случайных факторов. Обычно будут рассматриваться такие системы, которые в моменты времени $t_1, \ldots, t_n$ находятся в состояниях $E_1, E_2, \ldots, E_n$. Для каждого из этих моментов времени существует вероятность того, что система может перейти из состояния, в котором она находится, например, из состояния E_i, в любое другое состояние.

Мы уже указывали, что вероятностный процесс определяется семейством случайных величин $\{X\}$, зависящих от параметра t. Кроме того, уже указывалось, что число параметров может быть больше единицы, однако мы ограничимся одним параметром — временем. Если рассматривается дискретное время, как при бросании монеты, то записываем $X_1, X_2, \ldots$. Каждое значение X_i имеет функцию распределения, которая, в свою очередь, может быть дискретной или непрерывной.

Бросание монеты является примером конечного, или счетного семейства случайных величин, каждая из которых имеет два значения. При бросании монеты возможны два исхода и каждый из них имеет определенную вероятность появления. Например, $P(X_i=1)=p$ — вероятность выпадения надписи, с которой свяжем число 1. Выпадение герба, которое обозначим —1, имеет вероятность $P(X_i=-1)=q$. Должно выполняться условие $p+q=1$, поскольку обязательно появится один из этих исходов.

Заметим, что с течением времени эти вероятности не изменяются. Они будут одни и те же для всех i. Это является иллюстрацией *стационарного процесса,* для которого вероятности не зависят от времени.

Может представить интерес число появлений надписи при достаточно большом числе бросаний n[1]. Если это число разделить на n, а n устремить к бесконечности, то получим

$$\lim_{n\to\infty}\frac{X_1+\ldots+X_n}{n}=\lim_{n\to\infty}\frac{S_n}{n}.$$

Согласно закону больших чисел при увеличении n

$$P\left\{\left|\frac{S_n}{n}-p\right|<\varepsilon\right\}\to 1$$

при произвольном $\varepsilon>0$. Это означает, что при увеличении числа испытаний увеличивается точность приближения вероятности частотой.

Представляет интерес определить, как быстро $\frac{S_n}{n}-p$ стремится к нулю. Ответ на этот вопрос дает закон повторного логарифма:

$$P\left\{\limsup_{n\to\infty}\frac{\sqrt{n}\left(\frac{S_n}{n}\right)}{\sqrt{2pq\log\log n}}=1\right\}=1.$$

Кроме того, из центральной предельной теоремы следует, что при больших n величина $\frac{S_n-np}{\sqrt{npq}}$ имеет распределение, близкое к нормальному.

Число возможных исходов X_t для каждого t не обязательно должно быть конечным, как в приведенном примере, или даже счетным. Может существовать континуум исходов, как при случайном выборе действительных чисел. Если X_t — непрерывная случайная величина, то можно выбрать любое число.

---

[1] J. Doob, What is a Stochastic Process. *Am. Math. Monthly,* 1942, vol. 49, pp. 648—653.

Наконец, время t может рассматриваться как непрерывное; при этом распределение каждого значения может быть непрерывным или дискретным. Примером такого дискретного распределения, когда сам параметр может принимать любое значение, является пуассоновское. Число телефонных абонентов, производящих вызов в любой момент времени, имеет дискретное распределение, которое в любой момент времени t определяет вероятность появления события, заключающегося в том, что поступил вызов от одного абонента, от двух абонентов и т. д. Эти величины определяются экспериментально путем записывания в течение многих дней информации о том, сколько абонентов производит вызов в данном интервале времени, и подсчета частот. Часто допускают, что время принимает значения от $-\infty$ до $+\infty$, хотя это может и не иметь практического смысла, а затем после соответствующей интерпретации результатов получают требуемый ответ.

Допустим теперь, что время t принимает дискретные значения, и в результате каждого испытания может произойти конечное множество исходов, называемых состояниями и обозначаемых E_i $(i=1, \ldots, m)$. (В задаче на бросание монеты $m=2$.) Попытаемся ответить на некоторые важные для этого случая вопросы. Исход каждого испытания не зависит от исходов предыдущих испытаний; следовательно, вероятность появления каждого исхода не зависит от того, что было раньше.

Теперь допустим, следуя Маркову, что вводится некоторый вид ограниченной зависимости: условная вероятность исхода E_j в n-м испытании при условии, что исход $(n-1)$-го испытания был E_i $(i=1, 2, \ldots, m)$, равна p_{ij}; при дополнительной информации об исходах $(n-2)$-го, $(n-3)$-го и т. д. испытаний данная вероятность остается неизменной. Помимо *вероятностей перехода* p_{ij} требуется задание вероятностей $p_i(0)$ $(i=1, \ldots, m)$ получения i-го исхода в первом опыте, т. е. в момент времени $t=0$. Подобная схема приводит к весьма интересным результатам. Заметим, что вероятности p_{ij} описывают переходы $E_i \to E_j$ от одного испытания к другому, следующему непосредственно за ним.

Можно рассмотреть более общую схему, введя вероятности перехода для двух опытов, не следующих непосредственно друг за другом. Это означает, что появление событий в будущем может зависеть от появлений событий в некоторых или во всех опытах. Имеется много реальных физических систем, для которых требуется такое описание. Действительно, могут оказаться необходимыми вероятности, описывающие исходы нескольких испытаний, проводимых одновременно.

В приведенном выше простом случае вероятности начального состояния a_i и вероятности перехода p_{ij} (которые могут быть расположены в виде квадратной матрицы P, называемой стохастической матрицей) определяют цепь Маркова (которая называется конечной, если конечно число исходов). Эта матрица записывается в виде

$$P=\begin{bmatrix} p_{11} & p_{12} & \dots \\ p_{21} & p_{22} & \dots \\ \dots & \dots & \dots \end{bmatrix}.$$

В любом случае сумма элементов каждой строки матрицы равна единице, так как за один шаг мы обязательно попадем в какое-нибудь состояние. Не исключена возможность того, что при переходе от одного испытания к другому может сохраниться то же самое состояние. Если сумма элементов каждого столбца также равна единице, то матрица называется двойной стохастической.

Отметим, что переход из одного состояния в другое в последовательных опытах может происходить не непосредственно, а минуя ряд промежуточных состояний. Число шагов при переходе будем обозначать с помощью верхнего индекса. Имеем

$$p_{ij}^{(1)}=p_{ij},\ p_{ij}^{(2)}=\sum_k p_{ik}p_{kj},\ \dots,\ p_{ij}^{(n)}=\sum_k p_{ik}p_{kj}^{(n-1)}. \tag{3.18}$$

Чтобы получить вероятность перехода из состояния E_i в состояние E_j в момент n-го перехода, совместно с вероятностью перехода p_{ij} рассматривается вероятность того, что в момент $(n-1)$-го перехода система находится в некотором состоянии. Таким образом, в любой момент времени вероятность нахождения системы в определенном состоянии зависит только от знания состояния системы в предыдущий момент времени. Это основное свойство цепей Маркова и марковских процессов.

С помощью матриц вероятностей перехода последнее уравнение можно записать в виде $P^{(n)}=PP^{(n-1)}$. Можно написать и более общее соотношение $P^{(m+n)}=P^{(m)}P^{(n)}$, которому соответствует равенство элементов матриц

$$p_{ij}^{(n+m)}=\sum_k p_{ik}^{(m)}p_{kj}^{(n)}.$$

Вообще представляет интерес вероятность

$$P_j(n)=\sum_{i=0}^{\infty}P_i(0)\,p_{ij}^{(n)}.$$

Это вероятность того, что в момент времени n система находится в состоянии j независимо от того, в каком состоянии система находилась в предыдущий момент времени. Желательно, чтобы при $n\to\infty$ система асимптотически приближалась к стационарному состоянию, т. е. в этом случае слева существует предельный вектор, а справа — предельная переходная матрица, в которой вероятности не зависят от состояния i. Наличие этого свойства ожидается интуитивно. Это свойство эргодичности является важным при длительном протекании процесса, когда вероятности перестают зависеть от времени. Для многих задач массового обслуживания такие результаты имеют бóльшую ценность, чем нестационарные распределения.

Проиллюстрируем эти положения с помощью конечной цепи Маркова с матрицей вероятностей перехода

$$P=\begin{bmatrix} \frac{1}{6} & 0 & \frac{1}{6} & 0 & \frac{2}{3} \\ 0 & \frac{1}{4} & 0 & 0 & \frac{3}{4} \\ \frac{1}{4} & \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} \end{bmatrix}.$$

Множество состояний, в котором для каждой пары состояний существует отличная от нуля вероятность перехода из одного состояния в другое, есть класс эквивалентностей. В данном случае имеется три таких класса. Они объединяют соответственно: 1) первое и третье состояния; 2) второе и пятое состояния и 3) одно четвертое состояние. Например, $p_{25}=3/4$ и $p_{52}=1/3$; следовательно, система может оставаться во втором или в пятом состоянии или переходить из второго состояния в пятое и наоборот. Переход в другие состояния невозможен.

Заметим, что возможен переход из первого класса эквивалентностей во второй и из первого в третий. Если же переход во второй или третий класс произошел, то обратный переход в первый класс невозможен. Поэтому первый класс неустойчив. Он называется *неустановившимся классом* эквивалентностей, и его состояния являются неустойчивыми. Второй и третий классы находятся в равновесии, т. е. если система перешла в один из этих классов, то переход в другой класс невозможен. Эти классы являются замкнутыми. Переход из них в другие классы невозможен, т. е. вероятности p_{ij} равны нулю для значений i, входящих в этот класс, и значений j, не входящих в него. Эти два класса называются *эргодическими классами* эквивалентностей, а их состояния называются эргодическими. Третий класс имеет единственное эргодическое состояние, называемое поглощающим. Цепь, в которой каждое эргодическое состояние является поглощающим, называется *поглощающей цепью Маркова.* В такой цепи возможен переход из любого состояния в поглощающее через один или большее число шагов.

Заметим, что в общем случае подматрица вероятностей перехода между состояниями эргодического класса также является стохастической матрицей, т. е. сумма значений p_{ij} в строке равна единице. Стохастическая матрица называется *регулярной,* если все элементы некоторой ее степени положительны. .

Если подматрица, соответствующая эргодическому классу, регулярна (ее состояния также называются регулярными), то и сам класс называется регулярным. В противном случае эргодический класс эквивалентностей является *периодическим* (с периодическими состояниями).

Если P — стохастическая матрица регулярной цепи, то $\lim\limits_{n\to\infty} P^n = T$ (в гл. 5 будет показано, как находятся функции от матриц), и матрица T будет иметь в каждом столбце одинаковые положительные элементы. Таким образом, все векторы-строки матрицы T одинаковы. Если V — такой вектор-строка, то можно показать, что для любого вектора вероятностей W (вектора, все элементы которого имеют значения от 0 до 1 и в сумме составляют единицу) $\lim\limits_{n\to\infty} WP^n = V$. (Читатель может проверить это положение, воспользовавшись определением V.) Из этого свойства следует, что вероятность пребывания в данном состоянии не зависит от вектора вероятностей начального состояния a_i.

Кроме того, $VP = V$, что получим из определения T путем умножения справа на P. Среднее время, необходимое для возвращения в состояние, в котором система уже находилась ранее, для регулярной цепи Маркова равно обратной величине предельной вероятности пребывания в этом состоянии.

В случае поглощающих состояний представляют интерес вероятность перехода в это состояние, среднее время перехода и средняя длительность пребывания системы в каждом из неустойчивых состояний до перехода в поглощающее состояние. При этом результат может зависеть от вероятностей начальных состояний a_i.

Заметим, что в предыдущем примере в цепи было два замкнутых класса состояний. Если в цепи Маркова нет таких замкнутых классов, за исключением, разумеется, класса всех состояний, то в этом случае цепь называется неприводимой. В противном случае она называется разложимой.

Обозначим через $f_{ij}^{(n)}$ вероятность того, что, находясь в момент времени $t=0$ в состоянии E_i, система в момент времени n в первый раз перейдет в состояние E_j. Тогда

$$f_{ij}^{(n)} = p_{ij}^{(n)} - \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{ij}^{(n-k)},$$

$$f_{ij}^{(1)} = p_{ij}.$$

Вероятность того, что система когда-либо возвратится в состояние E_j, обозначим через

$$F_{ij}=\sum_{n=1}^{\infty} f_{ij}^{(n)}.$$

Таким образом, если E_j — начальное состояние, то F_{ij} — вероятность того, что система когда-либо возвратится в состояние E_j. Математическое ожидание времени первого перехода из состояния E_i в состояние E_j имеет вид

$$\mu_{ij}=\sum_{n=1}^{\infty} n f_{ij}^{(n)}$$

(время между двумя испытаниями принимается за единицу).

Феллер [236] предложил следующую классификацию состояний для цепи Маркова со счетным числом состояний.

Рекуррентное состояние при $F_{ij}=1$ (т. е. система достоверно возвращается в это состояние).

Рекуррентное нулевое состояние при $\mu_{jj}=\infty$.

Переходное состояние при $F_{jj}<1$ (т. е. возвращение в это состояние не является достоверным).

Периодическое состояние с периодом t, если возвращение возможно только через t, $2t$, ... шагов, где $t>1$ — наибольшее целое число, обладающее этим свойством, и, следовательно, $p_{jj}^{(n)}=0$, если n не кратно t.

Рекуррентное ненулевое апериодическое (т. е. не являющееся периодическим) состояние называется эргодическим состоянием.

Конечная цепь Маркова может не содержать нулевых состояний, и невозможно, чтобы все ее состояния были неустойчивыми. В самом деле, если M — число состояний, то по крайней мере для одного j имеем $p_{ij}^{(n)}\geqslant\frac{1}{M}$, так как сумма всех вероятностей, взятая по j, равна единице, а всего имеется M состояний. Следовательно, при $n\to\infty$ невозможно выполнение условия $p_{ij}^{(n)}\to 0$ для всех j.

Феллер приводит теоремы, в которых характеризуются различные виды состояний. Мы приведем некоторые положения из статьи Фостера [264], имеющие отношение к теории массового обслуживания; в последующем эти результаты будут нами использоваться. Имея в виду подобные приложения, мы при формулировке результатов заменим слово «цепь» словом «система».

Пусть требуется найти $\lim\limits_{n\to\infty} p_{ij}^{(n)}$.

Теорема 3.1. Неразложимая апериодическая марковская система является эргодической, если существует ненулевое решение системы уравнений

$$\sum_{i=0}^{\infty} x_i p_{ij}=x_j,\quad j=0,\ 1,\ 2,\ \ldots.$$

Теорема 3.2. Система является эргодической, если неравенства

$$\sum_{j=0}^{\infty} p_{ij} y_j \leqslant y_{i-1}, \quad i \neq 0,$$

имеют неотрицательное решение, причем $\sum_{j=0}^{\infty} p_{0j} y_j < \infty$.

Теорема 3.3. Если система является эргодической, то (конечные) математические ожидания времени d_j, истекшего от j-го до нулевого состояний, удовлетворяют условиям:

$$\sum_{j=1}^{\infty} p_{ij} d_j = d_i - 1$$

и

$$\sum_{j=1}^{\infty} p_{0j} d_j < \infty.$$

Теорема 3.4. Система является неустойчивой в том и только в том случае, если существует ограниченное непостоянное решение уравнений

$$\sum_{j=0}^{\infty} p_{ij} y_j = y_i.$$

Теорема 3.5. Система является рекуррентной, если существует такое решение $\{y_i\}$ неравенств

$$\sum_{j=0}^{\infty} p_{ij} y_j \leqslant y_i, \quad i \neq 0,$$

что при $i \to \infty$ $y_i \to \infty$.

Теорема 3.6. Система является неустойчивой в том и только в том случае, если существует такое ограниченное решение $\{y_i\}$ последних неравенств, что для некоторых i выполняется условие $y_i < y_0$.

Теорема 3.7. Если $\{p_n\}$ $(n=0, 1, \ldots)$ есть распределение вероятностей, причем $p_0 > 0$, то $\sum_{n=0}^{\infty} z^n p^n = z$ имеет корень ξ, значение которого заключено в пределах $0 < \xi < 1$, в том и только в том случае, если $\sum_{n=1}^{\infty} n p_n = 1$.

Пусть k_n — вероятность поступления n новых требований за время обслуживания одного требования в одноканальной системе массового обслуживания $M/G/1$ с пуассоновским входящим потоком с параметром λ и произвольным распределением времени обслуживания $B(t)$. Имеем (см. гл. 7)

$$k_n = \frac{1}{n!} \int_0^{\infty} e^{-\lambda t} (\lambda t)^n \, dB(t), \quad n = 0, 1, \ldots . \tag{3.19}$$

Матрица вероятностей перехода имеет вид

$$[p_{ij}] = \begin{bmatrix} k_0 & k_1 & k_2 & \dots \\ k_0 & k_1 & k_2 & \dots \\ 0 & k_0 & k_1 & \dots \\ 0 & 0 & k_0 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}.$$

Теорема 3.8. Система $M/G/1$ является эргодической в том и только в том случае, если выполняется условие

$$\rho \equiv \sum_{n=1}^{\infty} n k_n < 1.$$

Она будет рекуррентной в том и только в том случае, если $\rho \leqslant 1$.

Доказательство. Пусть $\rho < 1$; введем обозначение $y_j = \frac{j}{1-\rho}$; тогда y_j будет удовлетворять теореме 3.2.

С другой стороны, если система — эргодическая, то из матрицы вероятностей перехода видно, что μ_{ij} — математическое ожидание времени первого перехода из состояния E_i в состояние E_j — удовлетворяет условию

$$\mu_{i,\,i-1} = \mu_{10}, \quad i \neq 0.$$

Переход в состояние E_0 может произойти только из состояния E_i через состояние E_{i-1}. Следовательно,

$$\mu_{i0} = \mu_{i,\,i-1} + \mu_{i-1,\,0},$$

и по индукции имеем

$$\mu_{i0} = i\mu_{10}.$$

Воспользовавшись теоремой 3.3, получим

$$\mu_{10} \sum_{j=1}^{\infty} j p_{ij} = \mu_{10} - 1.$$

Поэтому

$$\rho\mu_{10} = \mu_{10} - 1$$

и

$$\rho = 1 - \frac{1}{\mu_{10}} < 1.$$

Пусть $\rho \leqslant 1$. Примем $y_j = j$; тогда $y_j \to \infty$ при $j \to \infty$ и условие теоремы 3.5 выполняется, следовательно, система является рекуррентной. С другой стороны, если $\rho > 1$, то к выражению $\sum_{n=0}^{\infty} k_n z^n = z$ применим теорему 3.7 и примем $y_j = \xi^j$. Теперь выполняется условие теоремы 3.4, $y_j \to 0$ при $j \to 0$ и $y_0 = 1$, следовательно, система является неустойчивой.

Заметим, что $\lim_{n \to \infty} p_{ij}^{(n)} = 0$ при $\rho \geqslant 1$ для всех i и j. Система является недиссипативной, если сумма строк матрицы предельных вероятностей равна единице. В противном случае она является диссипативной.

Для системы $GI/M/1$ также можно записать матрицу вероятностей перехода и доказать аналогичную теорему. Для применения изложенных выше положений процесс массового обслуживания должен быть выражен через цепь Маркова, имеющую вероятности перехода p_{ij}. Процессы массового обслуживания с непрерывным временем являются более трудными для исследования.

Если вероятности перехода p_{ij} становятся зависимыми от непрерывного параметра (времени t), то имеем марковский процесс, который является примером вероятностного процесса с непрерывным временем. Другими примерами вероятностного процесса является пуассоновский процесс и процесс Эрланга.

Вероятность перехода системы в момент времени t в состояние j при условии, что в момент времени s система находилась в состоянии i, обозначается через $p(j, t; i, s)$. Конечно, существуют и такие системы (например, в статистической механике), в которых информация обо всех предыдущих состояниях влияет на исход нового состояния.

Одномерный марковский процесс описывается уравнениями Чепмена—Колмогорова, которые являются обобщением уравнений, описывающих цепи Маркова.

$$p(j, t; i, s) = \sum_{k=0}^{\infty} p(j, t; k, u)\, p(k, u; i, s), \tag{3.20}$$

при $s < u < t$ и $i, j = 0, 1, 2, \ldots$. Из этого уравнения выводятся две системы уравнений:

1. Прямые уравнения, которые получим, взяв производную по t и заменив затем u на t.

2. Обратные уравнения, которые получим, взяв производную по s и заменив затем u на s.

Так, например, если переходы в состояние i в момент времени t могут осуществиться только из двух соседних состояний $i-1$ и $i+1$, как в процессе рождения и гибели, то обратные уравнения имеют вид

$$\begin{aligned} p_s(j, t; i, s) = p(j, t; i, s)\, p_s(i, s; i, s) + \\ + p(j, t; i-1, s)\, p_s(i-1, s; i, s) + \\ + p(j, t; i+1, s)\, p_s(i+1, s; i, s), \end{aligned} \tag{3.21}$$

где p_s — производная от p по s.

Если же возможны только переходы вида $E_i \to E_{i+1}$, как в случае пуассоновского процесса, то имеет место чистый процесс рождения (если же только $E_i \to E_{i+1}$, то имеет место чистый процесс гибели).

При выводе прямых уравнений принимаются во внимание пути достижения данного состояния; для обратных уравнений рассматриваются пути выхода из данного состояния. В общем случае решение одной из этих двух систем при одинаковых начальных условиях является также решением и для другой системы. Иногда легче решить систему обратных уравнений, чем систему прямых уравнений.

Пример 5. Покажем, что пуассоновский процесс, который является марковским процессом, удовлетворяет уравнению Чепмена—Колмогорова. Используя соответствующую форму записи пуассоновского процесса:

$$p(j,\ t;\ k,\ u) = e^{-\lambda(t-u)}\frac{[\lambda(t-u)]^{j-k}}{(j-k)!},$$

$$p(k,\ u;\ i,\ s) = e^{-\lambda(u-s)}\frac{[\lambda(u-s)]^{k-i}}{(k-i)!},$$

и производя подстановку в уравнение Чепмена—Колмогорова' получим

$$e^{-\lambda(t-s)}\sum_{k=i}^{j}\frac{[\lambda(t-u)]^{j-k}}{(j-k)!}\,\frac{[\lambda(u-s)]^{k-i}}{(k-i)!} =$$

$$= e^{-\lambda(t-s)}[\lambda(t-u)]^{j}[\lambda(u-s)]^{-i}\sum_{i}^{j}\left(\frac{u-s}{t-u}\right)^{k}\frac{1}{(j-k)!\,(k-i)!}.$$

Пусть $m = k - i$, тогда $k = m + i$ и

$$\sum_{m=0}^{j-i}\left(\frac{u-s}{t-u}\right)^{m+i}\frac{1}{(j-m-i)!\,m!} =$$

$$= \frac{\left[\frac{(u-s)}{(t-u)}\right]^{i}}{(j-i)!}\sum_{m=0}^{j-i}\left(\frac{u-s}{t-u}\right)^{m}\frac{(j-i)!}{(j-i-m)!\,m!} =$$

$$= \frac{1}{(j-i)!}\left(\frac{u-s}{t-u}\right)^{i}\left(1+\frac{u-s}{t-u}\right)^{j-i} =$$

$$= \frac{1}{(j-i)!}\left(\frac{u-s}{t-u}\right)^{i}\left(\frac{t-s}{t-u}\right)^{j-i}.$$

Первое выражение принимает вид

$$\frac{e^{-\lambda(t-s)}}{(j-i)!}\,\frac{[\lambda(t-u)]^{j}}{[\lambda(u-s)]^{i}}\left(\frac{u-s}{t-u}\right)^{i}\left(\frac{t-s}{t-u}\right)^{j-i} =$$

$$= \frac{e^{-\lambda(t-s)}}{(j-i)!}[\lambda(t-s)]^{j-i}, \qquad (3.22)$$

что и требовалось доказать.

Наконец, заметим, что пуассоновский процесс является однородным, он зависит только от длительности промежутка времени и не зависит от того, где находится начало отсчета (т. е. он не зависит от положения на временно́й оси). Таким образом, пуассоновский процесс есть однородный марковский процесс рождения.

Поскольку обычно будут рассматриваться начальные условия (т. е. начальное состояние E_i) в момент времени $t=0$, то во избежание громоздких обозначений запишем вероятность того, что в момент времени t система находится в состоянии E_j при условии, что в момент времени $t=0$ она находилась в состоянии E_i, в виде $P_{ij}(t)$. Существует такая матрица вероятностей перехода $P(t)=[P_{ij}(t)]$, что

$$0 \leqslant P_{ij}(t), \quad \sum_{j=0}^{\infty} P_{ij}(t)=1, \quad \lim_{t\to 0} P_{ij}(t)=\delta_{ij},$$

$$P_{ij}(s+t)=\sum_{k=0}^{\infty} P_{ik}(s)\,P_{kj}(t), \quad 0<s<\infty, \ 0<t<\infty.$$

Матрица, для которой выполняется неравенство $\sum_{0}^{\infty} P_{ij}(t)<1$, называется вырожденной.

Если обозначить через $P_i(0)$ вероятность того, что система в начальный момент времени находилась в состоянии i, то $\sum_{i=0}^{\infty} P_i(0)=1$. Вероятность того, что через промежуток времени t система будет находиться в состоянии j, равна

$$P_j(t)=\sum_{i=0}^{\infty} P_{ij}(t)\,P_i(0), \quad j=0, 1, \ldots . \tag{3.23}$$

При этих обозначениях коэффициенты матрицы $P(t)$ удовлетворяют прямому уравнению Чепмена—Колмогорова

$$\frac{dP_{ij}(t)}{dt}=\sum_{k=0}^{\infty} P_{ik}(t)\,a_{kj}, \quad i, j=0, 1, 2, \ldots , \tag{3.24}$$

где $A \equiv (a_{ij})$ — такая матрица постоянных вероятностей перехода, что

$$a_{ij} \geqslant 0, \quad i \neq j; \quad a_{ii} \leqslant 0, \quad \sum_{j=0}^{\infty} a_{ij} \leqslant 0$$

($a_{ij}dt$ — вероятность перехода из состояния E_i в состояние E_j за время dt, и, следовательно, A — инфинитезимальная матрица вероятностей перехода).

Обратное уравнение найдем, транспонировав матрицу A. В матричной записи эти уравнения имеют вид $P'(t)=P(t)A$ и $P'(t)= =AP(t)$ при $P(0)=I$. Если справедливы оба уравнения, то $P'_{ij}(0)=a_{ij}$.

Хилл [352] и Ройтер [719] исследовали решение этой задачи, применяя теорию полугрупп линейных ограниченных операторов.

Для процесса рождения и гибели матрица A приобретает особый вид:

$$a_{i,\,i+1}=\lambda_i,\quad a_{ii}=-(\lambda_i+\mu_i),\; a_{i,\,i-1}=\mu_i, \tag{3.25}$$

и $a_{ij}=0$, если $|i-j|>1$, где $\lambda_i>0$ при $i\geqslant 0$, $\mu_i>0$ при $i\geqslant 1$ и $\mu_0=0$.

Более подробное рассмотрение процесса рождения и гибели с помощью прямых и обратных уравнений будет дано в следующей главе.

В заключение заметим, что марковский процесс имеет простую структуру вероятностей, которые легко описываются, так как будущее состояние системы определяется только ее настоящим состоянием. Информация о ее состоянии в прошлом не оказывает влияния на ее состояние в будущем. Именно это свойство и определяет значение марковских процессов для теории массового обслуживания.

Задачи

1. Используя моментные производящие функции и свертку, покажите, что плотность распределения суммы взаимно независимых случайных величин

$$S_n=X_1+\ldots+X_n,$$

распределенных по одному и тому же закону $\mu e^{-\mu t}$, имеет вид

$$\frac{\mu(\mu t)^{n-1}e^{-\mu t}}{(n-1)!}\quad \text{при } t\geqslant 0.$$

2. Решите задачу 1 для двух случайных величин, имеющих плотность того же вида, но различные параметры μ_1 и μ_2.

3. Покажите, что преобразование Лапласа свертки двух функций равно произведению преобразований этих функций. Выполните это для двух последовательностей.

4. Если принять, что время обслуживания имеет экспоненциальное распределение, то вероятность того, что длительность обслуживания будет заключена в интервале $(t,\,t+dt)$, равна $\mu e^{-\mu t}dt$. Допустим, что имеется c каналов с различным экспоненциальным временем обслуживания, каждый с соответствующим параметром μ_n, и бесконечное множество требований, ожидающих обслуживания. Каждый раз, когда требование поступает в систему, оно направляется в один из каналов, а все остальные каналы бездействуют, т. е. требования обслуживаются поодиночке и только одним из каналов. Пусть a_n — вероятность того, что в момент поступления требования в систему оно направляется в n-й канал. Покажите, что плотность распределения промежутков времени между двумя последовательными требованиями, поступающими на обслуживание, равна

$$\sum_{n=1}^{c}a_n\mu_n e^{-\mu_n t},\quad \sum_{n=1}^{c}a_n=1.$$

Это выражение для гиперэкспоненциального распределения. Покажите, что его математическое ожидание равно

$$m=\sum_{n=1}^{c}\frac{a_n}{\mu_n},$$

а дисперсия —

$$\sigma^2 = 2\sum_{n=1}^{c} \frac{a_n}{\mu_n^2} - \left(\sum_{n=1}^{c} \frac{a_n}{\mu_n}\right)^2.$$

Заметим, что всегда выполняется условие $\frac{\sigma}{m} \geqslant 1$ и что для фиксированных m можно найти такие отличные от нуля значения a_n и μ_n ($n=1, \ldots, c$), что среднее квадратическое отклонение σ может быть сделано сколь угодно большим.

Пальм разделил распределения на два класса. Вместо функции распределения $F(t)$ он рассматривал функцию $1-F(t)$. Введем величину

$$\varepsilon \equiv \frac{2}{m^2}\int_0^\infty t[1-F(t)]\,dt = 1+\frac{\sigma^2}{m^2},$$

где m — математическое ожидание, вычисленное обычным способом для $1-F(t)$; σ^2 — дисперсия.

Всегда $\varepsilon \leqslant 1$. Если $1 \leqslant \varepsilon < 2$, то кривая распределения будет крутой, а при $\varepsilon > 2$ она будет пологой. Экспоненциальное распределение лежит на границе $\varepsilon=2$. Примером крутого распределения является вырожденное распределение времени обслуживания при $\varepsilon=1$. Можно показать, что при помощи гиперэкспоненциального распределения можно приближенно представить любую абсолютно монотонную функцию. Такие функции образуют важный подкласс пологих распределений. Для абсолютно монотонной функции $f(t)$ знак ее p-й производной определяется из выражения

$$(-1)^p f^{(p)}(t) \geqslant 0.$$

Сама функция имеет вид

$$\int_0^\infty e^{-xt}\,dH(x),$$

где $H(x)$ — ограниченная неубывающая функция при $x \geqslant 0$, $H(0)=0$. Действительно, любая функция такого вида абсолютно монотонна. Если такая функция описывает распределение времени обслуживания, то математическое ожидание оставшегося времени обслуживания, задаваемое выражением

$$\int_0^\infty \frac{f(x+t)\,dx}{f(t)},$$

увеличивается с течением времени. То же выражение убывает при увеличении времени, если $f(t)$ — крутое распределение. Приведите примеры для каждого случая.

5. Рассмотрите логарифм функции правдоподобия для выборки из n наблюдений $(x_1, \ldots, x_n)$, взятой из эрланговской совокупности:

$$\log \prod_{i=1}^{n} \frac{\lambda^k}{(k-1)!} x_i^{k-1} e^{-\lambda x_i} = nk \log \lambda -$$

$$- n \log \Gamma(k) - \lambda \sum_1^n x_i + (k-1)\sum_1^n \log x_i.$$

Найдите оценку наибольшего правдоподобия для λ и k. Оценку для k можно оставить в неявном виде.

6. Используя статью Фостера [264], покажите, что система $GI/M/1$ будет эргодической в том и только в том случае, если $\rho < 1$.

7. Напишите матрицу вероятностей перехода для четырех состояний с двумя классами эквивалентностей, один из которых является эргодическим.

8. Приведите пример разложимой цепи Маркова.

9. Найдите математическое ожидание и дисперсию распределений, приведенных в таблице.

10. При каком соотношении параметров распределения Пирсона III типа получается гамма-распределение, эрланговское и экспоненциальное распределения.

11. Приведите пример немарковского процесса.

12. Допустим, что прибытие свободных такси на стоянку образует пуассоновский процесс с параметром a. Предположим также, что прибытие пассажиров на стоянку такси соответствует другому пуассоновскому процессу с параметром b. Пусть в момент времени $t=0$ нет пассажиров, ожидающих такси, и нет машин, а в момент времени t имеется X пассажиров и Y такси. Пусть $Z=X-Y$ определяет длину очереди пассажиров. Покажите, что среднее значение Z равно $(b-a)t$, а дисперсия равна $(a+b)t$. Вычислите характеристическую функцию для Z и покажите, что все семиинварианты четных порядков равны $(a+b)t$, а семиинварианты нечетных порядков равны $(b-a)t$.

Если $b > a$, то, рассматривая среднее квадратическое отклонение случайной величины Z, покажите, что с течением времени число пассажиров будет неограниченно возрастать с вероятностью, приближающейся к единице. Рассмотрите также случай $a=b$. При $t \to \infty$ ни в одном из этих случаев устойчивое состояние не достигается.

13. В одноканальной системе с произвольным входящим потоком и произвольным временем обслуживания [157] интенсивность поступления требований равна λ, а интенсивность их обслуживания равна μ. Загрузка системы равна $\rho = \frac{\lambda}{\mu}$. Рассмотрим неустановившееся состояние системы, когда $\rho > 1$. Допустим, что процесс начинается в момент времени $t=0$, когда в очереди находится i требований. За время t в среднем будет обслужено μt требований. Однако за это время поступит еще λt требований. Всего за время t очередь увеличится на $\lambda t - \mu t = \mu t(\rho - 1)$ требований. Таким образом, среднее число требований, ожидающих в очереди в момент времени t, окажется равным $i + \mu t(\rho - 1)$, что совпадает с результатом, полученным для детерминированной системы массового обслуживания, рассмотренной в гл. 2 [1]. Покажите, что дисперсия распределения числа требований, находящихся в очереди в момент времени t, равна дисперсии числа требований, поступивших за время t [т. е. $\sigma_a^2(t)$], плюс дисперсия времени обслуживания (т. е. $\mu t C^2$, где C^2 — коэффициент вариации распределения времени обслуживания).

Если начальное число требований распределено по определенному закону, то в выражении для среднего числа требований, находящихся в очереди в момент времени t, значение i заменяется математическим ожиданием этого закона, а его дисперсия добавляется к дисперсии, вычисленной выше. Пусть $w(t)$ — плотность распределения времени ожидания в очереди в момент времени t, а $L_q(t)$ — плотность распределения числа требований, находящихся в очереди также в момент времени t. Покажите, что

$$E\,[w(t) \mid L_q(t)] \approx \frac{L_q(t)}{\mu},$$

$$\sigma^2\,[w(t) \mid L_q(t)] \approx \frac{L_q(t)\,C^2}{\mu^2}.$$

---

[1] Этот результат является приближенным. Его относительная погрешность стремится к нулю при $t \to \infty$. — *Прим. ред.*

Здесь вертикальная черта заменяет слова «при условии, что». Покажите, что

$$W_q = E[w(t)] \approx \frac{i}{\mu} + t(\rho - 1),$$

$$\sigma^2[w(t)] = E[L_q(t)]\,\sigma^2[w(t) \mid L_q(t)] + \sigma^2[L_q(t)]\,E[w(t) \mid L_q(t)] \approx$$

$$\approx \frac{iC^2}{\mu^2} + \frac{t}{\mu}\rho C^2 + \frac{\sigma_\alpha^2(t)}{\mu^2}.$$

Если входящие требования равномерно распределены в интервале $[0, T]$ с плотностью λ, то общее время ожидания в очереди всех требований, поступивших в этом интервале, приближенно равно

$$\lambda \int_0^T E[w(t)]\,dt = i\rho T + \frac{\rho(\rho-1)\mu T^2}{2}.$$

14. Случайная величина называется безгранично делимой, если она может быть представлена как сумма n независимых одинаково распределенных случайных величин, где n — любое натуральное число (см. Б. В. Гнеденко и А. Н. Колмогоров. Предельные распределения сумм независимых случайных переменных. Гостехиздат, М.—Л., 1949).

Функция распределения является безгранично делимой в том и только в том случае, если ее характеристическая функция равна n-й степени некоторой характеристической функции (зависящей от n) для любого натурального числа n. Последняя может быть однозначно представлена как корень n-й степени из характеристической функции безгранично делимой функции распределения, если учесть условия непрерывности и равенства 1 в нуле.

Покажите, что нормальный закон, закон Пуассона и закон гамма-распределения являются безгранично делимыми. Кроме того, проверьте на примерах этих распределений, что характеристическая функция безгранично делимой функции распределения нигде не обращается в нуль.

Покажите, что функция распределения суммы конечного числа независимых безгранично делимых случайных величин безгранично делима.

Часть II

МАРКОВСКИЕ МОДЕЛИ МАССОВОГО ОБСЛУЖИВАНИЯ

В двух главах второй части исследуются одноканальные и многоканальные системы массового обслуживания с пуассоновским входящим потоком и экспоненциальным временем обслуживания при обслуживании требований в порядке поступления. В гл. 4 рассматриваются случаи, когда параметры системы зависят от времени и числа требований, находящихся в очереди. В гл. 5 основное внимание уделяется применению матриц и операторных методов в теории массового обслуживания. Изложение начинается с марковских моделей вследствие их особого значения. Изучение подобных систем обеспечивает хорошее ознакомление с задачами массового обслуживания и с методами их формализации и решения. В тексте обычно даются общие решения для стационарного случая, а большое число частных случаев вынесено в задачи, помещенные в конце гл. 4.

Глава 4

ПРОЦЕСС РОЖДЕНИЯ И ГИБЕЛИ В ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

4.1. Введение

Изучение марковских моделей массового обслуживания (поступление требований в систему и окончание обслуживания имеют марковский характер) началось с телефонных линий. Поток вызовов есть пуассоновский процесс, а экспоненциальное распределение продолжительности телефонных разговоров подтверждается экспериментально. Вероятность поступления за малый промежуток времени более одного требования мала. Это условие наряду со свойством «отсутствия памяти» экспоненциального распределения находит широкое применение. Можно считать, что эти допущения справедливы также для процесса прибытия судов в большой порт и для процессов их разгрузки. Марковские модели массового обслуживания начали изучаться раньше других. Так как теория этих систем разработана достаточно полно, то исследование начнем с уравнений процесса рождения и гибели, в которых содержатся допущения, сделанные для пуассоновского потока. Мы уже видели в гл. 3, что уравнения процесса рождения и гибели составляются для вероятностей того, что в момент времени t в системе находится то или иное число требований. С помощью этих вероятностей путем дальнейших выкладок можно найти распределение времени ожидания.

Уравнения процесса рождения и гибели играют важную роль в теории массового обслуживания. Частным случаем этого процесса является процесс, соответствующий модели Эрланга. Устанавливая определенным образом значения коэффициентов, получим описание соответствующих систем массового обслуживания с пуассоновским входящим потоком и экспоненциальным временем обслуживания.

В этом параграфе будут приведены лишь основные уравнения. В следующих параграфах будут решены некоторые задачи массового обслуживания, первой из которых является задача для одноканальной системы. Результаты, полученные для стационарного состояния, будут сведены в таблицу.

1. Уравнения процесса рождения и гибели

«Прямые» уравнения процесса рождения и гибели имеют вид

$$P'_{in}(t)=\frac{dP_{in}(t)}{dt}=-(\lambda_n+\mu_n)P_{in}(t)+\lambda_{n-1}P_{i,\,n-1}(t)+\mu_{n+1}P_{i,\,n+1}(t),$$
$$t\geqslant 0,\quad n>0,\quad i=0,1,2,\ldots,$$
$$P'_{i0}(t)=\frac{dP_{i0}(t)}{dt}=-\lambda_0 P_{i0}(t)-\mu_1 P_{i1}(t),$$
$$n=0,\quad i=0,\,1,\,2,\,\ldots. \tag{4.1}$$

Здесь $P_{in}(t)$ — вероятность того, что в момент времени t система перейдет из состояния E_i, в котором она находилась в момент времени $t=0$, в состояние E_n.

Заметим, что

$$P_{in}(0)=\delta_{in}=\begin{cases}0, & i\neq n,\\ 1, & i=n,\end{cases}$$

где δ_{in} — символ Кронекера.

Применительно к теории массового обслуживания говорят, что система находится в состоянии E_n, если в данный момент времени в ней находится n требований. Вероятность перехода из состояния E_n в состояние E_{n+1} в течение временно́го интервала $(t,\ t+dt)$ равна $\lambda_n dt+o(dt)$, а вероятность перехода из состояния E_n в состояние E_{n-1} за такой же промежуток времени равна $\mu_n dt+o(dt)$. Существует связь между этими допущениями и теми, которые были сделаны для пуассоновского процесса и экспоненциального распределения. При $\mu_n=0$, $\lambda_n=\lambda$ уравнения (4.1) становятся уравнениями пуассоновского процесса. При $\lambda_n=\lambda$ и $\mu_n=\mu$ эти уравнения описывают одноканальную систему массового обслуживания.

С прямыми уравнениями связаны «обратные» уравнения:

$$P'_{in}(t)=-(\lambda_i-\mu_i)P_{in}(t)+\lambda_i P_{i+1,\,n}(t)+\mu_i P_{i-1,\,n}(t),$$
$$n>0,\quad t\geqslant 0, \tag{4.2}$$
$$P'_{0n}(t)=-\lambda_0 P_{0n}(t)+\lambda_0 P_{1,\,n}(t).$$

Решение прямых уравнений (удовлетворяющее заданным начальным условиям) является также решением и для обратных уравнений. Существует возможность сделать выбор, какую из этих двух систем следует решать.

На данном этапе целесообразно вновь ввести понятие о поглощающих и отражающих барьерах. Если рассмотреть, например, прямые уравнения, то мы обнаружим, что при $\lambda_0=0$ невозможен переход из состояния E_0 в состояние E_1, т. е. если система перешла в состояние E_0, то она навсегда остается в этом состоянии; следовательно, E_0 — *поглощающее* состояние. С другой стороны, если $\lambda_0 \neq 0$, то E_0 — *отражающее* состояние, и переход в состояние E_1 возможен. Число состояний может быть конечным. В этом случае последнее состояние E_N может быть поглощающим или отражающим. В любом случае не обязательно начинать с состояния E_0. Любое другое состояние может быть нижним поглощающим или нижним отражающим состоянием системы.

Представляет интерес рассмотреть вопросы существования решения задачи и его единственности. Мы кратко изложим некоторые относящиеся сюда результаты, приведенные в замечательной статье Ройтера [719]; этот автор исследовал проблему существования единственного решения уравнений процесса рождения и гибели.

Если $\sum_{k=0}^{\infty} a_i k = 0$, то матрица A, рассмотренная в гл. 3, называется консервативной. Рекуррентно определим

$$f_{ij}^0(t) \equiv 0,$$

$$f_{ij}^{n+1}(t) \equiv \delta_{ij} e^{-a_i t} + e^{-a_i t} \int_0^t \left[\sum_{k \neq i} a_{ik} f_{kj}^n(u) \right] e^{a_i u} du,$$

где $a_i \equiv -a_{ii} (\geqslant 0)$.

При увеличении n функция f_{ij}^n возрастает и стремится к конечному пределу

$$f_{ij}(t) \equiv \lim_{n \to \infty} f_{ij}^n(t).$$

Кроме того, удовлетворяется следующее соотношение:

$$f_{ij}^{n+1}(t) = \delta_{ij} e^{-a_j t} + e^{-a_j t} \int_0^t \left[\sum_{k \neq j} f_{ik}^n(u) a_{kj} \right] e^{a_j u} du.$$

Предельные функции f_{ij} определяют такой процесс $F = \{f_{ij}(t)\}$, что $f'_{ij}(0) = a_{ij}$. Как прямые, так и обратные уравнения справедливы для процесса F. Любой процесс $P_{ij}(t)$ при $P'_{ij}(0) = a_{ij}$ называется A-процессом. Процесс F естественно назвать минимальным A-процессом, так как можно показать, что для любого A-процесса $f_{ij}(t) \leqslant P_{ij}(t)$. A-процесс называется собственным, если $\sum_k P_{ik} = 1$; в противном случае он называется несобственным. Минимальный A-процесс F всегда обеспечивает решение как прямых, так и обратных уравнений. В статье Ройтера приведены необходимые и достаточные условия для того чтобы процесс F являлся единственным решением этих двух систем уравнений. Если матрица A консервативна, то любой A-процесс, в частности каждое решение прямых уравнений, есть решение и обратных уравнений.

Выберем коэффициенты a_{ij} так, как было указано в (3.25) при получении уравнений процесса рождения и гибели.

Теорема 4.1. Если матрица A с коэффициентами, определенными для процесса рождения и гибели, консервативна и если

$$R=\sum_{n=1}^{\infty}\left(\frac{1}{\lambda_n}+\frac{\mu_n}{\lambda_n\lambda_{n-1}}+\ldots+\frac{\mu_n\ldots\mu_2}{\lambda_n\ldots\lambda_2\lambda_1}\right),$$

$$S=\sum_{n=1}^{\infty}\left(\frac{1}{\mu_{n+1}}+\frac{\lambda_n}{\mu_{n+1}\mu_n}+\ldots+\frac{\lambda_n\ldots\lambda_1}{\mu_{n+1}\ldots\mu_2\mu_1}\right),$$

$$T=\sum_{n=1}^{\infty}\left(\frac{\mu_n\ldots\mu_2}{\lambda_n\ldots\lambda_2\lambda_1}+\frac{\lambda_n\ldots\lambda_1}{\mu_{n+1}\ldots\mu_2\mu_1}\right),$$

то:

1) при $R=\infty$ существует только один A-процесс; этот процесс является собственным и удовлетворяет прямым уравнениям;

2) при $R<\infty$ и $S=\infty$ существует бесконечное множество A-процессов; только один из них удовлетворяет прямым уравнениям, но он является несобственным;

3) при $R<\infty$ и $S<\infty$, или, что то же самое, при $T<\infty$, существует бесконечное множество процессов, удовлетворяющих прямым уравнениям. Из них только один является собственным.

Легко показать, что для системы массового обслуживания с c обслуживающими устройствами справедливо первое условие (см. ниже). Заметим, что $\lambda_0=0$ и матрица A консервативна. Каждый член ряда, определяющего R, не меньше соответствующего члена расходящегося ряда

$$\sum_{n=1}^{\infty}\frac{1}{\lambda_n}=\frac{1}{\lambda}\left(1+\frac{1}{2}+\ldots+\frac{1}{n}+\ldots+\frac{1}{c-1}+\frac{1}{c}+\frac{1}{c+1}+\ldots\right),$$

следовательно, $R=\infty$. Таким образом, существует единственное решение задачи, которое является собственным и удовлетворяет прямым уравнениям. (Для эрланговской модели $c=1$.)

Карлин и Мак-Грегор [410, 411] также получили необходимое и достаточное условие того, что уравнения процесса рождения и гибели имеют единственное решение $P_{in}(t)\geqslant 0$, удовлетворяющее начальным условиям и соотношению

$$\sum_{n=0}^{\infty}P_{in}(t)\leqslant 1.$$

Это условие выражается в виде

$$\sum_{n=0}^{\infty}\left(\pi_n+\frac{1}{\lambda_n\pi_n}\right)=\infty,$$

где

$$\pi_0=1, \quad \pi_n=\frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}, \quad n\geqslant 1.$$

Замечание. Имеем интегральное представление

$$P_{in}(t)=\pi_n\int\limits_0^\infty e^{-xt}R_i(x)\,R_n(x)\,d\psi(x).$$

Здесь

$$R_0(x)=1;$$

$$-xR_0(x)=-(\lambda_0+\mu_0)\,R_0(x)+\lambda_0R_1(x);$$

$$-xR_n(x)=\mu_nR_{n-1}(x)-(\lambda_n+\mu_n)\,R_n(x)+\lambda_nR_{n+1}(x), \quad n\geqslant 1,$$

а ψ — положительная регулярная мера на интервале $0\leqslant x<\infty$, удовлетворяющая соотношению

$$\int\limits_0^\infty R_i(x)\,R_n(x)\,d\psi(x)=\frac{\delta_{in}}{\pi_n}, \quad i,\ n=0,\ 1,\ 2,\ \ldots.$$

Подобная мера ψ называется решением проблемы моментов. Если

$$\rho=\frac{1}{\sum\limits_{n=0}^\infty\pi_n},$$

то, поскольку $R_n(0)=1$,

$$\lim_{n\to\infty}P_{in}(t)=\rho\pi_n\equiv p_n.$$

Здесь ρ — мера ψ точки $x=0$. Постоянная ρ принимается равной нулю, если ряд расходится. В данном случае имеем классическую эргодическую теорему, характеризующую предельное поведение вероятности $P_{in}(t)$ при $t\to\infty$, исходя из интегральных представлений. При соответствующей классификации состояний было показано, что система является эргодической в том и только в том случае, если $\sum\pi_n<\infty$, $\sum\frac{1}{\lambda_n\pi_n}=\infty$. Аналогичным образом выписываются и условия принадлежности к другим классам.

Во многих случаях вместо способа, основанного на классификации состояний, используется методика, аналогичная той, которая применяется в статистической механике. Для условия равновесия, если допускается его существование, необходимо, чтобы производная $\frac{dP_{in}(t)}{dt}$ равнялась нулю (т. е. чтобы вероятности не изменялись с течением времени и, следовательно, не зависели от времени). В этом случае можно найти решение системы дифференциальных уравнений для стационарного состояния.

Таким образом, вероятности стационарного состояния (если выполняются условия существования последнего) находятся путем решения этих уравнений при равных нулю производных по времени или путем использования идеи приведенного выше доказательства. Имеем

$$p_n = p_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i},$$

где p_0 определяется из условия $\sum_{n=0}^{\infty} p_n = 1$.

В качестве упражнения для подготовки к решению задач, приведенных в конце этой главы, читателю предлагается рекуррентным способом получить решение для стационарного состояния.

Вспомните полученное в гл. 2 решение задачи для частного случая:

$$p_n = p_0 \left(\frac{\lambda}{\mu}\right)^n, \quad p_0 = 1 - \frac{\lambda}{\mu}.$$

Перейдем теперь к решению задачи для переходного состояния и к другим вопросам теории массового обслуживания. При этом, чтобы показать возможные пути решения таких задач, будут использованы различные методы. Сущность процесса рождения и гибели такова, что положения этой главы применимы к системам массового обслуживания с пуассоновским входящим потоком и экспоненциальным временем обслуживания.

Нередко при анализе систем мы будем иметь дело с дифференциальными уравнениями, уравнениями в конечных разностях, преобразованиями Лапласа и теорией функций комплексного переменного. Важную роль в теории функций комплексного переменного играет теорема Руше, которая будет применяться довольно часто.

Теорема 4.2. Если две функции $f(z)$ и $g(z)$ аналитические в замкнутой области, ограниченной контуром C и удовлетворяют на контуре C условию $|g(z)| < |f(g)|$, то внутри C функции $f(z)$ и $f(z)+g(z)$ имеют одинаковое число нулей.

Доказательство этой теоремы можно найти в известной работе Титчмарша, посвященной этому вопросу [836][1].

4. 2. Случай 1. Одноканальная система (модель Эрланга); $\lambda_n = \lambda$, $\mu_n = \mu$.

Такой выбор параметров приводит к уравнениям, описывающим достаточно знакомую читателю систему массового обслуживания с одним обслуживающим устройством.

---

[1] См. Титчмарш Э. Теория функций. Гостехиздат, М.—Л., 1951, стр. 137—138. — *Прим. переводчика.*

С тех пор как в 1953 г. Кларк получил решение для переходного процесса, было предложено еще несколько способов решения этой задачи. Здесь будет изложен простой способ последовательного нахождения решения, основанный на усовершенствовании некоторых существующих методов.

До получения в явном виде преобразований Лапласа для производящей функции наш способ во многом следует методике Бейли.

1. Уравнения

Для упрощения обозначений опустим индекс i, примем $\lambda_n=\lambda$ и $\mu_n=\mu$. Получим уравнения для одноканальной системы с пуассоновским входящим потоком и экспоненциальным временем обслуживания для вероятностей того, что в момент времени t в системе будет n требований при условии, что в момент времени $t=0$ в системе было i требований:

$$\frac{dP_n}{dt}=-(\lambda+\mu)P_n(t)+\lambda P_{n-1}(t)+\mu P_{n+1}(t), \quad n\geqslant 1, \tag{4.3}$$

$$\frac{dP_0}{dt}=-\lambda P_0(t)+\mu P_1(t).$$

2. Решение уравнений

Введем производящую функцию для вероятностей $P_n(t)$

$$P(z, t)=\sum_{n=0}^{\infty} P_n(t) z^n, \tag{4.4}$$

которая, очевидно, сходится внутри единичного круга $|z|\leqslant 1$.

Умножая каждое уравнение на z^{n+1} и суммируя по n, получим

$$z\frac{\partial P(z, t)}{dt}=(1-z)\left[(\mu-\lambda z)P(z, t)-\mu P_0(t)\right]. \tag{4.5}$$

Заметим, что начальным условием является $P(z, 0)=z^i$, так как $P_n(0)=0$, за исключением случая $n=i$, когда $P_i(0)=1$.

Преобразуя это линейное дифференциальное уравнение первого порядка в частных производных по Лапласу, находим преобразование функции $P(z, t)$ в виде

$$P^*(z, s)=\frac{z^{i+1}-\mu(1-z)P_0^*(s)}{sz-(1-z)(\mu-\lambda z)}; \tag{4.6}$$

здесь

$$f^*(s)=\int_0^{\infty} e^{-st}f(t)\,dt.$$

Заметим, что

$$\int_0^\infty e^{-st}\frac{\partial P}{\partial t}dt = e^{-st}P\Big|_0^\infty + s\int_0^\infty e^{-st}Pdt = -z^i + sP^*. \qquad (4.7)$$

Несколько позже будет получено более удобное выражение для $P^*(z, s)$. Коэффициент $P_n^*(s)$ при z^n в разложении $P^*(z, s)$ в степенной ряд является преобразованием функции $P_n(t)$. Взяв обратное преобразование Лапласа, получим искомое значение $P_n(t)$ для переходного процесса.

Так как для $\mathrm{Re}(s)>0$ функция $P^*(z, t)$ сходится внутри единичного круга и на его границе, то нули знаменателя выражения (4.6), лежащие внутри круга $|z|<1$ и на его границе, должны совпадать с соответствующими нулями числителя. Нули $\alpha_k(s)$ знаменателя найдем, решив относительно z квадратное уравнение

$$\alpha_k(s) = \frac{\lambda+\mu+s \pm [(\lambda+\mu+s)^2 - 4\lambda\mu]^{\frac{1}{2}}}{2\lambda}, \quad k=1, 2. \qquad (4.8)$$

Здесь берется значение квадратного корня только с положительной действительной частью, т. е. $\alpha_k(s)$ имеет $\mathrm{Re}(s)>0$. Нуль $\alpha_1(s)$ имеет перед радикалом знак плюс.

Кроме того, согласно теореме Руше знаменатель $P^*(z, s)$ имеет один нуль внутри круга $|z|=1$. В данном случае

$$f(z) = (\lambda+\mu+s)z, \quad g(z) = \lambda z^2 + \mu,$$

$$|f(z)| = |\lambda+\mu+s| > |\lambda+\mu|, \quad |g(z)| = |\lambda+\mu|,$$

и поэтому $|g(z)| < |f(x)|$ на окружности $|z|=1$.

Так как функции $f(z)$ и $g(z)$ — аналитические внутри круга $|z|=1$ и на его границе, то функции $f(z)$ и $f(z)+g(z)$ имеют одно и то же число нулей. Этот нуль должен быть равен $\alpha_2(s)$, а так как $|\alpha_2(s)| < |\alpha_1(s)|$, то на окружности $|z|=1$ нулей нет. Заметим, что

$$\alpha_1+\alpha_2 = \frac{\lambda+\mu+s}{\lambda}, \quad \alpha_1\alpha_2 = \frac{\mu}{\lambda}, \quad s = -\lambda(1-\alpha_1)(1-\alpha_2).$$

Числитель выражения (4.6) при $z=\alpha_2(s)$ должен обращаться в нуль, иначе при этом значении z функция $P^*(z, s)$ не будет существовать. Отсюда найдем

$$P_0^*(s) = \frac{\alpha_2^{i+1}}{\mu(1-\alpha_2)}; \qquad (4.9)$$

подставив значение $P_0^*(s)$ в (4.6), получим

$$P^*(z, s) = \frac{z^{i+1} - \left[\dfrac{(1-z)\alpha_2^{i+1}}{(1-\alpha_2)}\right]}{-\lambda(z-\alpha_1)(z-\alpha_2)}. \qquad (4.10)$$

После умножения числителя и знаменателя на $1-\alpha_2$ числитель упрощается и разлагается на множители. Имеем

$$P^*(z, s)=\frac{(z-\alpha_2)\left(z^i+\alpha_2 z^{i-1}+\ldots+\alpha_2^i\right)-\\ -z\alpha_2(z-\alpha_2)\left(z^{i-1}+\alpha_2 z^{i-2}+\ldots+\alpha_2^{i-1}\right)}{\lambda\alpha_1(z-\alpha_2)\left(1-\frac{z}{\alpha_1}\right)(1-\alpha_2)}. \qquad (4.11)$$

Сокращаем числитель и знаменатель на $z-\alpha_2$, в числителе вычитаем и прибавляем α_2^{i+1}. Это дает возможность вынести за скобки множитель $1-\alpha_2$. Записав

$$\left(1-\frac{z}{\alpha_1}\right)^{-1}=\sum_{k=0}^{\infty}\left(\frac{z}{\alpha_1}\right)^k,$$

получим

$$P^*(z, s)=\frac{1}{\lambda\alpha_1}\left(z^i+\alpha_2 z^{i-1}+\ldots+\alpha_2^i\right)\sum_{k=0}^{\infty}\left(\frac{z}{\alpha_1}\right)^k+$$
$$+\frac{\alpha_2^{i+1}}{\lambda\alpha_1(1-\alpha_2)}\sum_{k=0}^{\infty}\left(\frac{z}{\alpha_1}\right)^k. \qquad (4.12)$$

Заметим, что $\left|\frac{z}{\alpha_1}\right|<1$.

Теперь $P_n^*(s)$ — коэффициент при z^n в преобразованном по Лапласу выражении (4.4). Второй член выражения (4.12) дает следующее слагаемое для этого коэффициента (равное коэффициенту при z^n в этом члене):

$$\frac{\alpha_2^{i+1}}{\lambda\alpha_1^{n+1}(1-\alpha_2)}=\frac{\alpha_2^{i+1}}{\lambda\alpha_1^{n+1}}\left(1+\alpha_2+\alpha_2^2+\ldots\right)=$$
$$=\frac{1}{\lambda}\left(\frac{\lambda}{\mu}\right)^{n+1}\sum_{k=n+i+2}^{\infty}\left(\frac{\mu}{\lambda}\right)^k\frac{1}{\alpha_1^k}. \qquad (4.13)$$

Здесь использовано то обстоятельство, что $|\alpha_2|<1$ и $\alpha_1\alpha_2=\frac{\mu}{\lambda}$. Из последнего соотношения следует, что при умножении каждого значения α_2 в степени m $(m=0, 1, 2, \ldots)$ на α_1^m получим $\left(\frac{\mu}{\lambda}\right)^m$, при этом каждый раз необходимо умножать знаменатель на α_1^m. Остальные коэффициенты при z^n получим из первого члена правой части выражения (4.12) после умножения множителя, записанного слева, на члены разложения, содержащие z в соответствующей степени, и сложения коэффициентов при z^n.

Если, например, $n \geqslant i$, то все члены выражения $\left(\frac{1}{\lambda\alpha_1}\right)(z^i + \alpha_2 z^{i-1} + \ldots + \alpha_2^i)$ войдут в этот коэффициент. Так, z^i умножается на член разложения, содержащий z^{n-i}, и коэффициент, равный в этом случае $\frac{1}{\alpha_1^{n-i}}$, умножается на стоящий слева множитель $\frac{1}{\lambda\alpha_1}$. К этому произведению добавляется коэффициент, полученный при умножении z^{i-1} на z^{n-i+1}, равный $\left(\frac{1}{\lambda\alpha_1}\right)\left(\frac{\alpha_2}{\alpha_1^{n-i+1}}\right)$, и т. д.

В общем виде при умножении z^{i-m} на z^{n-i+m} получим следующее слагаемое этого коэффициента:

$$\frac{1}{\lambda\alpha_1}\,\frac{\alpha_2^m}{\alpha_1^{n-i+m}} = \frac{1}{\lambda\alpha_1}\,\frac{\alpha_2^m\alpha_1^m}{\alpha_1^{n-i+2m}} = \frac{\left(\frac{\mu}{\lambda}\right)^m}{\lambda\alpha_1^{n-i+2m+1}}. \tag{4.14}$$

Следовательно, при $n \geqslant i$ имеем

$$P_n^*(s) = \frac{1}{\lambda}\left[\frac{1}{\alpha_1^{n-i+1}} + \frac{\frac{\mu}{\lambda}}{\alpha_1^{n-i+3}} + \frac{\left(\frac{\mu}{\lambda}\right)^2}{\alpha_1^{n-i+5}} + \ldots + \frac{\left(\frac{\mu}{\lambda}\right)^i}{\alpha_1^{n+i+1}} + \right.$$
$$\left. + \left(\frac{\lambda}{\mu}\right)^{n+1} \sum_{k=n+i+2}^{\infty} \frac{\left(\frac{\mu}{\lambda}\right)^k}{\alpha_1^k}\right]. \tag{4.15}$$

Теперь найдем $P_n(t)$ как обратное преобразование выражения (4.15):

$$P_n(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{st} P_n^*(s)\,ds. \tag{4.16}$$

Чтобы найти обратное преобразование, применим вначале следующую теорему. Если $f^*(s)$ — преобразование Лапласа функции $f(t)$, то $f^*(s+a)$ — преобразование Лапласа функции $e^{-at}f(t)$. Этот результат легко получить непосредственно. Обратное преобразование можно применить к каждому члену выражения для $P_n^*(s)$, т. е. распространить на всю правую часть.

Упражнение. Докажите два последних предложения.

Заметим, что в равенстве (4.8) к s прибавляется постоянная величина $\lambda+\mu$; следовательно, при обратном преобразовании каждого члена выражения (4.15) нужно применить приведенную выше теорему к функциям, преобразование Лапласа которых имеет вид

$(\alpha_1)^{-\nu}$. Выражение $\left[\frac{(s+\sqrt{s^2-4\lambda\mu})}{2\lambda}\right]^{-\nu}$ есть преобразование Лапласа функции

$$(2\lambda)^{-\nu}\nu\left(2\sqrt{\lambda\mu}\right)^{-\nu}t^{-1}I_\nu\left(2\sqrt{\lambda\mu}\,t\right)=\nu\left(\sqrt{\frac{\lambda}{\mu}}\right)^{\nu}t^{-1}I_\nu\left(2\sqrt{\lambda\mu}\,t\right). \quad (4.17)$$

Здесь $I_\nu(z)$ — обобщенная функция Бесселя первого рода:

$$I_\nu(z)=\sum_{k=0}^{\infty}\frac{\left(\frac{z}{2}\right)^{\nu+2k}}{k!\,\Gamma(\nu+k+1)},\quad I_\nu(z)=i^{-\nu}I_\nu(z), \quad (4.18)$$

$$I_\nu(z)=\frac{z^\nu}{2^\nu\nu!}\text{ при } z\to 0,\quad I_\nu(z)=\frac{e^z}{\sqrt{2\pi z}}\text{ при } z\to\infty. \quad (4.19)$$

Окончательно имеем

$$\begin{aligned}P_n(t)=\frac{e^{-(\lambda+\mu)t}}{\lambda}&\left[\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n-i+1}(n-i+1)\,t^{-1}I_{n-i+1}\left(2\sqrt{\lambda\mu}\,t\right)+\right.\\&+\frac{\mu}{\lambda}\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n-i+3}(n-i+3)\,t^{-1}I_{n-i+3}\left(2\sqrt{\lambda\mu}\,t\right)+\ldots+\\&+\left(\frac{\mu}{\lambda}\right)^{i}\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n+i+1}(n+i+1)\,t^{-1}I_{n+i+1}\left(2\sqrt{\lambda\mu}\,t\right)+\\&\left.+\left(\frac{\lambda}{\mu}\right)^{n+1}\sum_{k=n+i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^{k}kt^{-1}I_k\left(2\sqrt{\lambda\mu}\,t\right)\right].\end{aligned} \quad (4.20)$$

Подставляя известное соотношение

$$\frac{2\nu}{z}I_\nu(z)=I_{\nu-1}(z)-I_{\nu+1}(z), \quad (4.21)$$

получим

$$\begin{aligned}P_n(t)=\frac{e^{-(\lambda+\mu)t}}{\lambda}&\left\{\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n-i+1}\sqrt{\lambda\mu}\left[I_{n-i}\left(2\sqrt{\lambda\mu}\,t\right)-\right.\right.\\&\left.-I_{n-i+2}\left(2\sqrt{\lambda\mu}\,t\right)\right]+\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n-i+1}\sqrt{\lambda\mu}\left[I_{n-i+2}\left(2\sqrt{\lambda\mu}\,t\right)-\right.\\&\left.-I_{n-i+4}\left(2\sqrt{\lambda\mu}\,t\right)\right]+\ldots+\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n-i+1}\sqrt{\lambda\mu}\left[I_{n+i}\left(2\sqrt{\lambda\mu}\,t\right)-\right.\\&\left.-I_{n+i+2}\left(2\sqrt{\lambda\mu}\,t\right)\right]+\left(\frac{\lambda}{\mu}\right)^{n+1}\sum_{k=n+i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^{k}\sqrt{\lambda\mu}\left[I_{k-1}\left(2\sqrt{\lambda\mu}\,t\right)-\right.\\&\left.\left.-I_{k+1}\left(2\sqrt{\lambda\mu}\,t\right)\right]\right\}.\end{aligned} \quad (4.22)$$

Но

$$\left(\frac{\lambda}{\mu}\right)^{n+1}\sum_{k=n+i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^{k}\left[I_{k-1}(2\sqrt{\lambda\mu}\,t)-I_{k+1}(2\sqrt{\lambda\mu}\,t)\right]=$$

$$=\left(\frac{\lambda}{\mu}\right)^{n+1}\left[\left(\sqrt{\frac{\mu}{\lambda}}\right)^{n+i+2}I_{n+i+1}(2\sqrt{\lambda\mu}\,t)+\right.$$

$$+\sqrt{\frac{\mu}{\lambda}}\sum_{k=n+i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^{k}I_k(2\sqrt{\lambda\mu}\,t)+\left(\sqrt{\frac{\mu}{\lambda}}\right)^{n+i+1}I_{n+i+2}(2\sqrt{\lambda\mu}\,t)-$$

$$\left.-\sqrt{\frac{\lambda}{\mu}}\sum_{k=n+i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^{k}I_k(2\sqrt{\lambda\mu}\,t)\right]=$$

$$=\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n-1}I_{n+i+1}(2\sqrt{\lambda\mu}\,t)+\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n-i+1}I_{n+i+2}(2\sqrt{\lambda\mu}\,t)+$$

$$+\left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n}\sqrt{\frac{\lambda}{\mu}}\sum_{k=n+i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^{k}I_k(2\sqrt{\lambda\mu}\,t);\qquad(4.23)$$

поэтому второе слагаемое первого члена равенства (4.22) сокращается с первым слагаемым второго члена и т. д. При $n\geqslant i$ имеем

$$P_n(t)=\mathrm{e}^{-(\lambda+\mu)t}\left[\left(\sqrt{\frac{\mu}{\lambda}}\right)^{i-n}I_{n-i}(2\sqrt{\lambda\mu}\,t)+\right.$$

$$+\left(\sqrt{\frac{\mu}{\lambda}}\right)^{i-n+1}I_{n+i+1}(2\sqrt{\lambda\mu}\,t)+$$

$$\left.+\left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n}\sum_{k=n+i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^{k}I_k(2\sqrt{\lambda\mu}\,t)\right].\qquad(4.24)$$

Покажем теперь, что это выражение справедливо также и при $n<i$.

Отыскание коэффициента при z^n в первом выражении правой части равенства (4.12) начинается с умножения членов ряда $\sum_{k=0}^{\infty}\left(\frac{z}{\alpha}\right)^k$ на член множителя $\left(\frac{1}{\lambda\alpha_1}\right)(z_i+\alpha_2 z^{i-1}+\ldots+\alpha_2^i)$, содержащий z^n, при этом используются все члены, стоящие справа от z^n. Далее все операции выполняются аналогично. Следовательно, сокращения будут производиться таким же образом. Однако необходимо показать, что первый член при $n<i$ совпадает с первым членом равенства (4.24).

Первое слагаемое для коэффициента при z^n найдем путем умножения коэффициента при z^n во множителе, стоящем слева от знака суммирования, на первый член разложения (равный единице). Получим выражение

$$\frac{\alpha_2^{i-n}}{\lambda\alpha_1}=\frac{\left(\frac{\mu}{\lambda}\right)^{i-n}}{\lambda\alpha_1^{i-n+1}}\,. \qquad (4.25)$$

Соответствующее ему обратное преобразование имеет вид

$$\frac{e^{-(\lambda+\mu)t}}{\lambda}\left[\left(\frac{\mu}{\lambda}\right)^{i-n}\left(\sqrt{\frac{\lambda}{\mu}}\right)^{i-n+1}(i-n+1)\,t^{-1}I_{i-n+1}(2\sqrt{\lambda\mu}\,t)\right]=$$

$$=\frac{e^{-(\lambda+\mu)t}}{\lambda}\left\{\left(\sqrt{\frac{\lambda}{\mu}}\right)^{n-i+1}\sqrt{\lambda\mu}\left[I_{i-n}(2\sqrt{\lambda\mu}\,t)-\right.\right.$$

$$\left.\left.-I_{i-n+2}(2\sqrt{\lambda\mu}\,t)\right]\right\}. \qquad (4.26)$$

Кроме того, известно, что $I_{-n}(z)=I_n(z)$; поскольку, как было показано ранее, второе выражение сокращается, первые члены совпадают. Таким образом, решение найдено полностью.

3. Доказательство того, что сумма вероятностей равна единице

Покажем, что сумма вероятностей равна единице. Доказательство будет просто упражнением в изменении порядка суммирования.

$$\sum_{n=0}^{\infty}P_n(t)=e^{-(\lambda+\mu)t}\left[\sum_{m=-i}^{\infty}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t)+\right.$$

$$+\left(\sqrt{\frac{\lambda}{\mu}}\right)^{-2(i+1)}\sum_{m=i+1}^{\infty}\left(\frac{\lambda}{\mu}\right)^m I_m(2\sqrt{\lambda\mu}\,t)+$$

$$\left.+\left(1-\frac{\lambda}{\mu}\right)\sum_{n=0}^{\infty}\sum_{k=n+i+2}^{\infty}\left(\frac{\lambda}{\mu}\right)^n\left(\sqrt{\frac{\mu}{\lambda}}\right)^k I_k(2\sqrt{\lambda\mu}\,t)\right]. \qquad (4.27)$$

После изменения порядка суммирования последнее выражение примет вид

$$\left(1-\frac{\lambda}{\mu}\right)\sum_{k=i+2}^{\infty}\sum_{n=0}^{k-i-2}\left(\frac{\lambda}{\mu}\right)^n\left(\sqrt{\frac{\mu}{\lambda}}\right)^k I_k(2\sqrt{\lambda\mu}\,t)=$$

$$=\sum_{k=i+2}^{\infty}\left[1-\left(\frac{\lambda}{\mu}\right)^{k-i-1}\right]\left(\sqrt{\frac{\mu}{\lambda}}\right)^k I_k(2\sqrt{\lambda\mu}\,t)=$$

$$=\sum_{m=i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^m I_m(2\sqrt{\lambda\mu}\,t)-$$

$$-\left(\sqrt{\frac{\lambda}{\mu}}\right)^{-2(i+1)}\sum_{m=i+2}^{\infty}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t). \qquad (4.28)$$

Объединяя первое выражение правой части формулы (4.28) с первым слагаемым, стоящим в квадратных скобках равенства (4.27), используя то обстоятельство, что $I_n(z) = I_{-n}(z)$, и прибавляя затем второе выражение соотношения (4.28) ко второму слагаемому, стоящему в квадратных скобках равенства (4.27), получим

$$\sum_{m=0}^{\infty} P_m(t) = e^{-(\lambda+\mu)t}\left[\sum_{m=-i}^{\infty}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t) + \right.$$

$$\left. + \sum_{m=-\infty}^{-(i+2)}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t) + \left(\sqrt{\frac{\lambda}{\mu}}\right)^{-(i+1)} I_{-(i+1)}(2\sqrt{\lambda\mu}\,t)\right] =$$

$$= e^{-(\lambda+\mu)t}\sum_{m=-\infty}^{\infty}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t) =$$

$$= e^{-(\lambda+\mu)t}\exp\left[\sqrt{\lambda\mu}\left(\sqrt{\frac{\lambda}{\mu}} + \sqrt{\frac{\mu}{\lambda}}\right)t\right] = 1. \qquad (4.29)$$

Здесь использован тот факт, что

$$e^{\left(\frac{x}{2}\right)\left(y+\frac{1}{y}\right)} = \sum_{n=-\infty}^{\infty} y^n I_n(x) = \sum_{n=0}^{\infty} y^n I_n(x) + \sum_{n=1}^{\infty} y^{-n} I_n(x), \qquad (4.30)$$

так как $I_{-n}(x) = I_n(x)$, если n — целое число.

4. Решение для стационарного состояния

Упражняясь в действиях с асимптотическими разложениями, покажем, что полученное решение при $t \to \infty$ стремится к стационарному, если $\lambda < \mu$. Упростим вычисления, положив $i=0$. Для фиксированных значений λ и μ при $t \to \infty$ получим выражение

$$I_n(2\sqrt{\lambda\mu}\,t) \approx \frac{\exp(2\sqrt{\lambda\mu}\,t)}{\sqrt{2\pi(2\sqrt{\lambda\mu}\,t)}}, \qquad (4.31)$$

которое не зависит от n.

Подставим это значение I_n в первые два члена, стоящие в квадратных скобках выражения (4.24), а результат умножим на их общий множитель, записанный слева. Можно легко показать, что каждое из двух полученных выражений стремится к нулю при $t \to \infty$. При этом принимается во внимание то обстоятельство, что $\lambda+\mu > 2\sqrt{\lambda\mu}$. Это неравенство справедливо при $\lambda/\mu < 1$. Рассмотрим поведение последнего члена выражения (4.24) при $t \to \infty$. Заметим, что

$$\sum_{k=n+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}}\right)^k I_k(2\sqrt{\lambda\mu}\, t) = \sum_{k=0}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}}\right)^k I_k(2\sqrt{\lambda\mu}\, t) - \sum_{k=0}^{n+1} \left(\sqrt{\frac{\mu}{\lambda}}\right)^k I_k(2\sqrt{\lambda\mu}\, t). \tag{4.32}$$

После умножения на $e^{-(\lambda+\mu)t}$ конечная сумма в правой части при $t \to \infty$ стремится к нулю, следовательно, и каждый член, взятый отдельно, также стремится к нулю. Теперь рассмотрим разложение в ряд.

После умножения на $e^{-(\lambda+\mu)t}$ при $t \to \infty$ последняя сумма в правой части равенства (4.30) при $y = \sqrt{\frac{\mu}{\lambda}}$ и $x = \sqrt{\lambda\mu}\, t$ будет стремиться к нулю; к этому выводу приходим путем рассмотрения асимптотического выражения для $I_n(x)$ при условии $\frac{\lambda}{\mu} < 1$. С другой стороны, первая сумма в правой части выражения (4.30) и есть искомое разложение в ряд выражения (4.32). Окончательно находим

$$e^{-(\lambda+\mu)t} \sum_{n=0}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}}\right)^k I_k(2\sqrt{\lambda\mu}\, t) = e^{-(\lambda+\mu)t} \exp\left[\frac{1}{2} t (2\sqrt{\lambda\mu})\left(\sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}}\right)\right] = 1. \tag{4.33}$$

При $t \to \infty$

$$P_n(t) \to \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right). \tag{4.34}$$

5. Вероятность того, что число требований в системе не меньше заданного

Читатель может убедиться самостоятельно, что вероятность того, что число требований в системе больше или равно j, задается выражением

$$\sum_{n=j}^{\infty} P_n(t) = e^{-(\lambda+\mu)t} \left[\sum_{n=j-i}^{\infty} \left(\frac{\lambda}{\mu}\right)^n I_n(2\sqrt{\lambda\mu}\, t) + \sum_{n=j+i+1}^{\infty} \left(\frac{\mu}{\lambda}\right)^{n-2j} I_n(2\sqrt{\lambda\mu})\right]. \tag{4.35}$$

6. Другой способ решения

Конолли [145] получил решение данной задачи, применив преобразование Лапласа к системе уравнений, а не к уравнению для суммы $P(z, t)$, как это делали мы. Он ввел начальные условия в уравнение с $P'_i(t)$, где i — начальное состояние в момент времени $t=0$. Полученная бесконечная система дифференциальных уравнений линейна. Решение ее может быть записано в общем виде как

$$P_n^*(s) = A\alpha_1^n + B\alpha_2^n. \tag{4.36}$$

Здесь A и B не зависят от n.

Из условия $\sum_{n=0}^{\infty} P_n(t) = 1$, из которого следует, что

$$\sum_{n=0}^{\infty} P_n^*(s) = \frac{1}{s}, \tag{4.37}$$

заключаем, что последний бесконечный ряд будет расходиться, так как $|\alpha_1| > 1$, если не сделано разграничение между значениями $n \leqslant i$ и значениями $n \geqslant i$.

При $n \leqslant i$ используем равенство (4.36), а при $n \geqslant i$ воспользуемся выражением $P_n^*(s) = C\alpha_2^n$, где C не зависит от n. Эти два выражения должны совпадать при $n=i$. Это условие дает первое соотношение между A, B, и C. Еще одно соотношение получим из первого уравнения $(s+\lambda)P_0^*(s) = \mu P_1^*(s)$. Третье соотношение найдем из этого уравнения при $n=i$, в котором подставляется соответствующее решение, имея в виду, что $n=i$ есть критическая точка. Таким образом, решение определено при $n \leqslant i$ и при $n \geqslant i$.

7. Решение методом случайного блуждания

Случай 1 был также рассмотрен Чемпернауном [109], который использовал метод случайного блуждания.

Введем обозначения: q_t — длина очереди в момент времени t; λ — параметр пуассоновского входящего потока, b_t — число требований, поступивших за время t. Таким образом,

$$P(b_t = b) = e^{-\lambda t}\frac{(\lambda t)^b}{b} \equiv g_t(b). \tag{4.38}$$

Пусть среднее время обслуживания равно единице, т. е. $\mu=1$. Время обслуживания имеет экспоненциальное распределение. Если a_t — число требований, обслуженных за время $(0, t)$, — изменяется в зависимости от величины параметра t по закону Пуассона, то

$$P(a_t = a) = e^{-t}\frac{t^a}{a!} \equiv h_t(a). \tag{4.39}$$

Пусть

$$c_t = b_t - a_t, \tag{4.40}$$

а ν_t — наибольшее по абсолютной величине отрицательное значение c_t в интервале $0 \leqslant t' \leqslant t$. Тогда

$$\begin{aligned} q_t &= q_0 + c_t \quad \text{при } \nu_t + q_0 \geqslant 0, \\ q_t &= -\nu_t + c_t \quad \text{при } \nu_t + q_0 < 0. \end{aligned} \tag{4.41}$$

Необходимо найти $f_1(q) \equiv P(q_t = q)$.

Вначале необходимо определить $J_t(\nu|a, b)$ — функцию распределения для ν_t при условии, что $a_t = a$ и $b_t = b$. Все траектории a_T и b_T в интервале $0 \leqslant T \leqslant t$ таковы, что события $\{a_T = a\}$ и $\{b_T = b\}$ имеют одинаковую вероятность. Таким образом, $J_t(\nu|a, b)$ есть доля тех $\binom{a+b}{a}$ траекторий, содержащих b шагов

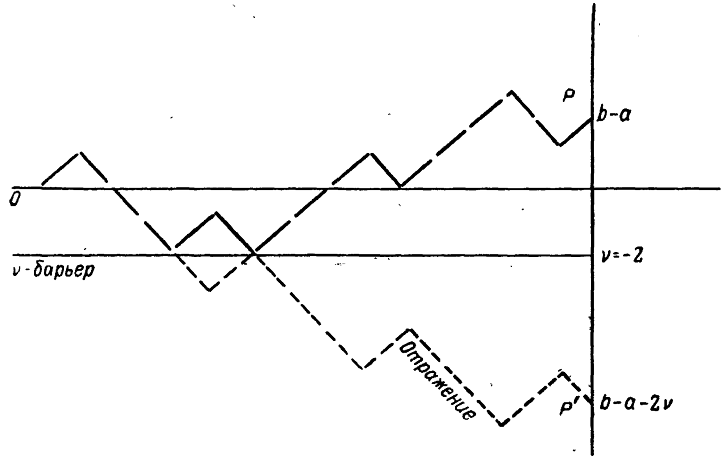

Рис. 4.1.

вверх и a шагов вниз, которые достигают барьера ν. При $0 \geqslant \nu \geqslant -a$ существует $\binom{a+b}{a+\nu}$ таких траекторий. Отсюда следует, что

$$J_t(\nu|a, b) = \frac{a!\,b!}{(a+\nu)!\,(b-\nu)!}, \quad 0 \geqslant \nu \geqslant -a, \tag{4.42}$$

$$J_t(\nu|a, b) = \begin{cases} 1, & 0 \leqslant \nu, \\ 0, & \nu < -a. \end{cases} \tag{4.43}$$

Рассмотрим случай, когда $\nu = -2$, $a = 7$ и $b = 9$, как показано на рис. 4.1. Все траектории идут из O в P; всем траекториям, которые достигают барьера ν, соответствуют отраженные траектории, ведущие из O в P', где P' — отражение точки P относительно барьера ν. Кроме того, каждая последовательность шагов, ведущая из O в P', есть траектория, являющаяся зеркальным отражением по отношению к траектории, идущей из O в P.

Чтобы попасть из O в P' за $a+b$ шагов, потребуется $b-\nu$ шагов вверх и $a+\nu$ шагов вниз. Поэтому таких траекторий всего $\binom{a+b}{a+\nu}$. В нашем случае имеется $\binom{16}{5}$ траекторий длиной в 16 шагов, идущих из O в P'. Следовательно, имеется и $\binom{16}{5}$ таких траекторий, идущих из O в P, которые всегда пересекаются с барьером $\nu=-2$. Отсюда следует соотношение $\dfrac{\binom{16}{5}}{\binom{16}{7}}$ для общего числа траекторий длиной в 16 шагов, идущих из O в P.

Пусть $F_t(q\,|\,a, b)$ — вероятность того, что $q_t \geqslant q$, при условии, что $a_t = a$ и $b_t = b$, и пусть $F_t(q)$ — вероятность того, что $q_t \geqslant q$; тогда

$$F_t(q) = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} h_t(a)\, g_t(b)\, F_t(q\,|\,a, b), \tag{4.44}$$

и искомое распределение имеет вид

$$f_t(q) = F_t(q) - F_t(q-1). \tag{4.45}$$

Из равенств (4.40) и (4.41) и соотношений $a_t = a$ и $b_t = b$ находим, что

$$F_t(q\,|\,a, b) = 1 \text{ при } q \leqslant q_0 + b - a,$$

а из равенств (4.41) и (4.43) имеем

$$F_t(q\,|\,a, b) = J_t(-a-1\,|\,a, b) = 0$$
$$\text{при } q > q_0 + b - a \text{ и } q > b.$$

Из равенства (4.42) и соотношений $a_t = a$ и $b_t = b$ имеем

$$F_t(q\,|\,a, b) = J_t(b - a - q\,|\,a, b) = \frac{a!\,b!}{(a+q)!\,(b-q)!}$$
$$\text{при } b \geqslant q > q_0 + b - a.$$

Например, для случая $q \geqslant q_0$

$$F_t(q) = e^{-(1+\lambda)t}\left[\sum_{a=0}^{\infty} \sum_{b=a+q-q_0}^{\infty} \frac{\lambda^a t^{a+b}}{a!\,b!} + \sum_{b=q}^{\infty} \sum_{a=b-q+q_0+1}^{\infty} \frac{\lambda^b t^{a+b}}{(a+q)!\,(b-q)!}\right] =$$

$$= e^{-(1+\lambda)t}\left[\sum_{c=q-q_0}^{\infty} \lambda^{\frac{c}{2}} I_c\left(2\lambda^{\frac{1}{2}} t\right) + \sum_{c=q+q_0+1}^{\infty} \lambda^{q-\frac{c}{2}} I_c\left(2\lambda^{\frac{1}{2}} t\right)\right], \tag{4.46}$$

а $f_t(q)$ находится по формуле (4.45), которая приводит к известному выражению (4.24). Решение для случая $q < q_0$ можно получить аналогично.

Замечание. Выражение для распределения длительности периода занятости см. в задаче 15, приведенной в конце главы.

4.3. Случай 2. Бесконечное число каналов; $\lambda_n = \lambda$, $\mu_n = n\mu$

Очевидно, что в этом случае очереди не образуется, так как поступающее требование немедленно направляется на обслуживание. Уравнения имеют вид (индекс i для упрощения снова опущен)

$$\frac{dP_n(t)}{dt} = -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \quad n \geqslant 1,$$

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t). \tag{4.47}$$

Введем производящую функцию для $P_n(t)$:

$$P(z, t) = \sum_{n=0}^{\infty} P_n(t) z^n.$$

Как и ранее, для производящей функции имеем следующее дифференциальное уравнение в частных производных:

$$\frac{\partial P}{\partial t} = (1-z)\left(-\lambda P + \mu \frac{\partial P}{\partial z}\right) \tag{4.48}$$

или в стандартной форме записи

$$\frac{\partial P}{\partial t} - (1-z)\mu \frac{\partial P}{\partial z} = -\lambda(1-z)P.$$

Найдем его решение обычным способом.

Соответствующие уравнения Лагранжа имеют вид

$$\frac{dt}{1} = \frac{dz}{-(1-z)\mu} = \frac{dP}{-\lambda(1-z)P}. \tag{4.49}$$

Здесь коэффициент при dt является коэффициентом при $\frac{\partial P}{\partial t}$, коэффициент при dz является обратной величиной коэффициента при $\frac{\partial P}{\partial z}$, а коэффициент при dP — обратной величиной к правой части уравнения.

При решении уравнений по методу Лагранжа первое уравнение используется для нахождения одного однопараметрического семейства поверхностей $u(z, t, P) = c_1$, а второе уравнение — семейства $v(z, t, P) = c_2$. Общее решение уравнения есть такая произвольная функция f от u и v, что $f(u, v) = 0$. Можно найти решение относительно v в виде $v = g(u)$. Затем вид функции g можно определить из начальных условий.

В нашем случае

$$\frac{dt}{1} = \frac{dz}{-(1-z)\mu},$$

откуда

$$u(z, t, P) \equiv (1-z)\,e^{-\mu t} = c_1.$$

Кроме того,

$$\frac{dz}{-(1-z)\mu}=\frac{dP}{-\lambda(1-z)P},$$

откуда

$$v\equiv Pe^{-\left(\frac{\lambda}{\mu}\right)z}=c_2.$$

Следовательно, общее решение имеет вид

$$P=e^{\left(\frac{\lambda}{\mu}\right)z}g\left[(1-z)e^{-\mu t}\right].$$

При $t=0$ имеем $P=z^i$, поскольку все значения $P_n(0)$ равны нулю, кроме одного, а именно $P_i(0)=1$, так как начальным состоянием системы является E_i. Поэтому

$$z^i=e^{\left(\frac{\lambda}{\mu}\right)z}g(1-z).$$

Пусть $y=1-z$, тогда

$$g(y)=e^{-\left(\frac{\lambda}{\mu}\right)(1-y)}(1-y)^i.$$

При произвольном значении t аргументом функции g является $(1-z)e^{-\mu t}$, поэтому в правой части этого уравнения нужно заменить y на значение этого аргумента. Получим $g\left[(1-z)e^{-\mu t}\right]$. Подставив это выражение в общее решение, получим

$$P(z,t)=e^{\left(\frac{\lambda}{\mu}\right)z}e^{-\left(\frac{\lambda}{\mu}\right)[1-(1-z)e^{-\mu t}]}\left[1-(1-z)e^{-\mu t}\right]^i=$$

$$=\left\{\exp\left[-\frac{\lambda}{\mu}(1-z)(1-e^{-\mu t})\right]\right\}\left[1-(1-z)e^{-\mu t}\right]^i. \quad (4.50)$$

Итак,

$$P_n(t)=\frac{1}{n!}\left.\frac{\partial^n P(z,t)}{\partial z^n}\right|_{z=0}.$$

Заметим, что $P(z, t)$ — произведение двух множителей. Если первый из них обозначить через $f(z)$, а второй — через $g(z)$, то для нахождения производной можно применить формулу Лейбница:

$$\frac{d^n f(z)g(z)}{dz^n}=\sum_{k=0}^{n}\binom{n}{k}f^{n-k}(z)g^k(z). \quad (4.51)$$

При $n<i$ имеем

$$P_n(t)=\frac{\exp\left[\left(-\frac{\lambda}{\mu}\right)(1-e^{-\mu t})\right]}{n}\times$$

$$\times\sum_{k=0}^{n}\binom{n}{k}\frac{i!}{(i-k)!}\left(\frac{\lambda}{\mu}\right)^{n-k}(1-e^{-\mu t})^{n+i-2k}e^{-\mu kt}. \quad (4.52)$$

Аналогичное выражение можно получить и для $n>i$. (Определение среднего числа требований, находящихся в системе, см. в задаче 13.)

Если начальное число требований есть величина случайная и распределена, например, по закону Пуассона с математическим ожиданием, равным m, то следует рассмотреть выражение

$$\sum_{i=0}^{\infty} \frac{m^i e^{-m}}{i!} P_{in}(t);$$

получим вероятности, которые зависят не от начального числа требований, находящихся в системе, а от его математического ожидания. Можно также найти новую производящую функцию

$$\sum_{i=0}^{\infty} \frac{m^i e^{-m}}{i!} P(z, t).$$

Упражнение. Вычислите введенную производящую функцию.

4.4. Случай 3. $\lambda_n=\lambda(t)$, $\mu_n=\mu(t)$ (параметры зависят от времени)

Кларк [124] исследовал одноканальную систему, в которой в момент времени $t=0$ находится начальное число требований i и параметры которой зависят от времени. При анализе задачи массового обслуживания, удовлетворяющей этим условиям, было принято, что параметр входящего потока есть периодическая функция времени. В этом параграфе и рассматривается такой случай. Выкладки довольно громоздки, поэтому мы приводим их в сокращенном виде. Читателю предлагается рассмотреть аналогичный случай, сформулированный в задаче 24. Вначале введем преобразование времени τ, определяемое следующим образом:

$$\tau \equiv \tau(t) = \int_0^t \mu(s)\, ds. \tag{4.53}$$

Затем, используя для упрощения вычислений масштаб времени τ, положим:

$$\rho(\tau) = \frac{\lambda(t)}{\mu(t)}, \tag{4.54}$$

$$R(\tau) \equiv \frac{\int_0^t \lambda(s)\, ds}{\int_0^t \mu(s)\, ds} = \frac{1}{\tau} \int_0^\tau \rho(s)\, ds, \tag{4.55}$$

$$\frac{d}{d\tau} P_0(\tau) = -\rho(\tau) P_0(\tau) + P_1(\tau), \tag{4.56}$$

$$\frac{d}{d\tau} P_n(\tau) = -[1+\rho(\tau)] P_n(\tau) + \rho(\tau) P_{n-1}(\tau) + P_{n+1}(\tau), \quad n>0,$$

$$P_n(0) = \delta_{in}.$$

Пусть $Q_n(\tau)=e^{\tau[1+R(\tau)]}P_n(\tau)$, $n=0, 1, \ldots$, тогда система уравнений примет вид

$$\frac{dQ_0(\tau)}{d\tau}=Q_0(\tau)+Q_1(\tau),$$

$$\frac{dQ_n(\tau)}{d\tau}=\rho(\tau)\,Q_{n-1}(\tau)+Q_{n+1}(\tau), \quad n>0. \tag{4.57}$$

Как и ранее, $Q_{in}(0)=\delta_{in}$. (Заметим, что индекс i, который характеризует начальное состояние, снова опущен для упрощения обозначений.)

Применение производящей функции

$$Q(z, \tau)=\sum_{n=0}^{\infty} Q_n(\tau)\frac{(z-\tau)^n}{n!} \tag{4.58}$$

приводит эту систему уравнений к гиперболическому дифференциальному уравнению в частных производных, которое называется телеграфным уравнением:

$$\frac{\partial^2 Q}{\partial\tau\partial z}=\rho(\tau)\,Q. \tag{4.59}$$

Граничные условия

$$\left.\frac{\partial Q(z, \tau)}{\partial\tau}\right|_{z=\tau}=Q(\tau, \tau)$$

находим, дифференцируя $Q(z, \tau)$ по τ, принимая $z=\tau$ и используя уравнение для $Q_0'(\tau)$. Кроме того, из условия начального состояния системы находим, что $Q(z, 0)=\frac{z^i}{i!}$. Пусть $f(\tau)=$ $=\frac{\partial Q(0, \tau)}{\partial\tau}$.

Дифференциальное уравнение (4.59) содержит функцию Римана (обобщенную функцию Бесселя)

$$I_0\left(2\left\{[R(\tau)\tau-R(\sigma)\sigma](z-\zeta)\right\}^{\frac{1}{2}}\right); \tag{4.60}$$

после интегрирования приходим к выражению

$$Q(z, \tau)=A_i(0, \tau, z)+\int_0^{\tau} A_0(\sigma, \tau, z)\,f_i(\sigma)\,d\sigma. \tag{4.61}$$

Здесь:

$$A_n(\sigma, \tau, z) = z^{\frac{n}{2}} [R(\tau)\tau - R(\sigma)\sigma]^{-\frac{n}{2}} I_n\left(2\{[R(\tau)\tau - R(\sigma)\sigma] z\}^{\frac{1}{2}}\right),$$

$$n = 0, \pm 1, \ldots,$$

$$\frac{\partial}{\partial z} A_n(\sigma, \tau, z) = A_{n-1}(\sigma, \tau, z);$$

$$\frac{\partial}{\partial \tau} A_n(\sigma, \tau, z) = \rho(\tau) A_{n+1}(\sigma, \tau, z), \tag{4.62}$$

$$A_n(0, 0, 0) = \delta_{on},$$

$$A_{-n}(\tau, \tau, \tau) = \begin{cases} 0, & n > 0, \\ 1, & n = 0. \end{cases}$$

Пусть

$$B_n(\sigma, \tau) = A_n(\sigma, \tau, \tau) - \rho(\tau) A_{n+1}(\sigma, \tau, \tau).$$

Тогда

$$f_i(\tau) = B_i(0, \tau) + \int_0^\tau B_0(\sigma, \tau) f_i(\sigma)\, d\sigma. \tag{4.63}$$

Это выражение можно использовать для получения решения данной задачи в явном виде, т. е.

$$P_n(\tau) = \exp\{-\tau[1 + R(\tau)]\} \left[A_{i-n}(0, \tau, \tau) + \right.$$

$$\left. + \int_0^\tau A_{-n}(\sigma, \tau, \tau) f_i(\sigma)\, d\sigma \right]. \tag{4.64}$$

Если $R(\tau) \equiv \rho$ — постоянная величина, т. е. $\mu(\tau) = \rho\lambda(\tau)$, то можно показать, что

$$P_{in}(\tau) = e^{-\tau(1+\rho)} \left[A_{i-n}(0, \tau, \tau) + \rho^n A_{n+i+1}(0, \tau, \tau) + \right.$$

$$\left. + (1 - \rho)\rho^n \sum_{k=n+i+2}^{\infty} A_k(0, \tau, \tau) \right]. \tag{4.65}$$

Заметим, что

$$P_{in}(\tau) = \rho^{-i} P_{0, i+n}(\tau) +$$

$$+ e^{-\tau(1+\rho)} [A_{i-n}(0, \tau, \tau) - \rho^n A_{i+n}(0, \tau, \tau)]. \tag{4.66}$$

Таким образом, вычисление $P_{in}(\tau)$ может проводиться с помощью таблиц для $P_{0n}(\tau)$.

Дифференцируя математическое ожидание и дисперсию

$$L(\tau) = \sum_{n=0}^{\infty} n P_n(\tau),$$

$$\sigma^2(\tau) = \sum_{n=0}^{\infty} [n - L(\tau)]^2 P_n(t),$$

используя систему дифференциально-разностных уравнений, записанных для τ, и учитывая, что $\sum_{n=0}^{\infty} P_n(t)=1$, получим

$$\frac{dL(\tau)}{d\tau}=\rho(\tau)-1+P_0(\tau),$$

или

$$L(\tau)=[R(\tau)-1]\,\tau+\int_0^{\tau} P_0(\sigma)\,d\sigma+i, \tag{4.67}$$

$$\frac{d\sigma^2(\tau)}{d\tau}=2\,\{\rho(\tau)+[\rho(\tau)-1]\,L(\tau)\}-[1+2L(\tau)]\,\frac{dL(\tau)}{d\tau}, \tag{4.68}$$

откуда

$$\sigma^2(\tau)=2\int_0^{\tau}\{\rho(\sigma)+[\rho(\sigma)-1]\,L(\sigma)\}\,d\sigma-L(\sigma)-[L(\sigma)]^2+i+i^2. \tag{4.69}$$

Если функция $\rho(\tau)$ ограничена для всех значений τ, то при $\tau\to\infty$ математическое ожидание и дисперсия равны $O(\tau)$.

Если $\rho>1$ — величина постоянная, то, воспользовавшись тем обстоятельством, что

$$\int_0^{\infty} e^{-\tau(1+\rho)}A_n(0,\ \tau,\ \tau)\,d\tau=\frac{1}{\rho^n(\rho-1)},$$

и полагая $\tau=\int_0^{t}\mu(s)\,ds\to\infty$ при $\tau\to\infty$, получим

$$\int_0^{\infty} P_0(\sigma)\,d\sigma=\frac{1}{\rho^i(\rho-1)}\ \text{или}\ \int_0^{\tau} P_0(\sigma)\,d\sigma=\frac{1}{\rho^i(\rho-1)}+O(1), \tag{4.70}$$

и

$$L(\tau)=(\rho-1)\,\tau+\frac{1}{\rho^i(\rho-1)}+i+O(1), \tag{4.71}$$

$$\sigma(\tau)=\sqrt{\tau(\rho+1)}+O(\sqrt{\tau}). \tag{4.72}$$

Если $\rho=1$, то при больших τ решение имеет вид

$$P_0(\tau)=e^{-2\tau}\,[I_i(2\tau)+I_{i+1}(2\tau)]\approx\frac{1}{\sqrt{\pi\tau}}\left[1+O\left(\frac{1}{\tau}\right)\right], \tag{4.73}$$

а, следовательно,

$$L(\tau)=\int_0^{\tau} P_0(\sigma)\,d\sigma+i=2\sqrt{\frac{\tau}{\pi}}+O(1), \tag{4.74}$$

$$\sigma(\tau)=\sqrt{2\tau\left(1-\frac{2}{\pi}\right)}+O(1). \tag{4.75}$$

4.5. Случай 4. $\lambda_n=\lambda(t)$, $\mu_n=n\mu$ (параметр входящего потока зависит от времени; бесконечно большое число каналов)

Для упрощения примем среднее время обслуживания равным одной временнóй единице, т. е. $\mu=1$. Здесь $P_n(t)$ — вероятность того, что в момент времени t будет занято ровно n каналов. Соответствующее уравнение имеет вид [862]

$$P'_n(t)=-[\lambda(t)+n]P_n(t)+(n+1)P_{n+1}(t)+\lambda(t)P_{n-1}(t). \qquad (4.76)$$

Введя производящую функцию

$$P(z,\ t)=\sum_{n=0}^{\infty}P_n(t)z^n,$$

получим линейное дифференциальное уравнение первого порядка:

$$\frac{\partial P}{\partial t}+(z-1)\frac{\partial P}{\partial z}=\lambda(t)(z-1)P. \qquad (4.77)$$

Если принять, что $\lambda(t)=\lambda$ — величина постоянная, то

$$\lim_{t\to\infty}P(z,\ t)=e^{\lambda(z-1)}. \qquad (4.78)$$

Отсюда

$$p_n=\lim_{t\to\infty}P_n(t)=\frac{\lambda^n}{n!}e^{-\lambda},$$

что совпадает с выражением для переменного параметра, т. е. независимо от начального состояния число занятых каналов в пределе будет распределено по закону Пуассона.

Если принять, что в момент времени $t=0$ также имеет место распределение Пуассона, то для $\lambda=\lambda(0)$ получим

$$P(z,\ 0)=e^{\lambda(z-1)}, \qquad (4.79)$$

и окончательное решение, использующее эти условия, имеет вид

$$P_n(t)=\frac{[\Lambda(t)]^n}{n!}e^{-\Lambda(t)},\quad \Lambda(t)=e^{-t}\left[\lambda+\int_0^t\lambda(t)\,e^t dt\right]. \qquad (4.80)$$

Упражнение. Выполните все промежуточные выкладки.

Заметим, что в частном случае, когда λ — величина постоянная и число каналов равно N, для стационарного состояния имеем

$$\sum_{n=0}^{N}p_n=1 \text{ и } p_n=\frac{\lambda^n\dfrac{e^{-\lambda}}{n!}}{e^{-\lambda}\sum\limits_{j=1}^{N}\dfrac{\lambda_j}{j!}}=\frac{\dfrac{\lambda^n}{n!}}{\sum\limits_{j=1}^{N}\dfrac{\lambda^j}{j!}}. \qquad (4.81)$$

Эта формула представляет собой распределение Эрланга.

Можно получить и более общее выражение, заменив λ на $\frac{\lambda}{\mu}\equiv\rho$.

4.6. Случай 5. Многофазовое обслуживание

1. Число фаз в системе

Рассмотрим одноканальную систему массового обслуживания с пуассоновским входящим потоком с параметром λ и предположим, что обслуживание является многофазовым. Время обслуживания требования представляет собой сумму случайного числа фаз, длительность каждой из которых распределена по экспоненциальному закону с параметром μ. Вероятность того, что время обслуживания требования будет состоять из j фаз, обозначим через c_j. Таким образом, в каждый момент времени систему можно характеризовать состоянием n, где n — общее число фаз обслуживания находящихся в ней требований. Если при $t=0$ система находится в i-м состоянии, и если $P_n(t)$ — вероятность n-го состояния системы в момент времени t, то

$$P'_0(t) = -\lambda P_0(t) + \mu P_1(t), \tag{4.82}$$

$$P'_n(t) = -(\lambda+\mu) P_n(t) + \mu P_{n+1}(t) + \lambda \sum_{j=1}^{n} c_j P_{n-j}(t), \quad n \geqslant 1.$$

Введя производящую функцию

$$P(z, t) = \sum_{n=0}^{\infty} P_n(t) z^n$$

при $P(z, 0) = z^i$, получим [534]:

$$z\frac{\partial P}{\partial t} = \left[\mu - (\lambda+\mu) z + \lambda z \sum_{j=1}^{\infty} c_j z^j\right] P + \mu (z-1) P_0(t).$$

Преобразование Лапласа функции $P(z, t)$ имеет вид

$$P^*(z, s) = \frac{z^{i+1} + \mu (z-1) P_0^*(s)}{(\lambda+\mu+s) z - \mu - \lambda z \sum_{j=1}^{\infty} c_j z^j}. \tag{4.83}$$

Заметим, что при $c_1 = 1$, $c_j = 0$ и $j \neq 1$ задача сводится к описанной для первого случая.

Как уже было показано ранее при помощи теоремы Руше, знаменатель имеет внутри единичного круга единственный простой нуль ξ, который является также нулем и для числителя. Это дает возможность вычислить $P_0(t)$, а затем и $P_n(t)$, используя исходную систему уравнений или применяя $P^*(z, s)$. Приведем способ определения вероятности $P_0(t)$, которая используется для вычисления других вероятностей.

Пусть

$$\rho=\frac{\lambda}{\mu},$$

$$\xi=\frac{\mu}{\lambda+\mu+s},$$

$$w=\frac{\lambda}{\lambda+\mu+s},$$

$$g(z)=z\sum_{j=1}^{\infty}c_j z^j.$$

Произведем подстановку $\xi^m=F(\xi)$, где m — произвольное целое число, а $F(\xi)$ — аналитическая функция от ξ, заданная разложением Лагранжа (см. задачу 25) [877],

$$F(\xi)=F(\zeta)+\sum_{n=1}^{\infty}\frac{w^n}{n!}\frac{\partial^{n-1}}{\partial\zeta^{n-1}}\{F'(\zeta)\,[g(\zeta)]^n\}. \tag{4.84}$$

Наконец, пусть

$$\left(\sum_{j=1}^{\infty}c_j\zeta^{j-1}\right)^n\equiv\sum_{j=0}^{\infty}b_{nj}\zeta^j; \tag{4.85}$$

тогда получим

$$\xi^m=\zeta^m+m\sum_{n=1}^{\infty}\frac{\rho}{n!}\sum_{j=0}^{\infty}\frac{(j+2n+m-1)!}{(j+n+m)!}\,b_{nj}\zeta^{j+2n+m}. \tag{4.86}$$

Необходимо найти обратное преобразование Лапласа этого выражения и использовать результат для вычисления обратного преобразования функции $P_0^*(s)$, для которой при разложении в ряд по степеням ξ потребуется вычисление обратного преобразования для всех членов ряда. Поэтому достаточно написать обратное преобразование правой части равенства (4.86) и, воспользовавшись им, найти $P_0(t)$. Первое преобразование имеет вид

$$\frac{m}{t}\left[\frac{(\mu t)^m}{m!}+\sum_{n=1}^{\infty}\frac{\rho^n}{n!}\sum_{j=0}^{\infty}b_{nj}\frac{(\mu t)^{j+2n+m}}{(j+n+m)!}\right]e^{-(1+\rho)\mu t}. \tag{4.87}$$

Отсюда

$$P_0(t)=\frac{1}{\mu t}\sum_{m=i+1}^{\infty}m\left[\frac{(\mu t)^m}{m!}+\sum_{n=1}^{\infty}\frac{(\rho\mu t)^n}{n!}\sum_{j=0}^{\infty}b_{nj}\frac{(\mu t)^{j+n+m}}{(j+n+m)!}\right]e^{-(1+\rho)\mu t}. \tag{4.88}$$

Кроме того, математическое ожидание состояния системы равно

$$L(t)=\left.\frac{\partial P}{\partial z}\right|_{z=1}=i+\left(\lambda\sum_{j=1}^{\infty}jc_j-\mu\right)t+\mu\int_0^t P_0(x)\,dx, \tag{4.89}$$

откуда среднее время ожидания в момент времени t равно $\frac{L(t)}{\lambda}$. Аналогично находится дисперсия времени ожидания.

При $c_j = \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases}$

время обслуживания имеет распределение Пирсона III типа (или распределение Эрланга) и

$$b_{n,\,n(k-1)} = 1,\quad b_{nj} = 0,\quad j \neq n(k-1).$$

Лучак получил для этого случая выражение

$$P_0(t) = 2 \sum_{m=i+1}^{\infty} m\rho^{-\left[\frac{(m-1)}{(k+1)}\right]} \frac{I_m^k(r)}{r} e^{-(1+\rho)\mu t}, \tag{4.90}$$

где

$$r = 2\rho^{\frac{1}{(1+k)}} \mu t, \quad I_m^k(r) = \sum_{n=0}^{\infty} \frac{\left(\frac{r}{2}\right)^{m+n(k+1)}}{n!\,(m+nk)!}.$$

Заметим, что при $k=1$ функция $I_m^k(r)$, свойства которой исследовал Лучак, есть обобщенная функция Бесселя первого рода. Читатель может попытаться найти $P_0(t)$ при $c_k = \alpha$, $c_{k+p} = \beta$ и $c_j = 0$, или, что то же самое, при $\alpha + \beta = 1$ для случая, когда

$$b_{n,\,qp+n(k-1)} = \binom{n}{q} \beta^q \alpha^{n-q},\quad q = 0, 1, \ldots, n.$$

2. Распределение длительности периода занятости

Период занятости начинается с момента поступления требования в свободный канал (т. е. начальное число требований i, ожидающих в момент времени $t=0$, равно единице) и оканчивается в тот момент, когда канал снова освобождается. Это приводит к системе, у которой E_0 — поглощающее состояние, т. е. как только это состояние достигается, процесс прекращается. Чтобы найти распределение длительности периода занятости, вычислим $\frac{dP_0}{dt}$. Это выражение для скорости изменения во времени вероятности пребывания системы в состоянии E_0 и есть искомая функция распределения длительности периода занятости. Она находится с помощью системы уравнений

$$\begin{aligned} P_n'(t) &= \mu P_{n+1}(t) - (\lambda+\mu) P_n(t) + \lambda \sum_{j=1}^{n-1} c_j P_{n-j}(t),\quad n \geqslant 2, \\ P_1'(t) &= \mu P_2(t) - (\lambda+\mu) P_1(t), \\ P_0'(t) &= \mu P_1(t). \end{aligned} \tag{4.91}$$

Производящая функция равна $P(z, t) = \sum_{n=1}^{\infty} P_n(t) z^n$. Аналогично находим ее преобразование

$$P^*(z, s) = \frac{z^{i+1} - \mu z P_1^*(s)}{(\lambda+\mu+s) - \mu - \lambda z \sum_{j=1}^{\infty} c_j z^j}. \tag{4.92}$$

В данном случае $P_1^*(s)=\frac{\xi^i}{\mu}$, где ξ — нуль, о котором говорилось выше,

$$\frac{\partial P_0}{\partial t}=\mu P_1(t)=\frac{i}{t}\left[\frac{t^i}{i!}+\sum_{n=1}^{\infty}\frac{(\rho t\mu)^n}{n!}\sum_{j=0}^{\infty}b_{nj}\frac{(\mu t)^{j+n+i}}{(j+n+i)!}\,e^{-(1+\rho)\mu t}\right].$$

При

$$c_j=\begin{cases}1, & j=k,\\ 0, & j\neq k,\end{cases}$$

это выражение принимает вид

$$\frac{dP_0(t)}{dt}=2i\mu\rho^{-\left[\frac{(i-1)}{(k+1)}\right]}\frac{I_i^k(r)}{r}\,e^{-(1+\rho)\mu t} \tag{4.93}$$

При $i=1$ это будет функция распределения длительности периода занятости.

3. Распределение времени ожидания

Пусть фазовая загрузка системы равна

$$\rho=\frac{\lambda}{\mu}\sum_{j=1}^{\infty}jc_j.$$

Обозначим через $w(x, t)dx$ вероятность того, что обслуживание требования, поступающего в момент времени t, закончится в промежутке $(x, x+dx)$. Если требования, поступившие до момента времени t, должны пройти n фаз обслуживания, то время ожидания данного требования равно сумме n независимых экспоненциально распределенных случайных величин; соответствующую вероятность необходимо умножить на вероятность $P_n(t)$ присутствия этих случайных величин. Затем следует произвести суммирование по всем n. Таким образом,

$$w(x, t)\,dx=\sum_{n=1}^{\infty}P_n(t)\,e^{-\mu x}\frac{(\mu x)^{n-1}}{(n-1)!}\mu dx+P_0(t)\,\delta(x)\,dx, \tag{4.94}$$

где $\delta(x)$ — дельта-функция.

Заметим, что распределение времени обслуживания одного требования (включая все фазы) равно

$$b(\tau)\,d\tau=\sum_{j=1}^{\infty}c_j e^{-\mu\tau}\frac{(\mu\tau)^{j-1}}{(j-1)!}\mu d\tau, \tag{4.95}$$

так как для каждого требования существует вероятность c_j того, что потребуется j фаз обслуживания. Обозначим преобразование Лапласа функции $b(\tau)$ через $b^*(s)$.

Введем производящую функцию для (4.82) и примем, что аргумент данной функции равен $z=\frac{\mu}{\mu+s}$, где s — параметр преобразования Лапласа $w^*(s, t)$ функции (4.94). Имеем

$$\frac{\partial w^*}{dt}+\{\lambda[1-b^*(s)]-s\}w^*=-sP_0(t), \tag{4.96}$$

откуда для стационарного состояния получим

$$w^*(s)=\frac{p_0}{1-\frac{\lambda}{s[1-b^*(s)]}},\quad p_0=1-\rho.$$

Если для каждого требования необходимо k фаз обслуживания, то $c_j=\delta_{ik}$, где δ_{ik} — символ Кронекера, и время обслуживания каждого требования имеет распределение Пирсона III типа:

$$b(\tau)\,d\tau=\mathrm{e}^{-\mu\tau}\frac{(\mu\tau)^{k-1}}{(k-1)!}\mu d\tau, \tag{4.97}$$

и

$$b^*(s)=\left[\frac{\mu}{\mu+s}\right]^k.$$

Пусть

$$\rho=\frac{\lambda k}{\mu}.$$

Тогда

$$w^*(s)=\frac{1-\rho}{1-\frac{\lambda}{s\left\{1-\left[\frac{\mu}{\mu+s}\right]^k\right\}}}. \tag{4.98}$$

Разложив преобразование Лапласа в ряд по степеням s, можно получить начальные моменты μ_i:

$$w^*(s)=\int_0^\infty \mathrm{e}^{-sx}w(x)\,dx=1-s\mu_1+\frac{s_2}{2!}\mu_2-\dots. \tag{4.99}$$

Следовательно, если разложить в ряд функцию (4.98), то, приняв математическое ожидание времени обслуживания равным единице и введя параметр a, получим

$$W\equiv\mu_1=\frac{a+1}{2a}\,\frac{\rho}{1-\rho}, \tag{4.100}$$

$$\sigma^2=\frac{a+1}{a}\,\frac{\rho}{1-\rho}\left(\frac{a+1}{4a}\,\frac{\rho}{1-\rho}+\frac{a+2}{3a}\right).$$

При $a=1$ и $k=1$ имеем экспоненциальное распределение времени обслуживания, а при $a=\infty$ и $k=\infty$ время обслуживания будет постоянным.

Упражнение. Найдите $w^*(s)$, W и σ^2 при

$$c_j = \frac{1-b}{b} b^j, \ j = 1, 2, \ldots,$$

и при

$$c_j = \frac{a^j e^{-a}}{(1-e^{-a}) j!}, \ j = 1, 2, \ldots .$$

Найденный результат сравните с результатом Гейвера [283].

Упражнение. Покажите, что преобразование Лапласа дельта-функции тождественно равно единице (см. § 4.8).

4.7. Случай 6. Несколько параллельных каналов; $\lambda_n = \lambda$; $\mu_n = n\mu$, $n \leqslant c$; $\mu_n = c\mu$, $c \leqslant n$

1. Решение уравнений

Допустим, что в момент времени $t=0$ в системе, состоящей из c параллельных каналов, ожидает i требований. Образуется общая очередь, требования поступают в свободные обслуживающие устройства по принципу «первым пришел — первым обслужен». Такая система описывается следующими уравнениями:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t),$$

$$\frac{dP_n(t)}{dt} = -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \ 1 \leqslant n < c,$$

$$\frac{dP_n(t)}{dt} = -(\lambda + c\mu) P_n(t) + \lambda P_{n-1}(t) + c\mu P_{n+1}(t), \ c \leqslant n, \tag{4.101}$$

с начальными условиями $P_n(0) = \delta_{in}$. Интенсивность обслуживания пропорциональна n (числу требований, ожидающих в очереди) при $n < c$ и пропорциональна c при $n \geqslant c$. Перепишем уравнения в следующем виде:

$$P_0'(t) = -(\lambda + c\mu) P_0(t) + c\mu P_0(t) + c\mu P_1(t) - (c-1)\mu P_1(t),$$

$$P_n'(t) = -(\lambda + c\mu) P_n(t) + (c-n)\mu P_n(t) + \lambda P_{n-1}(t) +$$
$$+ c\mu P_{n+1}(t) - (c-n-1)\mu P_{n+1}(t), \ 1 \leqslant n < c,$$

$$P_n'(t) = -(\lambda + c\mu) P_n(t) + \lambda P_{n-1}(t) + c\mu P_{n+1}(t), \ c \leqslant n. \tag{4.102}$$

Введем производящую функцию

$$P(z, t) \equiv \sum_{n=0}^{\infty} P_n(t) z^n, \ P(z, 0) = z^i. \tag{4.103}$$

Затем умножим второе и третье уравнения на z^n и просуммируем по всем n, включая и первое уравнение. Введя обозначение $P \equiv P(z, t)$, получим

$$\frac{\partial P}{\partial t} = -(\lambda + c\mu) P + \lambda z P + c\mu \frac{P - P_0(t)}{z} + c\mu P_0(t) -$$
$$- (c-1)\mu(1-z) P_1(z) - (c-2)\mu z(1-z) P_2(t) - \ldots -$$
$$- (c-n)\mu z^{n-1}(1-z) P_n(t) - \ldots - \mu z^{c-2}(1-z) P_{c-1}(t). \quad (4.104)$$

Возьмем преобразование Лапласа. Произведя группировку членов и решив уравнение относительно $P^*(z, s)$, получим

$$P^*(z, s) = \frac{z^{i+1} - \mu(1-z)\sum\limits_{n=0}^{c-1}(c-n) z^n P_n^*(s)}{sz - (1-z)(c\mu - \lambda z)}. \quad (4.105)$$

Применим теорему Руше к знаменателю правой части этого выражения, два нуля которого внутри единичного круга $|z| = 1$ и на его границе равны

$$\alpha_k = \frac{\lambda + c\mu + s \pm \sqrt{(\lambda + c\mu + s)^2 - 4c\mu k}}{2\lambda}, \; k = 1, 2. \quad (4.106)$$

Будем считать, что α_1 имеет положительный знак перед радикалом.

Так как функция P^* должна быть конечной внутри единичного круга $|z| = 1$ и на его окружности и так как $|\alpha_2| < 1$, то числитель должен иметь множитель $(z - \alpha_2)$, т. е. при $z = \alpha_2$ он обращается в нуль. Имеем

$$\sum_{n=0}^{c-1}(c-n)\alpha_2^n P_n^* = \frac{\alpha_2^{i+1}}{\mu(1-\alpha_2)}. \quad (4.107)$$

Чтобы определить P_n^* $(0 \leqslant n \leqslant c-1)$ и тем самым иметь возможность определить в явном виде P_n^* при $n \geqslant c$, потребуется c уравнений, содержащих c неизвестных P_n^* $(0 \leqslant n \leqslant c-1)$. Одно из них — уравнение (4.107), а остальные $(c-1)$ — уравнения (4.101). Заметим, что без уравнения (4.107) нельзя решить c уравнений, так как последнее из них не будет определено. Так как P_n^* определено, то $P_n(t)$ можно найти с помощью формулы обратного преобразования

$$P_n(t) = \frac{1}{2\pi i}\int\limits_{a-i\infty}^{a+i\infty} e^{st} P_n^*(s)\, ds.$$

Эти уравнения имеют вид

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t),$$
$$P_n'(t) = -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \quad (4.108)$$
$$1 \leqslant n \leqslant c-2.$$

Заметим, что второе уравнение справедливо только при $c \geqslant 3$. Случай $c=2$ нужно рассмотреть особо. Как было показано, при $c=1$ требуется только одно уравнение (4.107), с помощью которого определяется P_0^*. В данном случае задача состоит в том, чтобы, используя (4.107) и (4.108), найти преобразование Лапласа функции $P_n(t)$ $(0 \leqslant n \leqslant c-1)$ в явном виде. Используя (4.108), выражаем $P_n^*(s)$ $(0 \leqslant n \leqslant c-2)$ через $P_{c-1}^*(s)$, а затем, используя (4.107), определяем $P_{c-1}^*(s)$.

Введем производящую функцию

$$Q(z,\ t)=\sum_{n=0}^{c-2} P_n(t)\, z^n. \tag{4.109}$$

Допустим, что в момент времени $t=0$ начальное число требований $i > c-2$. Как будет показано далее, приведенное доказательство можно применить также и для случая $i \leqslant c$.

Умножив второе уравнение системы (4.108) на z^n, просуммировав по n и прибавив первое уравнение, получим

$$\frac{\partial Q}{\partial t}=-\lambda Q-\mu z\frac{\partial Q}{\partial z}+\mu\frac{\partial Q}{\partial z}+\lambda z Q-\lambda z^{c-1}P_{c-2}+\mu(c-1)\,z^{c-2}P_{c-1},$$

или в стандартной форме записи для линейного дифференциального уравнения первого порядка в частных производных

$$-\frac{\partial Q}{\partial t}+\mu(1-z)\frac{\partial Q}{\partial z}=\lambda(1-z)\,Q+ \\ +\lambda z^{c-1}P_{c-2}-\mu(c-1)\,z^{c-2}P_{c-1}. \tag{4.110}$$

Соответствующие уравнения Лагранжа имеют вид

$$\frac{dt}{-1}=\frac{dz}{\mu(1-z)}=\frac{dQ}{\lambda(1-z)\,Q+\lambda z^{c-1}P_{c-2}-\mu(c-1)\,z^{c-2}P_{c-1}}. \tag{4.111}$$

Из первого уравнения находим

$$c_1 e^{\mu t}=1-z;\ c_1=(1-z)\,e^{-\mu t};\ z=1-c_1 e^{\mu t}. \tag{4.112}$$

Подставив это значение z в третий член и рассмотрев уравнение, состоящее из второго и третьего членов, найдем

$$\frac{\partial Q}{\partial t}+\lambda c_1 e^{\mu t} Q=\mu(c-1)(1-c_1 e^{\mu t})^{c-2}P_{c-1}- \\ -\lambda(1-c_1 e^{\mu t})^{c-1}P_{c-2} \tag{4.113}$$

и

$$Q=e^{-\left(\frac{\lambda}{\mu}\right)c_1 e^{\mu t}}\int_0^t [\mu(c-1)(1-c_1 e^{\mu x})^{c-2}P_{c-1}(x)- \\ -\lambda(1-c_1 e^{\mu x})^{c-1}P_{c-2}(x)]\, e^{\left(\frac{\lambda}{\mu}\right)c_1 e^{\mu x}}\,dx. \tag{4.114}$$

Заметим, что постоянная интегрирования $c_2=f(c_1)=f[(1-z)e^{-\mu t}]$ для некоторой функции $f(c_1)$, которую нужно определить, обращается в нуль, так как $Q(z,\ 0)=0$. Это условие не выполняется при $i\leqslant c-2$, так как в этом случае

$$Q(z,\ 0)=z^i,\ f(1-z)=z^i e^{\left(\frac{\lambda}{\mu}\right)(1-z)}.$$

Пусть $y=1-z$, тогда $z=1-y$,

$$f(y)=e^{\left(\frac{\lambda}{\mu}\right)y}(1-y)^i$$

и

$$f[(1-z)e^{-\mu t}]=[1-(1-z)e^{-\mu t}]^i \exp\left[\left(\frac{\lambda}{\mu}\right)(1-z)e^{-\mu t}\right].$$

В этом случае, подставив значение c из (4.112), получим

$$Q(z,\ t)=e^{-\left(\frac{\lambda}{\mu}\right)(1-z)}\int_0^t \{\mu(c-1)\left[1-(1-z)e^{-\mu(t-x)}\right]^{c-2}P_{c-1}(x)-$$
$$-\lambda\left[1-(1-z)e^{-\mu(t-x)}\right]^{c-1}P_{c-2}(x)\}\exp\left[\frac{\lambda}{\mu}(1-z)e^{-\mu(t-x)}\right]dx+$$
$$+\exp\left[-\frac{\lambda}{\mu}(1-z)(1-e^{-\mu t})\right][1-(1-z)e^{-\mu t}]^i. \qquad (4.115)$$

Продолжая решение задачи для случая $i>c-2$, опустим второе выражение в правой части. Заметим, что первое выражение есть свертка, это удобно для применения преобразования Лапласа, к которому мы впоследствии обратимся.

Использовав результат из книги Эрдейи и др. ([218], стр. 147, формула 40) при $\nu=1$ и $\mu=-(c-2)$ после подстановки $y=\mu w$ и $dw=\frac{dy}{\mu}$ для $\operatorname{Re}(s)>0$ найдем:

$$\int_0^\infty e^{-sw}\left[1-(1-z)e^{-\mu w}\right]^{c-2}\exp\left[\frac{\lambda}{\mu}(1-z)e^{-\mu w}\right]dw=$$
$$=\int_0^\infty e^{-\left(\frac{s}{\mu}\right)y}\left[1-(1-z)e^{-y}\right]^{c-2}\exp\left[\frac{\lambda}{\mu}(1-z)e^{-y}\right]\frac{dy}{\mu}=$$
$$=\frac{1}{\mu}\,\frac{\Gamma\left(\frac{s}{\mu}\right)}{\Gamma\left(\frac{s}{\mu}+1\right)}\Phi_1\left[\frac{s}{\mu},\ -(c-2),\ 1,\ (1-z),\ \frac{\lambda}{\mu}(1-z)\right]. \qquad (4.116)$$

Здесь использовано гипергеометрическое распределение, имеющее вид

$$\Phi_1(\alpha,\ \beta,\ \gamma,\ x,\ y)=\sum_{n=0}^\infty\sum_{m=0}^\infty\frac{(\alpha)_{m+n}(\beta)_n}{(\gamma)_{m+n}\,m!\,n!}x^m y^n,$$
$$(\alpha)_0\equiv 1,\ (\alpha)_n\equiv\alpha(\alpha+1)\ldots(\alpha+n-1).$$

Отсюда

$$Q^*(z, s) = e^{-\left(\frac{\lambda}{\mu}\right)(1-z)} \left\{ (c-1) P^*_{c-1}(s) \frac{\Gamma\left(\frac{s}{\mu}\right)}{\Gamma\left(\frac{s}{\mu}+1\right)} \Phi_1 \left[\frac{s}{\mu}, -(c-2), 1, (1-z), \frac{\lambda}{\mu}(1-z)\right] - \right.$$

$$\left. - \frac{\lambda}{\mu} P^*_{c-2}(s) \frac{\Gamma\left(\frac{s}{\mu}\right)}{\Gamma\left(\frac{s}{\mu}+1\right)} \Phi_1 \left[\frac{s}{\mu}, -(c-1), 1, (1-z), \frac{\lambda}{\mu}(1-z)\right] \right\}. \quad (4.117)$$

Вначале путем дифференцирования найдем $P^*_{c-2}(s)$, выразим $P^*_{c-2}(s)$ через $P^*_{c-1}(s)$, а затем полученное значение $P^*_{c-2}(s)$ подставим в (4.117). Наконец, найдем функцию $P^*_n(s)$ $(0 \leqslant n \leqslant c-2)$, выраженную через $P^*_{c-1}(s)$. Затем полученное выражение подставим в (4.107) и найдем $P^*_{c-1}(s)$.

Имеем

$$P^*_{c-2}(s) = \frac{1}{(c-2)!} \frac{\partial^{c-2} Q^*}{\partial z^{c-2}} \bigg|_{z=0}.$$

Пусть

$$f(z) = e^{-\left(\frac{\lambda}{\mu}\right)(1-z)},$$

$$g(z) = \Phi_1 \left[\frac{s}{\mu}, -(c-2), 1, (1-z), \frac{\lambda}{\mu}(1-z)\right].$$

Для нахождения производной воспользуемся теоремой Лейбница. Получим

$$\frac{1}{(c-2)!} \frac{\partial^{c-2}}{\partial z^{c-2}} f(z) g(z) \bigg|_{z=0} =$$

$$= \frac{1}{(c-2)!} \sum_{k=0}^{c-2} \binom{c-2}{k} \left(\frac{\lambda}{\mu}\right)^{c-k-2} e^{-\frac{\lambda}{\mu}} \frac{d^k}{dz^k} \Phi_1 \left[\frac{s}{\mu}, -(c-2), 1, (1-z), \frac{\lambda}{\mu}(1-z)\right] \bigg|_{z=0}. \quad (4.118)$$

Таким образом,

$$P^*_{c-2}(s) = \frac{\Gamma\left(\frac{s}{\mu}\right)}{(c-2)!\,\Gamma\left(\frac{s}{\mu}+1\right)} \left\{ (c-1) P^*_{c-1}(s) \sum_{k=0}^{c-2} \binom{c-k}{k} \times \right.$$

$$\times \left(\frac{\lambda}{\mu}\right)^{c-k-2} e^{-\left(\frac{\lambda}{\mu}\right)} \frac{d^k}{dz^k} \Phi_1 \left[\frac{s}{\mu}, -(c-2), 1, (1-z), \frac{\lambda}{\mu}(1-z)\right] \bigg|_{z=0} -$$

$$\left. - \frac{\lambda}{\mu} P^*_{c-2}(s) \sum_{k=0}^{c-2} \binom{c-2}{k} \left(\frac{\lambda}{\mu}\right)^{c-k-2} e^{-\left(\frac{\lambda}{\mu}\right)} \frac{d^k}{dz^k} \Phi_1 \left[\frac{s}{\mu}, -(c-1), 1, (1-z), \frac{\lambda}{\mu}(1-z)\right] \bigg|_{z=0} \right\}. \quad (4.119)$$

Решив это уравнение относительно $P^*_{c-2}(s)$, получим

$$P^*_{c-2}(s)=$$

$$=\frac{\dfrac{\Gamma\left(\frac{s}{\mu}\right)(c-1)}{(c-2)!\,\Gamma\left(\frac{s}{\mu}+1\right)}P^*_{c-1}(s)\sum_{k=0}^{c-2}\binom{c-2}{k}\left(\frac{\lambda}{\mu}\right)^{c-k-2}e^{-\left(\frac{\lambda}{\mu}\right)}\times\left(\frac{d^k}{dz^k}\right)\Phi_1\left[\left(\frac{s}{\mu}\right),\,-(c-2),\,1,\,(1-z),\,\frac{\lambda}{\mu}(1-z)\right]\Big|_{z=0}}{1+\left(\frac{\lambda}{\mu}\right)\dfrac{\Gamma\left(\frac{s}{\mu}\right)}{(c-2)!\,\Gamma\left(\frac{s}{\mu}+1\right)}\sum_{k=0}^{c-2}\binom{c-2}{k}\left(\frac{\lambda}{\mu}\right)^{c-k-2}e^{-\left(\frac{\lambda}{\mu}\right)}\times\left(\frac{d^k}{dz^k}\right)\Phi_1\left[\left(\frac{s}{\mu}\right),\,-(c-1),\,1,\,(1-z),\,\left(\frac{\lambda}{\mu}\right)(1-z)\right]\Big|_{z=0}}. \quad (4.120)$$

При $n\leqslant c-2$ имеем

$$P^*_n(s)=\frac{\Gamma\left(\frac{s}{\mu}\right)}{n!\,\Gamma\left(\frac{s}{\mu}+1\right)}\left\{(c-1)P^*_{c-1}(s)\sum_{k=0}^{n}\binom{n}{k}\left(\frac{\lambda}{\mu}\right)^{n-k}e^{-\left(\frac{\lambda}{\mu}\right)}\times\right.$$

$$\times\frac{d^k}{dz^k}\Phi_1\left[\frac{s}{\mu},\,-(c-2),\,1,\,(1-z),\,\frac{\lambda}{\mu}(1-z)\right]\Big|_{z=0}-$$

$$-\frac{\lambda}{\mu}P^*_{c-2}(s)\sum_{k=0}^{n}\binom{n}{k}\left(\frac{\lambda}{\mu}\right)^{n-k}e^{-\left(\frac{\lambda}{\mu}\right)}\times$$

$$\left.\times\frac{d^k}{dz^k}\Phi_1\left[\frac{s}{\mu},\,-(c-1),\,1,\,(1-z),\,\frac{\lambda}{\mu}(1-z)\right]\Big|_{z=0}\right\}. \quad (4.121)$$

Таким образом,

$$P^*_{c-1}(s)=\frac{\alpha_2^{i+1}}{1-\alpha_2}\div\left(\alpha_2^{c-1}+\sum_{n=0}^{c-3}(c-n)\alpha_2^n\frac{\Gamma\left(\frac{s}{\mu}\right)}{n!\,\Gamma\left(\frac{s}{\mu}+1\right)}\right)\times$$

$$\times\left\{(c-1)\sum_{k=0}^{n}\binom{n}{k}\left(\frac{\lambda}{\mu}\right)^{n-k}e^{-\left(\frac{\lambda}{\mu}\right)}\times\right.$$

$$\times\frac{d^k}{dz^k}\Phi_1\left[\frac{s}{\mu},\,-(c-2),\,1,\,(1-z),\,\frac{\lambda}{\mu}(1-z)\right]\Big|_{z=0}-$$

$$-\frac{\lambda}{\mu}\frac{P^*_{c-2}(s)}{P^*_{c-1}(s)}\sum_{k=0}^{n}\binom{n}{k}\left(\frac{\lambda}{\mu}\right)^{n-k}e^{-\left(\frac{\lambda}{\mu}\right)}\times$$

$$\left.\times\frac{d^k}{dz^k}\Phi_1\left[\frac{s}{\mu},\,-(c-1),\,1,\,(1-z),\,\frac{\lambda}{\mu}(1-z)\right]\Big|_{z=0}\right\}+$$

$$\left.+2\alpha_2^{c-2}\frac{P^*_{c-2}(s)}{P^*_{c-1}(s)}\right). \quad (4.122)$$

Теперь это выражение необходимо подставить в (4.120) и (4.121), чтобы получить в явном виде значение $P_n^*(s)$ при $0 \leqslant n \leqslant c-2$. Возвращаясь к (4.105), заметим, что $P_n^*(s)$ для остальных значений n можно найти как коэффициент при z^n. Чтобы выполнить это, запишем знаменатель в виде $-\lambda(z-\alpha_1)\times\times(z-\alpha_2)$ и применим к произведению $(z-\alpha_1)^{-1}(z-\alpha_2)^{-1}$ теорему Лейбница. Имеем

$$\frac{1}{m!}\frac{d^m}{dz^m}(z-\alpha_1)^{-1}(z-\alpha_2)^{-1}\Big|_{z=0} = \frac{1}{m!}\sum_{k=0}^{m}\binom{m}{k}\times$$

$$\times(-1)^k k!\,(z-\alpha_1)^{-(k+1)}(-1)^{m-k}(m-k)!\,(z-\alpha_2)^{-(m-k-1)}\Big|_{z=0} =$$

$$= \frac{\lambda}{c\mu\alpha_2^{m-2}}\,\frac{1-\left(\frac{\alpha_2}{\alpha_1}\right)^{m+1}}{1-\left(\frac{\alpha_2}{\alpha_1}\right)}. \tag{4.123}$$

При $n \geqslant c$ коэффициент при z^n найдем, взяв сумму коэффициентов, полученных при умножении знаменателя выражения (4.105) на z в соответствующей степени. Получим выражение

$$P_n^*(s) = \frac{\mu}{\lambda}\left[\sum_{j=0}^{c-1}(c-j)P_j^*(s)\frac{\lambda}{c\mu\alpha^{n-j-2}}\cdot\frac{1-\left(\frac{\alpha_2}{\alpha_1}\right)^{n-j+1}}{1-\frac{\alpha_2}{\alpha_1}} - \right.$$

$$\left. -(c-j)P_j^*(s)\frac{\lambda}{c\mu\alpha^{n-i-3}}\,\frac{1-\left(\frac{\alpha_2}{\alpha_1}\right)^{n-j}}{1-\frac{\alpha_2}{\alpha_1}}\right] -$$

$$-\frac{1}{c\mu\alpha_2^{n-i-3}}\,\frac{1-\left(\frac{\alpha_2}{\alpha_1}\right)^{n-i}}{1-\frac{\alpha_2}{\alpha_1}},\quad n \geqslant c. \tag{4.124}$$

(Чтобы избежать путаницы при суммировании, в (4.105) вместо n используем индекс j). Подставим в (4.124) полученное ранее выражение для $P_j^*(s)$ $(0 \leqslant j \leqslant c-1)$.

Из определения преобразования Лапласа как производящей функции моментов можно найти моменты путем разложения в ряд в окрестности точки $s=0$.

2. Стационарное решение

Вероятность p_n для стационарного случая можно найти непосредственно из уравнений, приравняв производные нулю, или из соотношения

$$\lim_{t\to\infty} P_n(t) = \lim_{s\to 0} sP_n^*(s).$$

Имеем (см. задачу 1)

$$p_n=\begin{cases}\dfrac{p_0}{n!}\left(\dfrac{\lambda}{\mu}\right)^n, & 1\leqslant n\leqslant c,\\[2mm] \dfrac{p_0}{c!}\left(\dfrac{\lambda}{c\mu}\right)^n c^c, & c<n,\end{cases} \tag{4.125}$$

$$\sum_{n=0}^{\infty} p_n=1,$$

$$p_0=\frac{1}{\sum\limits_{n=0}^{c-1}\dfrac{(c\rho)^n}{n!}+\dfrac{(c\rho)^c}{c!\,(1-\rho)}}, \tag{4.126}$$

где $\rho=\dfrac{\lambda}{c\mu}$.

Среднее число требований, находящихся в системе, и среднее число требований, ожидающих в очереди, равны, соответственно,

$$L=\frac{(\rho c)^{c+1}}{(c-1)!\sum\limits_{n=0}^{c}\left[\dfrac{(\rho c)^n}{n!}\right][(c-n)^2-n]} \tag{4.127}$$

и

$$L_q=\frac{\rho\,(c\rho)^c}{c!\,(1-\rho)^2}p_0. \tag{4.128}$$

Вероятность того, что в системе находится c или более требований, а следовательно, поступающие требования должны ожидать, равна

$$P(>0)=\frac{(c\rho)^c}{c!\,(1-\rho)}p_0. \tag{4.129}$$

И, наконец, вероятность того, что время ожидания больше t, составляет

$$P(>t)=\exp\left[-c\mu t\,(1-\rho)\right]P(>0). \tag{4.130}$$

Упражнение. Дайте вывод последних шести выражений.

Среднее время ожидания (исключая ожидание в процессе обслуживания) равно

$$W_q=\frac{L_q}{\lambda}.$$

Пример. Как и в случае одноканальной системы,

$$P(<t)=\sum_{n=0}^{c-1}p_n+\sum_{n=0}^{\infty}p_{c+n}W_n(t),$$

где $W_n(t)$ — распределение времени ожидания $(n+1)$-го требования.

Значение $W_n(t)$ найдем следующим образом. Вероятность того, что к моменту времени t закончится обслуживание требования (уже находящегося на обслуживании), равна $1-e^{-\mu t}$. Вероятность того, что к моменту времени t обслуживание будет все еще продолжаться, равна $e^{-\mu t}$. Вероятность того, что к моменту времени t все c каналов будут заняты теми же самыми требованиями, составляет $e^{-c\mu t}$. Вероятность того, что освободится любой из каналов, равна $1-e^{-c\mu t}$. Чтобы на обслуживание поступило $(n+1)$-е требование, подобное событие должно произойти n раз. Таким образом, необходимо взять n-кратную свертку функции $1-e^{-c\mu t}$. Следовательно, функция распределения времени ожидания $(n+1)$-го требования имеет вид

$$W_0(t)=1-e^{-c\mu t},$$

$$W_n(t)=\int_0^t W_{n-1}(x)\,d_x W_0(t-x)=1-e^{-c\mu t}\sum_{i=0}^{n}\frac{(c\mu t)^i}{i!},$$

что можно проверить с помощью метода индукции.

Упражнение. Выполните вычисления для нахождения $P(<t)$. Ограничьтесь получением результата для $c=1$.

3. Период занятости

Обычно, если имеется c каналов, то для нахождения распределения длительности периода занятости при любом промежуточном числе каналов в соответствующую точку следует поместить поглощающий барьер. Например, в случае $k\leqslant c$ каналов имеем

$$P'_{k-1}(t)=k\mu P_k(t),$$

$$P'_n(t)=-(\lambda+n\mu)P_n(t)+\lambda P_{n-1}(t)+ \\ +(n+1)\mu P_{n+1}(t),\quad k\leqslant n<c. \qquad (4.131)$$

Отбрасывая член $\lambda P_{k-1}(t)$ при $n=k$, получим

$$P'_n(t)=-(\lambda+c\mu)P_n(t)+\lambda P_{n-1}(t)+c\mu P_{n+1}(t),\quad c\leqslant n.$$

Начальное число требований i должно равняться рассматриваемому числу занятых каналов.

4.8. Случай 7. Поглощающие барьеры; $\lambda_0=0$, $\lambda_n=0$, $\mu_n=0$ при $n\geqslant N$ (модель размножения бактерий)

Допустим, что интенсивность пуассоновского входящего потока, поступающего в одноканальную систему, пропорциональна числу уже имеющихся требований, т. е. $\lambda_n=n\lambda$. Допустим также, что время обслуживания имеет экспоненциальное распределение с интенсивностью обслуживания, также пропорциональной числу имеющихся требований, т. е. $\mu_n = n\mu$. Предположим, что как только число требований становится равным N (N — очень большое число), процесс сразу же прекращается. Предположим также, что, кроме того, процесс прекращается, как только система освобождается. Таким образом, E_N и E_0— поглощающие состояния. В уравнениях процесса рождения и гибели оба эти условия выражены равенствами $\lambda_0 = 0$ и $\lambda_n = 0$, $\mu_n = 0$ при $n \geqslant N$.

Допустим, что нас интересует $Q_n(t)$ — функция распределения времени, необходимого для того, чтобы система перешла в состояние E_N, если в момент времени $t=0$ она находилась в состоянии E_n; будем считать, что $\lambda > \mu$. Случай малых значений N [747] рассматривается в следующей главе. Найдем распределение времени, необходимого для превращения начальной дозы бактерий объема n, имеющейся в момент времени $t=0$, до инкубационного объема N. Предположим, что скорость размножения и гибели бактерий в организме пропорциональна имеющемуся числу их.

По определению,

$$Q_n(t) = \frac{dP_{nN}(t)}{dt}. \qquad (4.132)$$

Подставив это выражение в обратные уравнения процесса рождения и гибели, получим

$$Q'_n(t) = -n(\lambda + \mu) Q_n(t) + n\mu Q_{n-1}(t) + n\lambda Q_{n+1}(t),$$

$$n = 1, \ldots, N-1, \qquad (4.133)$$

$$Q_0(t) = Q_N(t) = 0,$$

$$Q_n(t) = \delta_{n,\,N-1}\lambda(N-1). \qquad (4.134)$$

Выражение (4.134) найдем из последнего уравнения системы

$$P'_{nN}(t) = (N-1)\lambda P_{n,\,N-1}(t), \qquad (4.135)$$

положив $t=0$ и воспользовавшись значением $Q_n(0)$.

Введя производящую функцию

$$Q(z, t) = \sum_{n=1}^{N-1} Q_n(t) z^{n-1},$$

получим линейное дифференциальное уравнение первого порядка в частных производных

$$-\frac{\partial Q}{\partial t} + [\mu z^2 - (\lambda + \mu) z + \lambda] \frac{\partial Q}{\partial z} = (\lambda + \mu -$$

$$- 2\mu z) Q + \mu N Q_{N-1}(t) z^{N-1}. \qquad (4.136)$$

Заметим, что в этом уравнении появляется величина $Q_{N-1}(t)$, которая является одной из неизвестных. Заметим также, что

$$Q_{N-1}(0) = \lambda (N-1).$$

Непосредственно интегрируя эти уравнения, можно показать, что

$$\int_0^{\infty} Q_n(t)\,dt = q_n = \frac{1-\left(\frac{\mu}{\lambda}\right)^n}{1-\left(\frac{\mu}{\lambda}\right)^N}. \tag{4.137}$$

Таким образом, путем интегрирования получим постоянную q_n, которая зависит от n; решая полученные уравнения в конечных разностях, находим значение интеграла.

В момент времени $t=0$ значение $Q_{N-1}(t)$ равно $\lambda(N-1)$. При больших N оно должно быстро уменьшаться таким образом, чтобы величина интеграла от этой функции не превосходила единицу. Действительно, если $N\to\infty$, то $Q_{N-1}(t)$ ведет себя как дельта-функция. Дельта-функция всюду равна нулю, кроме начальной точки, в которой она равна бесконечности. Интеграл от дельта-функции равен единице. Таким образом, из определения дельта-функции имеем

$$\int_0^{\infty} \delta(x) = 1,$$

$$\int_0^{t} f(x)\,\delta(x)\,dx = f(0), \tag{4.138}$$

если для некоторой функции $f(x)$ этот интеграл существует. Здесь употребляется правосторонняя дельта-функция.

Решая линейное дифференциальное уравнение в частных производных, получим

$$Q(z, t) = -N\mu\alpha^2 \int_0^{t} Q_{N-1}(s)\, e^{-\alpha(t-s)} \times$$

$$\times \frac{\{\lambda[1-e^{-\alpha(t-s)}] - z[\mu - e^{-\alpha(t-s)}]\}^{N-1}}{\{[\lambda - \mu e^{-\alpha(t-s)}] - \mu z[1-e^{-\alpha(t-s)}]\}^{N-1}}\, ds +$$

$$+ (N-1)\lambda\alpha^2 e^{-\alpha t} \frac{[\lambda(1-e^{-\alpha t}) - z(\mu - \lambda e^{-\alpha t})]^{N-2}}{[(\lambda - \mu e^{-\alpha t}) - \mu z(1-e^{-\alpha t})]^{N}}, \tag{4.139}$$

где $\alpha = \lambda - \mu$.

Учитывая замечание относительно значения $Q_{N-1}(s)$ при больших N, можно заменить интеграл значением подынтегральной функции при $s=0$. Затем можно найти

$$Q_n(t) = \frac{1}{(n-1)!} \frac{\partial^{n-1} Q(z, t)}{\partial z^{n-1}} \bigg|_{z=0},$$

и, воспользовавшись теоремой Лейбница для отыскания производной, вычислить производную от Q по z.

Получим довольно сложные выражения, которые можно упростить, имея в виду, что решение содержит множитель вида $N^p e^{-p\alpha t}$ (p — целое положительное число). Если N^p меньше $e^{-\alpha pt}$, то результат равен нулю; если же N^p больше $e^{-\alpha pt}$, то при больших N этот множитель велик. [Используя выражение для $Q_1(t)$, можно показать, что $\alpha t \approx \log N$.] Окончательно получим следующее решение исходной системы уравнений для больших N и $\beta = \frac{\mu}{\lambda}$ (см. задачу 21):

$$\frac{Q_n(t)}{\lambda} = (1-\beta)^3 \frac{(1-e^{-\alpha t})^{N-n-1}}{(1-\beta e^{-\alpha t})^{N+n}} \sum_{k=0}^{n-1} \binom{n}{k+1} \times$$

$$\times \beta^{n-k-1} (1-\beta)^{2k} \frac{(Ne^{-\alpha t})^{k+1}}{(k+1)!} =$$

$$= (1-\beta) \beta^n \exp\left(\frac{-\beta x}{1-x}\right) \frac{x}{n!} \frac{dL_n(-x)}{dx}, \qquad (4.140)$$

где

$$x = \frac{(1-\beta)^2}{\beta} Ne^{-\alpha t},$$

а $L_n(x)$ — полином Лагерра.

Теперь можно произвести вычисление и нормировку математического ожидания и дисперсии для $Q_n(t)$ [так как $Q_n(t)$ — условная вероятность]. Заметим, что

$$L_n(x) \equiv e^x \frac{d^n}{dx^n} (x^n - e^{-x}).$$

Замечание. В формуле (4.140) под знаком суммирования содержатся распределения экстремальных значений, одним из которых является хорошо известное распределение Фишера—Типпетта.

Формулировка и решение нескольких задач, рассмотренных в этой главе, обеспечит тем, кто взял на себя труд проследить за всеми выкладками, глубокое понимание вопроса и отсутствие затруднений при обращении к подобным задачам массового обслуживания. Задачи, приведенные в конце главы, дают хорошую возможность проверить свои успехи.

Задачи

1. Проверьте решения для следующих частных случаев стационарного процесса рождения и гибели и в каждом случае покажите справедливость указанного соответствия системам массового обслуживания.

а. Терпеливые клиенты

Неограниченный входящий поток			
Число каналов	$c=1$	$c>1$	$c=\infty$
Значения параметров	$\lambda_i=\lambda\ (i=0,\ 1,\ \ldots)$ $\mu_i=\mu\ (i=1,\ 2,\ \ldots)$	$\lambda_i=\lambda$ $\mu=\begin{cases} i\mu, & i\leqslant c, \\ c\mu, & i>c \end{cases}$	$\lambda_i=\lambda$ $\mu_i=i\mu$
Решение	$p_n=p_0\left(\frac{\lambda}{\mu}\right)^n$	$p_n=p_0\prod_{i=1}^{n}\frac{\lambda}{i\mu}=\frac{p_0}{n!}\left(\frac{\lambda}{\mu}\right)^n,\ 1\leqslant n\leqslant c$ $p_n=p_0\prod_{i=1}^{c}\frac{\lambda}{i\mu}\prod_{i=c+1}^{n}\frac{\lambda}{c\mu}=$ $=\frac{p_0}{c!}\left(\frac{\lambda}{\mu}\right)^n\frac{1}{c^{n-c}},\quad n\geqslant c$	$p_n=\frac{p_0}{n!}\left(\frac{\lambda}{\mu}\right)^n$

Продолжение

Ограниченный входящий поток[1]			
Число каналов	$c = 1$	$c > 1$	$c = \infty$
Значения параметров	$\lambda_i = (m - i)\lambda \ (i = 0, 1, \ldots, m)$ $\mu_i = \mu \ (i = 1, 2, \ldots, m)$	$\lambda_i = (m - i)\lambda$ $\mu_i = \begin{cases} i\mu, & 1 \leqslant i < c, \\ c\mu, & c \leqslant i \leqslant m \end{cases}$	$\lambda_i = (m - i)\lambda$ $\mu_i = i\mu$
Решение	$p_n = p_0 \prod\limits_{i=1}^{n} \frac{(m - i + 1)\lambda}{\mu} =$ $= p_0 \left(\frac{\lambda}{\mu}\right)^n \frac{m!}{(m - n)!}$	$p_n = p_0 \prod\limits_{i=1}^{n} \frac{(m - i + 1)\lambda}{i\mu} =$ $= p_0 \left(\frac{\lambda}{\mu}\right)^n \frac{m!}{n!\,(m - n)!}, \ 1 \leqslant n < c$ $p_n = p_0 \prod\limits_{i=1}^{c} \frac{(m - i + 1)\lambda}{i\mu} \prod\limits_{i=c+1}^{n} \frac{(m - i + 1)\lambda}{c\mu} =$ $= p_0 \left(\frac{\lambda}{\mu}\right)^n \binom{m}{n} \frac{n!}{c!\, c^{n-c}}, \ c \leqslant n \leqslant m,$ $n \geqslant c$	$p_n = p_0 \prod\limits_{i=1}^{n} \frac{(m - i + 1)\lambda}{i\mu} =$ $= p_0 \left(\frac{\lambda}{\mu}\right)^n \binom{m}{n}$

[1] Для ограниченного входящего потока λ является характеристикой требований, поступающих из определенного источника; эта величина постоянна в интервалах, когда обслуживается определенное число требований из данного источника.

б. Нетерпеливые клиенты

	Неограниченный входящий поток		
Число каналов	$c=1$	$c>1$	$c=\infty$
Значения параметров	$\lambda_i=\lambda$ $\mu_i=\mu+c_i$	$\lambda_i=\lambda$ $\mu_i=\begin{cases} i(\mu+c_i), & i\leqslant c, \\ c\mu, & i>c \end{cases}$	$\lambda_i=\lambda$ $\mu_i=i\mu$
Решение	$p_n=p_0\lambda^n\prod\limits_{i=1}^{n}\frac{1}{\mu+c_{i-1}}$	Найдите p_n	Найдите p_n
	Ограниченный входящий поток		
Значения параметров	$\lambda_i=(m-i)\lambda$ $\mu_i=\mu+c_i$	$\lambda_i=(m-i)\lambda$ $\mu_i=\begin{cases} i(\mu+c_i), & i\leqslant c \\ c(\mu+c_i), & i>c \end{cases}$	$\lambda_i=(m-i)\lambda$ $\mu_i=i(\mu+c_i)$
Решение	$p_n=p_0\lambda^n\prod\limits_{i=1}^{n}\frac{m-i+1}{\mu+c_i}$	Найдите p_n	Найдите p_n

Продолжение

в. Дополнительное поступление требований. Нетерпеливые клиенты

Неограниченный входящий поток			
Число каналов	$c = 1$	$c > 1$	$c = \infty$
Значения параметров	$\lambda_i = \lambda + d_i$ $\mu_i = \mu + c_i$	$\lambda_i = \lambda + d_i$ $\mu_i = \begin{cases} i(\mu + c_i), & i \leqslant c, \\ c\mu, & i > c \end{cases}$	$\lambda_i = \lambda + d_i$ $\mu_i = i(\mu + c_i)$
Решение	$p_n = p_0 \prod_{i=1}^{n} \frac{\lambda + d_{i-1}}{\mu + c_i}$	Найдите p_n	Найдите p_n
	Ограниченный входящий поток		
Значения параметров	$\lambda_i = (m - i)(\lambda + d_i)$ $\mu_i = \mu + c_i$	$\lambda_i = (m - i)(\lambda + d_i)$ $\mu_i = \begin{cases} i(\mu + c_i), & i \leqslant c, \\ c(\mu + c_i), & i > c \end{cases}$	$\lambda_i = (m - i)(\lambda + d_i)$ $\mu_i = i(\mu + c_i)$
Решение	Найдите p_n	Найдите p_n	Найдите p_n

2. Рассмотрев следующие показатели эффективности функционирования c-канальной системы $M/M/c$ при $\rho = \frac{\lambda}{c\mu}$ в стационарном состоянии, покажите, что:

а) Математическое ожидание числа требований, ожидающих в очереди, равно

$$\sum_{i=c}^{\infty}(i-c)\,p_i = p_c\,\frac{\rho}{(1-\rho)^2}.$$

б) Вероятность того, что заняты все c каналов, равна

$$\sum_{i=0}^{\infty} p_i = p_c\,\frac{1}{1-\rho}.$$

в) Вероятность того, что имеются требования, ожидающие начала обслуживания, равна

$$\sum_{i=c+1}^{\infty} p_i = p_c\,\frac{\rho}{1-\rho}.$$

г) Среднее число требований, ожидающих начала обслуживания, составляет

$$\frac{\sum\limits_{i=c+1}^{\infty}(i-c)\,p_n}{\sum\limits_{i=c+1}^{\infty} p_i} = \frac{1}{1-\rho}.$$

д) Среднее время ожидания в очереди для всех требований равно

$$\frac{1}{\lambda}\sum_{i=c}^{\infty}(i-c)\,p_i = p_c\,\frac{1}{c\mu(1-\rho)^2}.$$

е) Среднее время ожидания начала обслуживания для тех требований, которые действительно ожидают, составляет

$$\frac{1}{1-\rho}\cdot\frac{1}{c\mu} = \frac{1}{c\mu-\lambda}.$$

ж) Вероятность того, что в c-канальной системе занято ровно $n \leqslant c$ каналов, равна p. Определите p_0 из условия

$$\sum_{n=1}^{c} p_n = 1.$$

3. Покажите, что среднее время ожидания для требований, ожидающих начала обслуживания, равно $\frac{W_q}{P(>0)}$.

4. Покажите справедливость следующих формул и рассмотрите случай двухканальной системы:

а) среднее число обслуженных требований равно

$$\sum_{n=0}^{c-1} np_n + c\sum_{n=c}^{\infty} p_n,$$

б) среднее число свободных каналов равно c минус среднее число обслуживаемых требований;

в) коэффициент простоя каналов равен отношению среднего числа свободных каналов к общему числу каналов c;

г) коэффициент использования системы равен отношению среднего числа требований в системе к числу каналов;

д) если входящий поток конечен и общее число требований $m>c$, то, заменив в выражении для коэффициента использования системы c на m, получим коэффициент простоя требований вследствие обслуживания;

е) коэффициент простоя требований вследствие ожидания для случая, когда в системе находится m требований и все каналы заняты, получим, разделив среднее число требований, ожидающих начала обслуживания, на m;

ж) общий коэффициент простоя равен сумме коэффициентов простоя вследствие обслуживания и вследствие ожидания;

з) для случая конечной совокупности среднее число потерянных требований равно

$$\frac{\mu}{\lambda}\sum_{n=0}^{c-1} np_n + c\sum_{n=c}^{m} p_n.$$

Если совокупность состоит из станков, требующих обслуживания, то коэффициент использования станков получим, разделив эту величину на m.

5. Выведите закон Пуассона, рассмотрев одноканальную систему $M/M/1$ при $\mu=0$.

6. Для стационарного состояния двухканальной системы с пуассоновским входящим потоком с параметром λ и экспоненциальным временем обслуживания с параметром μ при обслуживании требований в порядке поступления найдите необходимое увеличение интенсивности обслуживания $\Delta\mu$, которое даст такой же эффект, что и добавление еще одного канала (с таким же распределением времени обслуживания), т. е. найдите $\Delta\mu$ при условии, что

$$W_q(3, \mu) = W_q(2, \mu + \Delta\mu).$$

В данном случае можно сравнить стоимость дополнительного обслуживающего устройства с расходами на увеличение объема обслуживания имеющимися устройствами. Каким образом?

7. Выведите закон Пуассона с помощью обратных уравнений

$$P'_{in}(t) = -\lambda P_{in}(t) + \lambda P_{i+1,\,n}(t), \quad P_{in}(0) = \delta_{in}.$$

Покажите, что производящая функция

$$P(z, t) \equiv \sum_{i=0}^{\infty} P_{in}(t)\, z^i$$

удовлетворяет уравнению

$$\frac{\partial P}{\partial t} + \lambda\left(1 - \frac{1}{z}\right)P = -\frac{\lambda}{z} P_{0n}(t).$$

Покажите, что математическое ожидание случайной величины с распределением $P_{in}(t)$ равно $\lambda t+i$, а дисперсия — λt.

8. Пусть X_1 и X_2 — случайные величины, которые распределены по закону Пуассона с параметрами λ_1 и λ_2. Покажите, что $Y=X_1-X_2$ имеет следующее распределение:

$$P(Y = x) = P(X_1 = X_2 + x) = \sum_{k=0}^{\infty} P(X_1 = k + x)\, P(X_2 = k) =$$

$$= e^{-(\lambda_1+\lambda_2)}\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{x}{2}} I_x\left(2\sqrt{\lambda_1\lambda_2}\right).$$

Покажите, что характеристическая функция этого распределения имеет вид

$$\varphi(t) = e^{(\lambda_1+\lambda_2)(\cos t - 1) + i(\lambda_1-\lambda_2)\sin t}, \quad i = \sqrt{-1},$$

математическое ожидание равно $\lambda_1 - \lambda_2$, а дисперсия равна $\lambda_1 + \lambda_2$. Пусть $Y = \sum_{i=1}^{n} a_i X_i$, где a_i равно $+1$ или -1, а величина X_i распределена по закону Пуассона с параметром λ_i. Покажите, что

$$P(Y = x) = \exp\left(-\sum_{j=1}^{n} \lambda_j\right)\left(\frac{\sum\limits_k \lambda_k}{\sum\limits_m \lambda_m}\right)^{\frac{x}{2}} I_x\left(2\sqrt{\sum_k \lambda_k \sum_m \lambda_m}\right),$$

где индекс j распространяется на все переменные, k — на переменные с положительным знаком, а m — на переменные с отрицательным знаком.

Введя функцию правдоподобия для n наблюдений, равную

$$e^{-n(\lambda_1+\lambda_2)}\left(\frac{\lambda_1}{\lambda_2}\right)^{\sum \frac{x_i}{2}} \prod_{i=1}^{n} I_{x_i}\left(2\sqrt{\lambda_1\lambda_2}\right),$$

покажите, что $\hat{\lambda}_1 - \hat{\lambda}_2 = \sum \frac{x_i}{n}$ есть оценка для математического ожидания m, а

$$\hat{\lambda}_1 + \hat{\lambda}_2 = \frac{\sum (x_i - m)^2}{n}$$

— оценка для дисперсии.

9. Найдите преобразование Лапласа решения задачи в случае двухканальной системы с одинаковым экспоненциальным временем обслуживания в каждом канале и покажите, что оно может быть выражено в виде разности двух функций, первая из которых есть производящая функция для одноканальной системы, в которой μ заменено на 2μ. Покажите, что [748]

$$P_0^*(s) = \frac{\alpha_2^{i+1}}{(1-\alpha_2)\left[2\mu + (s+\lambda)\alpha_2\right]},$$

$$P_1^*(s) = \frac{(s+\lambda)\alpha_2^{i+1}}{\mu(1-\alpha_2)\left[2\mu + (s+\lambda)\alpha_2\right]}.$$

Заметим, что в этом случае в выражения для корней α_1 и α_2 входит 2μ. Проверьте значения P_0 и P_1 для стационарного состояния.

10. Найдите производящую функцию и укажите способ решения задачи для двухканальной системы с экспоненциальным временем обслуживания с различными параметрами μ_1 и μ_2. Заметим, что в этих уравнениях необходимо принимать во внимание то обстоятельство, что поступающее требование может попасть в любой канал. Поэтому необходимо ввести $P_1(1, 0, t)$ и $P_1(0, 1, t)$ — вероятности того, что в системе имеется одно требование, которое находится соответственно в первом или втором каналах. Вероятность $P_1(t)$ равна их сумме.

11. Решите задачу для поглощающих барьеров, аналогичную приведенной в тексте, с той лишь особенностью, что в данном случае $\lambda_n = \lambda$ и $\mu_n = \mu$. Воспользуйтесь преобразованием Лапласа функции $Q(z, t)$, определение которой дано в последнем параграфе этой главы. Так как Q — полином, то преобразование Лапласа существует для всех z. Таким образом, два нуля знаменателя преобразования Лапласа $Q^*(z, s)$ сокращаются с соответствующими нулями числителя. Из двух уравнений, получаемых при этом, находятся значения $Q_{N-1}^*(s)$ и $Q_1^*(s)$. Подставив эти значения в выражение

$$Q^*(z, s) = \frac{\mu Q_{N-1}^* z^N - \lambda z^{N-1} + \lambda Q_1^*}{\mu z^2 - (\lambda + \mu + s) z + \lambda}$$

и произведя упрощения, получим

$$Q^*(z,\ s)=\frac{\lambda}{\mu}\left(\alpha_1^n-\alpha_2^N\right)^{-1}\sum_{n=1}^{N-1}\left(\frac{\lambda}{\mu}\right)^{N-n-1}\left(\alpha_1^n-\alpha_2^n\right)z^{n-1},$$

где α_1 и α_2 — нули.

Отсюда $Q_n^*(s)$ находится как коэффициент при z^{n-1}. Затем по способу, указанному в книге Эрдейи и др. [218], находим обратное преобразование.

12. Покажите, что решение уравнений

$$P_0'(t)=q\lambda P_1(t),$$

$$P_1'(t)=-\lambda P_1(t)+q\lambda P_2(t),$$

$$P_n'(t)=-\lambda P_n(t)+p\lambda P_{n-1}(t)+q\lambda P_{n-1}(t),\quad n\geqslant 2,$$

$(p+q=1)$, описывающих систему с поглощающим барьером в начальном состоянии и начальным числом требований i в момент времени $t=0$, для нестационарного случая имеет вид

$$P_n(t)=\mathrm{e}^{-\lambda t}\left(\sqrt{\frac{p}{q}}\right)^{n-i+1}\lambda\sqrt{pq}\,\left[I_{n-i}\left(2\lambda\sqrt{pq}\,t\right)-\right.$$
$$\left.-I_{n+i}\left(2\lambda\sqrt{pq}\,t\right)\right],\quad n\geqslant 0,$$

$$P_0(t)=\int_0^t \mathrm{e}^{-\lambda(t-u)}\left(\sqrt{\frac{p}{q}}\right)^{i}\frac{i}{t-u}I_k\left[2\lambda\sqrt{pq}\,(t-u)\right]du.$$

Перейдя к пределу, найдите решение для стационарного состояния.

13. Вычислив

$$\left.\frac{\partial P(z,\ t)}{\partial z}\right|_{z=1},$$

покажите, что в случае бесконечно большого числа каналов среднее число требований, находящихся в системе, равно

$$L=\frac{\lambda}{\mu}\left(1-\mathrm{e}^{-\mu t}\right)+i\mathrm{e}^{-\mu t}.$$

Для стационарного состояния

$$L=\frac{\lambda}{\mu}.$$

Вычислите также среднее число требований, находящихся в системе, с помощью новой производящей функции, учитывающей распределение начального числа требований. Заметим, что решение можно получить также, изменив выражение для L. Значение i заменяется его математическим ожиданием.

Предполагая, что место для ожидания ограничено и максимальное число требований равно N, введите в уравнения верхний поглощающий барьер и найдите производящую функцию для этого случая. Определите среднее число требований, находящихся в системе. Положив $N=0$, найдите соответствующее выражение для системы с потерями, а затем, приняв $N\to\infty$, найдите это выражение для системы с неограниченным местом для ожидания (эта величина получена выше).

14. Решите уравнения процессов рождения и гибели при

$$\lambda_n=(m-n)\lambda,\quad \mu_n=n\mu,\quad n=0,\ldots,\ m,$$

т. е. для случая ограниченного числа требований. Покажите, что решение дифференциального уравнения

$$\frac{\partial P}{\partial z} = \frac{m\lambda P}{\mu + \lambda z}$$

для производящей функции имеет вид

$$P(t, z) = \left(\frac{\mu + \lambda z}{\mu + \lambda}\right)^m.$$

Покажите, что $p_n = \binom{m}{n}\left(\frac{\lambda}{\lambda+\mu}\right)^n\left(\frac{\mu}{\lambda+\mu}\right)^{m-n}$ является решением уравнения для стационарного состояния.

15. Решив данную задачу, найдем распределение длительности периода занятости одноканальной системы. Решим задачу для случая, когда E_0 — поглощающее состояние. Положим, что в момент времени $t=0$ в системе находится i требований

$$P_0'(t) = \mu P_1(t),$$

$$P_1'(t) = -(\lambda + \mu) P_1(t) + \mu P_2(t),$$

$$P_n'(t) = -(\lambda + \mu) P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t), \quad n \geqslant 2.$$

Найдите следующее выражение для преобразованной по Лапласу производящей функции

$$P^*(z, s) = \frac{z^{i+1} - (1-z)(\mu - \lambda z)\left(\frac{\alpha_2^i}{s}\right)}{-\lambda(z - \alpha_1)(z - \alpha_2)},$$

где α_1 и α_2 — те же нули, что и в случае одноканальной системы.

Покажите, что

$$P_n'(t) = \left(\sqrt{\frac{\mu}{\lambda}}\right)^{i-n} e^{-(\lambda+\mu)t}\left[I_{i-n}(2\sqrt{\lambda\mu}\, t) - I_{i+n}(2\sqrt{\lambda\mu}\, t)\right], \quad n \geqslant 1,$$

$$P_0'(t) = \frac{i\left(\sqrt{\frac{\mu}{\lambda}}\right)^i e^{-(\lambda+\mu)t} I_i(2\sqrt{\lambda\mu}\, t)}{t}.$$

Период занятости начинается с момента поступления требования в свободный канал и оканчивается, когда этот канал снова освобождается. Это приводит к системе, в которой E_0 является поглощающим состоянием; таким образом, как только это состояние достигается, процесс прекращается. Заметим, что функция $P_0'(t)$ при $i=1$ описывает распределение длительности периода занятости для одноканальной системы. Она является мерой скорости достижения состояния E_0.

16. Для вывода распределения $Q_h(t)$ времени t, которое истекло до того момента, когда длина очереди в первый раз стала равной h $(h < i$, где i — начальное число требований в момент времени $t=0)$, рассмотрим следующий процесс, который обрывается, как только достигается состояние E_h. Система уравнений имеет вид

$$P_h'(t) = \mu P_{h+1}(t),$$

$$P_{h+1}'(t) = -(\lambda + \mu) P_{h+1}(t) + \mu P_{h+2}(t),$$

$$P_n'(t) = -(\lambda + \mu) P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t), \quad n \geqslant h + 2.$$

(Эти выражения можно получить непосредственно из задачи 15.) Покажите, что дифференциальное уравнение для производящей функции

$$P(z, t) = \sum_{n=h}^{\infty} P_n(t) z^n$$

имеет вид

$$z \frac{\partial P(z, t)}{\partial t} = (1 - z)(\mu - \lambda z)[P(z, t) - P_h(t) z^h],$$

а ее преобразование Лапласа выражается как

$$P^*(z, s) = \frac{z^{i+1} - z^h (1 - z)(\mu - \lambda z) P_h^*(s)}{sz - (1 - z)(\mu - \lambda s)}.$$

Покажите также, что

$$P_h(s) = \frac{\alpha_2^{i-h}}{s}$$

и

$$Q_h^*(s) = P_h^{*\prime}(s) = \alpha_2^{i-h},$$

откуда

$$Q_h(t) = \frac{(i-h)\left(\sqrt{\frac{\lambda}{\mu}}\right)^{i-h} e^{-(\lambda+\mu)t} I_{i-h}(2\sqrt{\lambda\mu}\, t)}{t}.$$

Распределение длительности периода занятости одноканальной системы получим, положив $i=1$, $h=0$. Первые два момента этого распределения равны

$$E(t) = \frac{i-h}{\mu - \lambda},$$

$$\sigma^2(t) = \frac{(i-h)(\mu+\lambda)}{(\mu-\lambda)^2}.$$

17. Распределение времени $Q_N(t)$, истекшего с того момента, когда в очереди находилось i требований, до момента, когда стало N требований (когда процесс обрывается), найдем, решив систему уравнений

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t),$$

$$P_n'(t) = -(\lambda+\mu) P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t), \quad 1 \leqslant n \leqslant N-2,$$

$$P_{N-1}'(t) = -(\lambda+\mu) P_{N-1}(t) + \lambda P_{N-2}(t),$$

$$P_N'(t) = \lambda P_{N-1}(t).$$

Пусть

$$P(z, t) = \sum_{n=0}^{N} P_n(t) z^n, \quad P(z, 0) = z^i.$$

Покажите, что

$$z \frac{\partial P}{\partial t} = (1 - z)\left\{(\mu - \lambda z)\left[P(z, t) - P_N(t) z^N\right] - \mu P_0(t)\right\}$$

и

$$P^*(z, s) = \frac{z^{i+1} - (1 - z)\left[\mu P_0^*(s) + (\mu - \lambda z) z^N P_N^*(s)\right]}{sz - (1 - z)(\mu - \lambda z)}.$$

Эта функция выражается полиномом относительно z и конечна для всех z, включая оба нуля знаменателя. Следовательно, $P_0(s)$ и $P_N(s)$ получим, приравняв числитель нулю и подставив значения обоих нулей α_1 и α_2 знаменателя (для каждого из которых числитель равен нулю). Имеем

$$Q_N^*(s) = P_N^{*\prime}(s) = sP_N^*(s) = \frac{\lambda\left(\alpha_2^{i+1} - \alpha_1^{i+1}\right) - \mu\left(\alpha_2^i - \alpha_1^i\right)}{\lambda\left(\alpha_2^{N+1} - \alpha_2^{N+1}\right) - \mu\left(\alpha_2^N - \alpha_1^N\right)}.$$

$$E(t) = \begin{cases} \dfrac{N-i}{\lambda-\mu} - \mu\left[\dfrac{\left(\frac{\mu}{\lambda}\right)^i - \left(\frac{\mu}{\lambda}\right)^N}{(\lambda-\mu)^2}\right] & \text{при } \dfrac{\lambda}{\mu} < 1,\ \dfrac{\lambda}{\mu} > 1, \\ \dfrac{N-i}{2\mu}(N+i+1) & \text{при } \dfrac{\lambda}{\mu} = 1. \end{cases}$$

18. Решите уравнения для процесса рождения и гибели при $\lambda_n = n\lambda$ и $\mu_n = n\mu$ (возрастание по линейному закону). В момент времени $t=0$ начальным состоянием является E_i. Заметим, что начальное состояние является поглощающим. Покажите, что производящая функция

$$P(z, t) = \sum_{n=0}^{\infty} P_n(t) z^n, \quad P(z, 0) = z^i,$$

удовлетворяет уравнению

$$\frac{\partial P}{\partial t} = (z-1)(\lambda z - \mu)\frac{\partial P}{\partial z}.$$

Решив это линейное дифференциальное уравнение первого порядка в частных производных, получим

$$P(z, t) = \left\{ \frac{\mu\left[1 - e^{(\lambda-\mu)t}\right] - z\left[\lambda - \mu e^{(\lambda-\mu)t}\right]}{\mu - \lambda e^{(\lambda-\mu)t} - \lambda z\left[1 - e^{(\lambda-\mu)t}\right]} \right\}^i.$$

Покажите, что при $i = 1$

$$P_n(t) = [1 - P_0(t)]\left[1 - \frac{\lambda - \lambda e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}}\right]\left[\frac{\lambda - \lambda e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}}\right]^{n-1}, \quad n \geqslant 1,$$

$$P_0(t) = \frac{\mu e^{(\lambda-\mu)t}}{\lambda e^{(\lambda-\mu)t} - \mu}.$$

Найдите $\lim\limits_{t\to\infty} P_n(t)$.

Покажите, что

$$L = E(n) = e^{(\lambda-\mu)t}, \quad \sigma^2(n) = \frac{\lambda+\mu}{\lambda-\mu} e^{(\lambda-\mu)t}\left[e^{(\lambda-\mu)t} - 1\right].$$

Заметим, что если $\lambda < \mu$, то при $t \to \infty$ $P(z, t) \to 1$ и процесс непременно обрывается. При $i = 1$ и $\mu = 0$

$$P_n(t) = e^{-\lambda t}\left(1 - e^{-\lambda t}\right)^{n-1}, \quad n \geqslant 1.$$

При $t \to \infty$ эта вероятность стремится к нулю, давая предельную вероятность обрыва процесса.

19. Пусть в уравнениях процесса рождения и гибели $\lambda_n = \lambda$ и $\mu_n = a + b(n-1)$, где a и b — постоянные. Покажите, что производящая функция удовлетворяет уравнению

$$\frac{\partial P}{\partial t} = (1-z)\left[(a - \lambda z)\frac{\partial P}{\partial t} + bz\frac{\partial^2 P}{\partial z^2}\right].$$

Если $a=0$, а λ и b положительны, то при $t\to\infty$ вероятность обрыва процесса равна нулю. Покажите также, что

$$E(n)=\frac{\frac{\lambda}{b}}{1-e^{-\frac{\lambda}{b}}}.$$

20. Для системы $M/M/1$ автокорреляционная функция длины очереди $n(t)$ согласно эргодической теореме (которая устанавливает эквивалентность усреднения по пространству реализаций и по времени для эргодического процесса), равна

$$\psi(t)=\lim\frac{1}{T}\int_0^T n(x)\,n(x+t)\,dx=\sum_{i=0}^{\infty}p_i\left[i\sum_{n=0}^{\infty}nP_{in}(t)\right]=$$

$$=\frac{\lambda^2}{(\mu-\lambda)^2}+(\mu-\lambda)\frac{\lambda\mu}{\pi}\int_0^{2\pi}\sin^2\theta\,\frac{e^{-wt}}{w^3}\,d\theta,$$

где $w=\mu+\lambda-2\sqrt{\lambda\mu}\cos\theta$.

Покажите, что при $t=0$ среднее выражение представляет собой второй момент распределения вероятностей того, что в очереди находится данное число требований. Применяя аппроксимацию Кларка ко второй сумме среднего выражения для больших t, найдите приближенное значение $\psi(t)$. Покажите, что при $t\to\infty$ оно равно L^2, где L — средняя длина очереди.

Косинус-преобразование разности $\psi-L^2$, зависящая от времени часть которого есть четная функция от t, имеет вид

$$w(f)=4\int_0^{\infty}[\psi(t)-L^2]\cos(2\pi ft)\,dt=4\lambda\mu\,\frac{\mu-\lambda}{\pi}\int_0^{2\pi}\frac{\sin^2\theta\,d\theta}{w^2(w^2+4\pi^2f^2)}.$$

Покажите, что

$$w(f)\to\begin{cases}\dfrac{4\lambda\mu(\mu+\lambda)}{(\mu-\lambda)^4} & \text{при } f\to 0,\\[2ex] \dfrac{4\mu}{(2\pi f)^2} & \text{при } f\to 0.\end{cases}$$

Величина $w(f)$ есть средний квадрат частотного спектра флуктуаций длины очереди относительно среднего значения. Обозначая среднее значение через $E[\cdot]$, имеем

$$w(f)=\lim\frac{2}{T}E[\,|S(f)|^2],$$

где

$$S(f)=\int_0^T[n(t)-L]\,e^{-2\pi ift}dt.$$

Интегрируя вначале по f, а затем по θ, покажите, что

$$\int_0^{\infty}w(f)\,df=E[(n-L)^2]=\frac{\lambda\mu}{(\mu-\lambda)^2}.$$

Морз [590] показал, что при $\lambda \to \mu$ низкочастотная составляющая разности $n(t) - L$ быстро возрастает, в то время как высокочастотные составляющие почти не изменяются. Таким образом, как только приближается насыщение, $n(t)$ после отклонения медленно возвращается к своему среднему значению. Покажите справедливость приближенных выражений

$$\psi(t) \approx \frac{\lambda^2}{(\mu - \lambda)^2} + \frac{\lambda\mu}{(\mu - \lambda)^2} \exp\left[\frac{-(\mu - \lambda)^2 t}{\lambda}\right]$$

и

$$w(f) \approx \frac{4\lambda^2\mu}{(\mu - \lambda)^4} + \lambda^2 (2\pi f)^2.$$

Определим время релаксации для случайных флуктуаций длины очереди относительно L как математическое ожидание времени, необходимого для того, чтобы отклонение длины очереди относительно L изменилось от среднего квадратического до значения, равного 1/e этого отклонения. Воспользовавшись выражением для среднего значения $\psi(t)$, покажите, что время релаксации приближенно равно $\frac{2\lambda}{(\mu - \lambda)^2}$. Это выражение более чувствительно к насыщению (т. е. к такому состоянию, когда $\lambda \to \mu$), чем выражение для L.

21. Произведем следующее упрощение равенства (4.140), которое соответствует допущениям, сделанным для произведения величин N^p и $e^{-\alpha tp}$:

$$\log \frac{(1 - e^{-\alpha t})^{N-n-1}}{(1 - \beta e^{-\alpha t})^{N+n}} = (N - n - 1) \log (1 - e^{-\alpha t}) -$$

$$- (N + n) \log (1 - \beta e^{-\alpha t}) = (N - n - 1) \left(-e^{-\alpha t} - \frac{e^{-2\alpha t}}{2} - \ldots\right) -$$

$$- (N + n) \left(-\beta e^{-\alpha t} - \frac{\beta^2 e^{-2\alpha t}}{2} - \ldots\right) = -N (1 - \beta) e^{-\alpha t}.$$

Всеми остальными произведениями можно пренебречь.

Пусть

$$m e^{-y} = N (1 - \beta) e^{-\alpha t},$$

$$t = \frac{1}{\alpha} [y + \log N + \log (1 - \beta) - \log m].$$

Тогда

$$e^{-m\alpha t} = \frac{m^m e^{-my}}{N^m (1 - \beta)^m}.$$

Пусть $m = k + 1$; подставив это значение в (4.140), найдите

$$\sum_{k=0}^{n-1} \binom{n}{k+1} (1 - \beta)^{k+2} \beta^{n-(k+1)} \frac{(k+1)^{k+1}}{k!} e^{-(k+1)y - (k+1)e^{-y}}.$$

Последние два множителя представляют собой известную функцию распределения экстремального значения

$$T_m(y) = \frac{m^m}{(m-1)!} e^{-my - me^{-y}}$$

(при $m=1$ это распределение Фишера—Типпетта), моменты которого равны

$$\int_{-\infty}^{\infty} T_m(y)\,dy = 1,$$

$$\int_{-\infty}^{\infty} yT_m(y)\,dy = \log m + \gamma - \sum_{j=1}^{m-1} \frac{1}{j},$$

$$\int_{-\infty}^{\infty} y^2 T_m(y)\,dy = \sum_{j=m}^{\infty} \frac{1}{j^2} + \left(\log m + \gamma - \sum_{j=1}^{m-1} \frac{1}{j}\right)^2,$$

где γ — постоянная Эйлера, равная 0,5772156649....

В выражениях для первого и второго моментов при $m=1$ отсутствуют последние члены, содержащие знак суммирования. Проверьте справедливость полученных выражений.

22. Согласно Эпстейну распределения экстремального значения появляются при выборочном методе следующим способом. Если выборка, состоящая из n независимых наблюдений $(x_1, \ldots, x_n)$, взята из совокупности с функцией распределения $F(x)$ и плотностью вероятности $f(x)$ и если

$$y_n = \min(x_1, \ldots, x_n),$$
$$z_n = \max(x_1, \ldots, x_n),$$

то

$$P(y_n > y) = P(x_1 > y,\, x_2 > y, \ldots, x_n > y) = [1 - F(y)]^n$$

и

$$G_n(y) = P(y_n \leqslant y) = 1 - [1 - F(y)]^n,$$
$$g_n(y) = G_n'(y) = nf(y)\,[1 - F(y)]^{n-1}.$$

Кроме того,

$$H_n(z) = P(z_n \leqslant z) = P(x_1 \leqslant z, \ldots, x_n \leqslant z) = [F(z)]^n,$$
$$h_n(z) = H'(z) = nf(z)\,[F(z)]^{n-1}.$$

Таким образом, если $F(x) = 1 - e^{-\lambda x}$, $x \geqslant 0$, то

$$g_n(y) = n\lambda e^{-n\lambda y} \text{ при } y \geqslant 0,$$
$$g_n(y) = 0 \quad \text{при } y < 0,$$
$$h_n(y) = n\lambda e^{-\lambda z}(1 - e^{-\lambda z})^{n-1} \text{ при } z \geqslant 0,$$
$$h_n(y) = 0 \text{ при } z < 0.$$

Для больших n можно ввести случайную величину $\eta_n = nF(y_n)$; при $0 \leqslant u \leqslant n$ получим

$$\Gamma_n(u) \equiv P(\eta_n \leqslant u) = P[nF(y_n) \leqslant u] = P\left[y_n \leqslant F^{-1}\left(\frac{u}{n}\right)\right] =$$

$$= 1 - \left\{1 - F\left[F^{-1}\left(\frac{u}{n}\right)\right]\right\}^n = 1 - \left(1 - \frac{u}{n}\right)^n, \quad u \geqslant 0.$$

При $n \to \infty$ имеем $\Gamma_n(u) \to 1 - e^u \equiv \Gamma(u)$. Таким образом, плотность вероятности также сходится, т. е.

$$\gamma(u) \equiv \lim_{n\to\infty} \gamma_n(u) = e^{-u}, \quad u \geqslant 0.$$

Кроме того,

$$y_n \to y \equiv F^{-1}\left(\frac{\eta}{n}\right).$$

Покажите, что если

$$F(x) = \begin{cases} 0, & x < 0, \\ x^\alpha, & 0 \leqslant x \leqslant 1, \\ 1, & x > 1, \end{cases}$$

то для больших n распределение y_n равно $\Gamma_n(y) = 1 - e^{-ny^\alpha}$ при $y \geqslant 0$ и равно нулю при $y < 0$.

Для z_n при больших n употребляем выражение

$$\xi_n = n\,[1 - F(z_n)].$$

При $0 \leqslant v \leqslant n$

$$\Lambda_n(v) \equiv \mathrm{P}(\xi_n \leqslant v) = \mathrm{P}\left[F(z_n) \geqslant 1 - \frac{v}{n}\right] = \mathrm{P}\left[z_n \geqslant F^{-1}\left(1 - \frac{v}{n}\right)\right] =$$

$$= 1 - H_n\left[F^{-1}\left(1 - \frac{v}{n}\right)\right] = 1 - \left\{F\left[F^{-1}\left(1 - \frac{v}{n}\right)\right]\right\}^n = 1 - \left(1 - \frac{v}{n}\right)^n,$$

$$v \geqslant 0.$$

При $n \to \infty$ имеем $\Lambda_n(v) \to 1 - e^{-v}$.

Таким образом, случайная величина ξ_n сходится по вероятности к случайной величине ξ, а плотность вероятности последней составляет

$$\lambda(v) = \lim_{n \to \infty} \lambda_n(v) = e^{-v}, \quad v \geqslant 0.$$

Следовательно, $z_n \to z$, где

$$z = F^{-1}\left(1 - \frac{\xi}{n}\right).$$

Если $F(x) = 1 - e^{-\lambda x}$ при $x \geqslant 0$ и $F(x) = 0$ при $x < 0$, то $\xi_n = ne^{-\lambda z_n}$ и

$$z_n \approx \frac{\log n}{\lambda} - \frac{\log \xi_n}{\lambda}.$$

Тогда

$$\begin{aligned} &\Lambda_n(z) = e^{-ne^{-\lambda z}} && \text{при } z \geqslant 0, \\ &\Lambda_n(z) = 0 && \text{при } z < 0, \\ &\lambda_n(z) = n\lambda e^{-\lambda z - ne^{-\lambda z}} && \text{при } z \geqslant 0, \\ &\lambda_n(z) = 0 && \text{при } z < 0, \end{aligned}$$

что является распределением I типа для максимального значения.

Наиболее вероятное значение $h_n(z)$ получим, решив уравнение $h_n'(z) = 0$.

Наряду с другими возможны следующие применения теории экстремальных значений:

1) она дает масштаб для линейного вычерчивания функции распределения наблюденных значений, т. е. $\log \log \left[\frac{1}{\Lambda(s)}\right]$ в зависимости от z или $\log z$, и 2) указывает, например, что более крупные образцы более чувствительны к действующим силам. Покажите, что если $f(x) = \frac{1}{x^2}$ при $x \leqslant -1$ и $f(x) = 0$

при $x > -1$, то $\gamma_n(y) = \left(\frac{n}{y^2}\right) e^{\frac{n}{y}}$ при $y < 0$ и $\gamma_n(y) = 0$ при $y \geqslant 0$, что является асимптотическим распределением II типа для минимального значения. Если

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{x}{A}, & 0 \leqslant x \leqslant A, \\ 1, & x > A, \end{cases}$$

то покажите, что

$$h_n(z) = \frac{nz^{n-1}}{A^n}, \quad 0 \leqslant z < A,$$

и $h_n(z) = 0$ при $z < 0$ и $z > A$.

23. Используя теорему Ройтера, покажите существование и единственность решений для одноканальной системы и системы с бесконечным числом каналов.

24. Решите уравнения процесса рождения и гибели, заменив λ_n на $\lambda(t)$ и μ_n на $n\mu(t)$, а также в случае $n\lambda(t)$ и $n\mu(t)$. Введите пуассоновское распределение начального числа требований и вычислите соответствующие вероятности.

25. Пусть $\xi = \zeta + wg(\xi)$. Дифференцируя по двум независимым переменным, вначале по w, а затем по ζ, покажите, что

$$\frac{\partial \xi}{\partial w} = g(\xi) \frac{\partial \xi}{\partial \zeta}.$$

Пусть $u \equiv F(\xi)$, положим, что $\xi = F(u)$, и, следовательно, $g(\xi) = g[f(u)] \equiv \equiv \Phi(u)$. Тогда

$$\frac{\partial u}{\partial w} = \frac{du}{d\xi}\frac{\partial \xi}{\partial u}, \quad \frac{\partial u}{\partial \zeta} = \frac{du}{d\xi}\frac{\partial \xi}{\partial \zeta}.$$

Покажите, что

$$\frac{\partial u}{\partial w} = g(\xi) \frac{\partial u}{\partial \zeta}. \tag{1}$$

Далее имеем

$$\frac{\partial}{\partial w}\left[\Phi(u)\frac{\partial u}{\partial \zeta}\right] = \frac{\partial}{\partial \zeta}\left[\Phi(u)\frac{\partial u}{\partial w}\right],$$

так как каждая часть равна

$$\Phi'(u)\frac{\partial u}{\partial w}\frac{\partial u}{\partial \zeta} + \Phi(u)\frac{\partial^2 u}{\partial w \partial \zeta}.$$

Отсюда, воспользовавшись формулой (1), получим

$$\frac{\partial^2 u}{\partial w^2} = \frac{\partial}{\partial w}\left[g(\xi)\frac{\partial u}{\partial \zeta}\right] = \frac{\partial}{\partial \zeta}\left[g(\xi)\frac{\partial u}{\partial w}\right] = \frac{\partial}{\partial \zeta}\left\{[g(\xi)]^2\frac{\partial u}{\partial \zeta}\right\}.$$

С помощью метода математической индукции находим

$$\frac{\partial^n(u)}{\partial w^n} = \frac{\partial^{n-1}}{\partial \zeta^{n-1}}\left\{[g(\xi)]^n\frac{\partial u}{\partial \zeta}\right\}.$$

Дайте вывод формулы Лагранжа, приведенной в тексте, разложив u в ряд по степеням w, что законно при малых значениях w, т. е.

$$u = F(\xi) = \sum_{n=0}^{\infty} \frac{w^n}{n!}\frac{\partial^n F}{\partial w^n}\bigg|_{w=0}.$$

Глава 5

СЛУЧАЙ ОГРАНИЧЕННОГО БЛУЖДАНИЯ

5.1. Введение

К решению уравнений процесса рождения и гибели могут применяться матричные методы. Так как эти дифференциальные уравнения и уравнения в конечных разностях линейны, то можно использовать матричный метод и по крайней мере записать решение в общем виде. Чтобы получить решение в явном виде, нужно найти характеристические корни, или собственные значения матрицы коэффициентов.

Материал, изложенный в этой главе, имеет двоякое назначение. Во-первых, читатель ознакомится с применением матричных методов в теории массового обслуживания, а во-вторых, приводятся примеры, поясняющие, как применяются эти методы. Наряду с методами, приведенными в предыдущей главе, существует много других алгебраических и аналитических методов, которые применяются для решения задач массового обслуживания, это требует определенного мастерства.

В конце главы кратко рассматривается решение бесконечной системы уравнений процесса рождения и гибели.

Глава начинается с решения уравнений процесса рождения и гибели для системы с поглощающими барьерами при таких положениях верхнего барьера, для которых можно получить решение в явном виде. Эта задача в значительной степени связана с проблемой разрешимости уравнений. В процессе решения уравнений кратко излагаются необходимые теоретические положения [268].

5.2. Вывод уравнений

Мы приведем уравнения процесса рождения и гибели с отражающими и поглощающими барьерами. Напомним, что переход из отражающего состояния возможен только в одном направлении: или в высшее или в низшее состояние. Таким образом, оно ведет себя как отражающий экран. Если система достигла поглощающего состояния, то выйти из него она не может. Это ловушка, или поглощающий экран. В каждом рассматриваемом случае существует два барьера, один в точке E_0, а другой в точке E_N, причем оба барьера в первом случае являются отражающими, а во втором случае — поглощающими.

При отражающих барьерах уравнения принимают вид

$$P_0'(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t),$$

$$P_n'(t) = -(\lambda_n + \mu_n) P_n(t) + \lambda_{n-1} P_{n-1}(t) + \mu_{n+1} P_{n+1}(t),$$

$$n = 1, \ldots, N-1, \tag{5.1}$$

$$P_N'(t) = -\mu_N P_N(t) + \lambda_{N-1} P_{N-1}(t).$$

Уравнения при поглощающих барьерах получим, подставив в (5.1) $\lambda_0 = 0$ и $\mu_N = 0$:

$$P_0'(t) = \mu_1 P_1(t),$$

$$P_n'(t) = -(\lambda_n + \mu_n) P_n(t) + \lambda_{n-1} P_{n-1}(t) + \mu_{n+1} P_{n+1}(t), \tag{5.2}$$

$$n = 1, \ldots, N-1,$$

$$P_N'(t) = \lambda_{N-1} P_{N-1}(t).$$

Если в момент времени $t = 0$ система находится в состоянии E_i, то как для (5.1), так и для (5.2) имеем

$$P_{in}(0) = \delta_{in} \equiv \begin{cases} 0, & n = i, \\ 1, & n \neq i, \end{cases}$$

где δ_{in} — символ Кронекера.

Для удобства в приведенных уравнениях в $P_{in}(t)$ опущен индекс i, показывающий зависимость от начальных условий.

Упражнение. Выпишите уравнения для процесса рождения и гибели с отражающим барьером в точке E_0 и поглощающим барьером в точке E_N, и наоборот.

5.3. Способ решения

Как будет показано в следующих параграфах, полученные системы уравнений можно представить в матричной записи в виде $\overline{P}'(t) = B\overline{P}$. Здесь $\overline{P}(t)$ — транспонированная матрица к $P(t)$, а B — транспонированная матрица к матрице A параметров процесса рождения и гибели. Проведем вычисления, используя матрицу B. Можно также провести анализ, воспользовавшись соотношением $P'(t) = P(t)A$. Заметим, что решение уравнения

$$\frac{d\overline{P}(t)}{dt} = B\overline{P}(t)$$

можно записать в виде

$$\overline{P}(t)=e^{Bt}C,$$

где матрица C определяется из начальных условий, т. е. при $t=0$.

Отсюда следует, что C — единичная матрица, следовательно, $\overline{P}(t)=e^{Bt}$. Эти положения относятся ко всем случаям, рассматриваемым далее. Решение для конечной квадратной матрицы B n-го порядка можно записать в виде

$$P(t)=e^{Bt}=\sum_{j=0}^{\infty}\frac{1}{j!}(Bt)^j.$$

В случае бесконечной матрицы ряд не всегда сходится, особенно когда элементы матрицы неограниченно возрастают.

При наличии с обеих сторон поглощающих или отражающих барьеров B — конечная квадратная матрица с постоянными элементами. Заметим, что для поглощающих и отражающих барьеров соответствующие коэффициенты матрицы B, т. е. λ_0, λ_N и μ_N, равны нулю.

Рассматриваемая здесь матрица B — конечная подматрица коэффициентов N-го порядка. Для ее получения берутся первые $N+1$ уравнений процесса рождения и гибели (4.1) и отбрасывается один член последнего уравнения. Эта подматрица имеет вид

$$B=\begin{bmatrix} -\lambda_0 & \mu_1 & 0 & 0 & \dots & 0 \\ \lambda_0 & -(\lambda_1+\mu_1) & \mu_2 & 0 & \dots & 0 \\ 0 & \lambda_1 & -(\lambda_2+\mu_2) & \mu_3 & \dots & 0 \\ 0 & 0 & \lambda_2 & -(\lambda_3+\mu_3) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \mu_N \\ 0 & 0 & 0 & 0 & \dots & -(\lambda_N+\mu_N) \end{bmatrix}.$$

При $\lambda_N=0$ матрица B становится матрицей коэффициентов системы уравнений (5.1). При $\lambda_0=0$, $\lambda_N=0$ и $\mu_N=0$ матрица B становится матрицей коэффициентов системы уравнений (5.2). В свою очередь,

$$\overline{P}(t)=\begin{bmatrix} P_{00}(t) & P_{10}(t) & \dots & P_{N0}(t) \\ P_{01}(t) & P_{11}(t) & \dots & P_{N1}(t) \\ \dots & \dots & \dots & \dots \\ P_{0N}(t) & P_{1N}(t) & \dots & P_{NN}(t) \end{bmatrix}.$$

Для определения $\overline{P'}(t)$ берется производная от каждого элемента матрицы $\overline{P}(t)$.

На основании теории ограниченных операторов применительно к целой функции $f(B)$ от ограниченного оператора B или с помощью известной теоремы Сильвестра [268] получим следующее спектральное представление:

$$f(B)=\sum_{i=1}^{k}\sum_{m=0}^{m_i-1}\frac{(B-\alpha_i I)^m}{m!}f^{(m)}(\alpha_i)\,Z(\alpha_i). \tag{5.3}$$

Здесь k — число различных характеристических корней матрицы B;

m_i — кратность i-го корня α_i характеристического уравнения $|\alpha I-B|=0$;

$f^{(m)}$ — общий формальный вид производной m-го порядка от f, вычисленной при α_i;

$Z(\alpha_i)$ — полные ортогональные идемпотентные матрицы для матрицы B.

Это означает, что имеют место соотношения:

$$\sum_{i=1}^{k}Z(\alpha_i)=I,\quad Z(\alpha_i)\,Z(\alpha_j)=0\quad (i\neq j),\quad Z^2(\alpha_i)=Z(\alpha_i),$$

где I — единичная матрица, а 0 — нулевая матрица.

Более детальное рассмотрение этого вопроса предлагается выполнить читателю.

Если все характеристические корни различны, то для матрицы B n-го порядка имеем

$$f(B)=\sum_{i=1}^{n}f(\alpha_i)\,Z(\alpha_i), \tag{5.4}$$

где

$$Z(\alpha_i)=\frac{\prod\limits_{j\neq i}(\alpha_j I-B)}{\prod\limits_{j\neq i}(\alpha_j-\alpha_i)}. \tag{5.5}$$

Если $f(B)=e^{Bt}$ и характеристические корни матрицы B различны, то существует спектральное разложение функции $f(B)$:

$$e^{Bt}=\sum_{i=1}^{n}e^{\alpha_i t}Z(\alpha_i). \tag{5.6}$$

Случай кратных характеристических корней выводится из формы теоремы Сильвестра. Если для краткости запишем

$$f(B)=\sum_{i=1}^{k}T(\alpha_i), \tag{5.7}$$

где k — число различных корней, то

$$T(\alpha_i)=f(\alpha_i)\,Z_{m_i-1}(\alpha_i)+f'(\alpha_i)\,Z_{m_i-2}(\alpha_i)+$$
$$+\frac{f''(\alpha_i)}{2!}Z_{m_i-3}(\alpha_i)+\ldots \tag{5.8}$$

Здесь m_i показывает кратность корня α_i, а

$$Z_{m_i}(\alpha_i) = \frac{1}{m_i!} \frac{d^{m_i}}{d\alpha^{m_i}} \frac{F(\alpha)}{\Delta_i(\alpha)}\bigg|_{\alpha=\alpha_i}, \tag{5.9}$$

где

$$F^{(m)}(\alpha_i) = m!(-1)^{n-m-1}(\alpha_i I - B)^{m_i-m-1} \prod_{j \neq i}(\alpha_j I - B) \tag{5.10}$$

— производная m-го порядка от F,

$$\Delta_i(\alpha) = \prod_{j \neq i}(\alpha - \alpha_j). \tag{5.11}$$

Теперь применим эти положения к частному случаю процесса рождения и гибели с поглощающими барьерами в точках E_0 и E_2, так как в этом случае можно вычислить характеристические корни матрицы в явном виде. Затем выполним это для случая поглощающих барьеров в точках E_0 и E_3, E_0 и E_4, E_0 и E_5.

Рассмотрим вначале случаи, когда $\lambda_n = n\lambda$ и $\mu_n = n\mu$. Пусть $N = 2$. Начнем с решения для такого поглощающего состояния, для которого при сделанных ранее допущениях имеем

$$B = \begin{bmatrix} 0 & \mu & 0 \\ 0 & -(\lambda+\mu) & 0 \\ 0 & \lambda & 0 \end{bmatrix}$$

и

$$\alpha I - B = \begin{bmatrix} \alpha & -\mu & 0 \\ 0 & \alpha+\lambda+\mu & 0 \\ 0 & -\lambda & \alpha \end{bmatrix}.$$

Прежде чем переходить к решению этой задачи, необходимо кратко пояснить ее физический смысл. Для одноканальной системы эта задача соответствует пуассоновскому потоку, интенсивность поступления требований в котором зависит от числа уже имеющихся требований, и экспоненциальному времени обслуживания с интенсивностью обслуживания, также зависящей от числа ожидающих требований. Однако как только число требований достигает заданной величины N, процесс полностью прекращается. Это происходит также и в том случае, когда очередь отсутствует. Если бактерии, находящиеся в организме, подчиняются приведенным выше законам размножения (поступление требований) и гибели (обслуживание), то N будет соответствовать числу бактерий, вызывающему заболевание организма, а нуль — случаю, когда бактерии полностью уничтожены и организм здоров. Скорость размножения и гибели бактерий пропорциональна имеющемуся числу их. Заметим, что эта задача для больших N уже рассматривалась в конце предыдущей главы. Здесь же рассматриваются случаи малых значений N.

Приравняв определитель нулю, получим характеристическое уравнение $|\alpha I - B| = 0$. Оно имеет вид

$$\alpha^2 [\alpha + (\lambda + \mu)] = 0.$$

Характеристические корни матрицы B равны $\alpha_1 = \alpha_2 = 0$ и $\alpha_3 = -(\lambda + \mu)$.

Из (5.7) имеем

$$e^{Bt} = f(B) = \sum_{i=1}^{k} T(\alpha_i) = T(0) + T[-(\lambda + \mu)]. \qquad (5.12)$$

Кратность первого корня равна $m_i = 2$. Кроме того, $f(0) = e^{0t} = 1$,

$$f'(0) = te^{0t} = t;$$

таким образом, из (5.8) получим

$$T(0) = f(0) Z_1(0) + f'(0) Z_0(0) + \frac{f''(0)}{2!} Z_{-1}(0) + \ldots =$$

$$= Z_1(0) + tZ_0(0). \qquad (5.13)$$

Производные от f берутся по B как переменной. Из (5.9) имеем

$$Z_1(0) = \frac{d^{(1)}}{d\alpha^{(1)}} \frac{F(\alpha)}{\Delta_1(\alpha)} \bigg|_{\alpha=0} = \frac{\Delta_1(0) F'(0) - F(0) \Delta_1'(0)}{[\Delta_1(0)]^2} \qquad (5.14)$$

и

$$Z_0(0) = \frac{d^{(0)}}{d\alpha^{(0)}} \frac{F(\alpha)}{\Delta_1(\alpha)} \bigg|_{\alpha=0} = \frac{F^{(0)}(0)}{\Delta_1(0)}. \qquad (5.15)$$

Из (5.10) находим

$$F^{(1)}(0) = 1!(-1]^1 (0I - B)^0 \prod_{i \neq j} (\alpha_j I - B) \qquad (5.16)$$

и

$$F^{(0)}(0) = \begin{cases} (-1)[-(\lambda + \mu) I - B] \\ (-1)^2 (0I - B)[-(\lambda + \mu) I - B] \\ (-B)[-(\lambda + \mu) I - B]. \end{cases} \qquad (5.17)$$

Наконец, из (5.11) имеем

$$\Delta_1(0) = [0 - (-\lambda - \mu)] = \lambda + \mu,$$
$$\Delta_1'(0) = 1.$$

Теперь можно вычислить

$$T(0)=\frac{-(\lambda+\mu)\,[-(\lambda+\mu)I-B]-(-B)\,[-(\lambda+\mu)I-B]}{(\lambda+\mu)^2}+$$
$$+\frac{t(-B)\,[-(\lambda+\mu)I-B]}{\lambda+\mu}. \tag{5.18}$$

Но

$$-(\lambda+\mu)I-B=\begin{bmatrix}-(\lambda+\mu) & -\mu & 0\\ 0 & 0 & 0\\ 0 & -\lambda & -(\lambda+\mu)\end{bmatrix}, \tag{5.19}$$

$$0I-B=-B=\begin{bmatrix}0 & -\mu & 0\\ 0 & \lambda+\mu & 0\\ 0 & -\lambda & 0\end{bmatrix} \tag{5.20}$$

и

$$(-B)\,[-(\lambda+\mu)I-B]=0. \tag{5.21}$$

Поэтому

$$T(0)=\frac{1}{\lambda+\mu}\begin{bmatrix}\lambda+\mu & \mu & 0\\ 0 & 0 & 0\\ 0 & \lambda & \lambda+\mu\end{bmatrix}. \tag{5.22}$$

Таким же образом находим $T[-(\lambda+\mu)]$. Так как в этом случае кратность корня равна единице и $f(-\lambda-\mu)=e^{-(\lambda+\mu)t}$, то имеем

$$T[-(\lambda+\mu)]=e^{-(\lambda+\mu)t}Z_0[-(\lambda+\mu)], \tag{5.23}$$

где

$$Z_0[-(\lambda+\mu)]=\frac{d^{(0)}}{d\alpha^{(0)}}\frac{F[-(\lambda+\mu)]}{\Delta_2[-(\lambda+\mu)]}=\frac{F^{(0)}[-(\lambda+\mu)]}{\Delta_2[-(\lambda+\mu)]}, \tag{5.24}$$

$$F^{(0)}[-(\lambda+\mu)]=(-B)^2 \tag{5.25}$$

и

$$\Delta_2[-(\lambda+\mu)]=[-(\lambda+\mu)-0]^2=(\lambda+\mu)^2. \tag{5.26}$$

Следовательно,

$$T[-(\lambda+\mu)]=\frac{e^{-(\lambda+\mu)t}}{(\lambda+\mu)^2}\begin{bmatrix}0 & -\mu(\lambda+\mu) & 0\\ 0 & (\lambda+\mu)^2 & 0\\ 0 & -\lambda(\lambda+\mu) & 0\end{bmatrix}=$$
$$=\frac{e^{-(\lambda+\mu)t}}{(\lambda+\mu)}\begin{bmatrix}0 & -\mu & 0\\ 0 & \lambda+\mu & 0\\ 0 & -\lambda & 0\end{bmatrix}. \tag{5.27}$$

Окончательно находим

$$e^{Bt}=\begin{bmatrix} P_{00}(t) & P_{10}(t) & P_{20}(t) \\ P_{01}(t) & P_{11}(t) & P_{21}(t) \\ P_{02}(t) & P_{12}(t) & P_{22}(t) \end{bmatrix}=$$

$$=\frac{1}{\lambda+\mu}\begin{bmatrix} \lambda+\mu & -\mu e^{-(\lambda+\mu)t}+\mu & 0 \\ 0 & e^{-(\lambda+\mu)t}(\lambda+\mu) & 0 \\ 0 & -\lambda e^{-(\lambda+\mu)t}+\lambda & \lambda+\mu \end{bmatrix}. \quad (5.28)$$

Имеем $P_{00}(t)=1$, $P_{01}(t)=0$,

$$P_{12}(t)=\frac{\lambda-\lambda e^{-(\lambda+\mu)t}}{\lambda+\mu} \text{ и т. д.}$$

Заметим, что сумма элементов каждого столбца равна единице. Кроме того, положив $t=0$, при начальных условиях находим единичную матрицу.

Для случая поглощающего барьера при $N=3$ (как и прежде, $\lambda_n=n\lambda$, $\mu_n=n\mu$), решив характеристическое уравнение

$$\alpha^4+(3\lambda+3\mu)\alpha^3+2(\lambda^2+\lambda\mu+\mu^2)\alpha^2=0,$$

находим следующие характеристические корни:

$$\alpha_1=\alpha_2=0, \quad \alpha_3=\frac{-(3\lambda+3\mu)+\sqrt{\lambda^2+10\lambda\mu+\mu^2}}{2}$$

и

$$\alpha_4=\frac{-(3\lambda+3\mu)-\sqrt{\lambda^2+10\lambda\mu+\mu^2}}{2}. \quad (5.29)$$

Чтобы найти e^{Bt}, поступим так же, как и при $N=2$. Имеем:

$$e^{Bt}=\sum_{i=1}^{k} T(\alpha_i)=T(0)+T(\alpha_3)+T(\alpha_4), \quad (5.30)$$

$$T(0)=Z_1(0)+tZ_0(0), \quad (5.31)$$

$$Z_1(0)=\frac{\Delta_1(0)F'(0)-F(0)\Delta_1'(0)}{[\Delta_1(0)]^2}, \quad Z_0(0)=\frac{F(0)}{\Delta_1(0)}, \quad (5.32)$$

$$F'(0)=f(\alpha_3)f(\alpha_4), \quad F(0)=-f(0)f(\alpha_3)f(\alpha_4), \quad (5.33)$$

$$\Delta_1(0)=\alpha_3\alpha_4, \quad \Delta_1'(0)=-(\alpha_3+\alpha_4), \quad (5.34)$$

Следовательно,

$$T(0)=\frac{\alpha_3\alpha_4 f(\alpha_3) f(\alpha_4)-(\alpha_3+\alpha_4) f(0) f(\alpha_3) f(\alpha_4)-\alpha_3\alpha_4 t f(0) f(\alpha_3) f(\alpha_4)}{\alpha_3^2\alpha_4^2}. \quad (5.35)$$

В свою очередь,

$$T(\alpha_3)=e^{\alpha_3 t} Z_0(\alpha_3), \quad (5.36)$$

$$Z_0(\alpha_3)=\frac{F(\alpha_3)}{\Delta_2(\alpha_3)}, \quad (5.37)$$

$$F(\alpha_3)=-[f(0)]^2 f(\alpha_4), \quad (5.38)$$

$$\Delta_2(\alpha_3)=\alpha_3^3-\alpha_3\alpha_4. \quad (5.39)$$

Значит,

$$T(\alpha_3)=-e^{\alpha_3 t}\frac{[f(0)]^2 f(\alpha_4)}{\alpha_3^2(\alpha_3-\alpha_4)}. \quad (5.40)$$

Поменяв местами α_3 и α_4 в $T(\alpha_3)$, можно легко найти $T(\alpha_4)$:

$$T(\alpha_4)=-e^{\alpha_4 t}\frac{[f(0)]^2 f(\alpha_3)}{\alpha_4^2(\alpha_4-\alpha_3)}. \quad (5.41)$$

И, наконец, приняв обозначение

$$D=\frac{1}{\alpha_3^2\alpha_4^2(\alpha_3-\alpha_4)}, \quad (5.42)$$

получим:

$$P_{00}(t)=[\alpha_3^2\alpha_4^2(\alpha_3-\alpha_4)]D, \quad (5.43)$$

$$P_{01}(t)=P_{02}(t)=P_{03}(t)=0,$$

$$\begin{aligned}P_{10}(t)=\{&-\alpha_3\alpha_4(\alpha_3-\alpha_4)\mu(\lambda+\mu+\alpha_3+\alpha_4)+\\&+e^{\alpha_3 t}\alpha_4^2\mu[(\lambda+\mu)(\lambda+\mu+\alpha_4)+2\lambda\mu]-\\&-e^{\alpha_4 t}\alpha_3^2\mu[(\lambda+\mu)(\lambda+\mu+\alpha_3)+2\lambda\mu]\}D,\end{aligned}$$

$$\begin{aligned}P_{11}(t)=\{&-e^{\alpha_3 t}\alpha_4^2[2\lambda\mu(4\lambda+4\mu+\alpha_4)+(\lambda+\mu)^2(\lambda+\mu+\alpha_4)]+\\&+e^{\alpha_4 t}\alpha_3^2[2\lambda\mu(4\lambda+4\mu+\alpha_3)+(\lambda+\mu)^2(\lambda+\mu+\alpha_3)]\}D,\end{aligned}$$

$$\begin{aligned}P_{12}(t)=\{&e^{\alpha_3 t}\alpha_4^2\lambda[(\lambda+\mu)(7\lambda+7\mu+3\alpha_4)+2\lambda\mu]-\\&-e^{\alpha_4 t}\alpha_3^2\lambda[(\lambda+\mu)(7\lambda+7\mu+3\alpha_3)+2\lambda\mu]\}D,\end{aligned}$$

$$\begin{aligned}P_{13}(t)=[&\alpha_3\alpha_4(\alpha_3-\alpha_4)2\lambda^2-e^{\alpha_3 t}\alpha_4^2 2\lambda^2(3\lambda+3\mu+\alpha_4)+\\&+e^{\alpha_4 t}\alpha_3^2 2\lambda^2(3\lambda+3\mu+\alpha_3)]D,\end{aligned}$$

$$P_{20}(t) = [\alpha_3\alpha_4(\alpha_3 - \alpha_4)\,2\mu^2 - e^{\alpha_3 t}\alpha_4^2 2\mu^2(3\lambda + 3\mu + \alpha_4) +$$
$$+ e^{\alpha_4 t}\alpha_3^2 2\mu^2(3\lambda + 3\mu + \alpha_3)]\,D,$$

$$P_{21}(t) = \{e^{\alpha_3 t}\alpha_4^2 2\mu\,[(\lambda + \mu)(7\lambda + 7\mu + 3\alpha_4) + 2\lambda\mu] -$$
$$- e^{\alpha_4 t}\alpha_3^2 2\mu\,[(\lambda + \mu)(7\lambda + 7\mu + 3\alpha_3) + 2\lambda\mu]\}\,D,$$

$$P_{22}(t) = \{-e^{\alpha_3 t}\alpha_4^2\,[2\lambda\mu(5\lambda + 5\mu + \alpha_4) +$$
$$+ (2\lambda + 2\mu + \alpha_4)(2\lambda + 2\mu)^2] + e^{\alpha_4 t}\alpha_3^2\,[2\lambda\mu(5\lambda + 5\mu + \alpha_3) +$$
$$+ (2\lambda + 2\mu + \alpha_3)(2\lambda + 2\mu)^2]\}\,D,$$

$$P_{23}(t) = \{-\alpha_3\alpha_4(\alpha_3 - \alpha_4)\,2\lambda(2\lambda + 2\mu + \alpha_3 + \alpha_4) +$$
$$+ e^{\alpha_3 t}\alpha_4^2 2\lambda\,[(2\lambda + 2\mu)(2\lambda + 2\mu + \alpha_4) + 2\lambda\mu] -$$
$$- e^{\alpha_4 t}\alpha_3^2 2\lambda\,[(2\lambda + 2\mu)(2\lambda + 2\mu + \alpha_3) + 2\lambda\mu]\}\,D,$$

$$P_{30}(t) = P_{31}(t) = P_{32}(t) = 0$$

и

$$P_{33}(t) = [\alpha_3^2\alpha_4^2(\alpha_3 - \alpha_4)]\,D.$$

5.4. Общий случай

Решим задачу для конечной подматрицы коэффициентов B N-го порядка при $N=2$. Для этого случая можно легко найти соответствующие решения для отражающего и поглощающего барьеров.

Характеристическое уравнение $|\alpha I - B| = 0$ при $N = 2$ имеет вид

$$\alpha^3 + (\lambda_0 + \lambda_1 + \lambda_2 + \mu_1 + \mu_2)\,\alpha^2 +$$
$$+ (\lambda_0\lambda_1 + \lambda_0\lambda_2 + \lambda_0\mu_2 + \lambda_1\lambda_2 + \lambda_2\mu_1 + \mu_1\mu_2)\,\alpha + \lambda_0\lambda_1\lambda_2 = 0. \quad (5.44)$$

Для решения этого кубического уравнения положим

$$\alpha = \beta - \frac{b}{3}.$$

Тогда исходное уравнение примет вид

$$\beta^3 + p\beta + q = 0,$$

где

$$p = c - \frac{b^2}{3}$$

и

$$q = d - \frac{bc}{3} + \frac{2b^3}{27}.$$

Таким образом,

$$p = \frac{1}{3}(\lambda_0\lambda_1 + \lambda_0\lambda_2 + \lambda_0\mu_2 + \lambda_1\lambda_2 + \lambda_2\mu_1 + \mu_1\mu_2 - 2\lambda_0\mu_1 - \lambda_1^2 - \\ - 2\lambda_1\mu_1 - 2\lambda_1\mu_2 - \lambda_2^2 - 2\lambda_2\mu_2 - \mu_1^2 - \mu_2^2 - \lambda_0^2) \qquad (5.45)$$

и

$$q = \frac{1}{27}[2(\lambda_1 + \mu_1)^3 + 2(\lambda_2 + \mu_2)^3 - \lambda_0^2(3\lambda_2 + 3\mu_2 + 3\lambda_1 - 2\lambda_0 - 6\mu_1) - \\ - 3\lambda_1^2(\lambda_2 + \lambda_0 - 2\mu_2) - 3\lambda_2^2(\lambda_0 + \lambda_1 + \mu_1) - \\ - 3\mu_1^2(\lambda_2 + \mu_2 - 2\lambda_0) - 3\mu_2^2(\lambda_0 - 2\lambda_1 + \mu_1) - \\ - 3\lambda_0(-4\lambda_1\lambda_2 + 2\lambda_2\mu_1 + 2\mu_1\mu_2 + 2\lambda_1\mu_2 + 2\lambda_2\mu_2 - \lambda_1\mu_1) - \\ - 3\lambda_1(2\lambda_2\mu_1 - \mu_1\mu_2 + \lambda_2\mu_2) - 6\lambda_2\mu_1\mu_2]. \qquad (5.46)$$

Пусть

$$n = \sqrt{-\frac{3}{4}p}$$

и

$$\cos 3\theta = -\frac{q}{2}\left(-\frac{27}{p^3}\right)^{\frac{1}{2}},$$

тогда

$$\theta = \frac{1}{3}\arccos\left[-\frac{q}{2}\left(-\frac{27}{p^3}\right)^{\frac{1}{2}}\right].$$

Решая обычным способом приведенное кубическое уравнение, находим три корня:

$$\beta_1 = \frac{1}{n}\cos\theta, \quad \beta_2 = \frac{1}{n}\cos(\theta + 120^\circ), \quad \beta_3 = \frac{1}{n}\cos(\theta + 240^\circ).$$

Искомые корни равны

$$\alpha_1 = \beta_1 - \frac{\lambda_0 + \lambda_1 + \lambda_2 + \mu_1 + \mu_2}{3},$$

$$\alpha_2 = \beta_2 - \frac{\lambda_0 + \lambda_1 + \lambda_2 + \mu_1 + \mu_2}{3},$$

$$\alpha_3 = \beta_3 - \frac{\lambda_0 + \lambda_1 + \lambda_2 + \mu_1 + \mu_2}{3}.$$

При $N=2$ решение имеет вид

$$
\left.
\begin{aligned}
P_{00}(t) &= \frac{e^{\alpha_1 t}\left[(\lambda_0+\alpha_2)(\lambda_0+\alpha_3)+\lambda_0\mu_1\right]}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} + \\
&\quad + \frac{e^{\alpha_2 t}\left[(\lambda_0+\alpha_1)(\lambda_0+\alpha_3)+\lambda_0\mu_1\right]}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} + \\
&\quad + \frac{e^{\alpha_3 t}\left[(\lambda_0+\alpha_1)(\lambda_0+\alpha_2)+\lambda_0\mu_1\right]}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}, \\
P_{01}(t) &= \frac{e^{\alpha_1 t}\lambda_0\left[-(\lambda_0+\alpha_3)-(\lambda_1+\mu_1+\alpha_2)\right]}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} - \\
&\quad - \frac{e^{\alpha_2 t}\lambda_0\left[(\lambda_0+\alpha_3)+(\lambda_1+\mu_1+\alpha_1)\right]}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} - \\
&\quad - \frac{e^{\alpha_3 t}\lambda_0\left[(\lambda_0+\alpha_2)+(\lambda_1+\mu_1+\alpha_1)\right]}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}, \\
P_{02}(t) &= \frac{e^{\alpha_1 t}\lambda_0\lambda_1}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} + \\
&\quad + \frac{e^{\alpha_2 t}\lambda_0\lambda_1}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} + \frac{e^{\alpha_3 t}\lambda_0\lambda_1}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}, \\
P_{10}(t) &= -\frac{e^{\alpha_1 t}\mu_1\left[(\lambda_0+\alpha_2)+(\lambda_1+\mu_1+\alpha_3)\right]}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} - \\
&\quad - \frac{e^{\alpha_2 t}\mu_1\left[(\lambda_0+\alpha_1)+(\lambda_1+\mu_1+\alpha_3)\right]}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} - \\
&\quad - \frac{e^{\alpha_3 t}\mu_1\left[(\lambda_0+\alpha_1)+(\lambda_1+\mu_1+\alpha_2)\right]}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}, \\
P_{11}(t) &= \frac{e^{\alpha_1 t}\left[\lambda_0\mu_1+(\lambda_1+\mu_1+\alpha_2)(\lambda_1+\mu_1+\alpha_3)+\lambda_1\mu_2\right]}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} + \\
&\quad + \frac{e^{\alpha_2 t}\left[\lambda_0\mu_1+(\lambda_1+\mu_1+\alpha_1)(\lambda_1+\mu_1+\alpha_3)+\lambda_1\mu_2\right]}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} + \\
&\quad + \frac{e^{\alpha_3 t}\left[\lambda_0\mu_1+(\lambda_1+\mu_1+\alpha_1)(\lambda_1+\mu_1+\alpha_2)+\lambda_1\mu_2\right]}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}, \\
P_{12}(t) &= -\frac{e^{\alpha_1 t}\lambda_1(\lambda_1+\mu_1+\alpha_3+\lambda_2+\mu_2+\alpha_2)}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} - \\
&\quad - \frac{e^{\alpha_2 t}\lambda_1(\lambda_1+\mu_1+\alpha_1+\lambda_2+\mu_2+\alpha_3)}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} - \\
&\quad - \frac{e^{\alpha_3 t}\lambda_1(\lambda_1+\mu_1+\alpha_1+\lambda_2+\mu_2+\alpha_2)}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}, \\
P_{20}(t) &= \frac{e^{\alpha_1 t}\mu_1\mu_2}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} + \frac{e^{\alpha_2 t}\mu_1\mu_2}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} + \\
&\quad + \frac{e^{\alpha_3 t}\mu_1\mu_2}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)},
\end{aligned}
\right\}
\qquad (5.47)
$$

$$\left.\begin{aligned}
P_{21}(t) = &-\frac{e^{\alpha_1 t}\mu_2(\lambda_1+\mu_1+\alpha_2+\lambda_2+\mu_2+\alpha_3)}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} - \\
&-\frac{e^{\alpha_2 t}\mu_2(\lambda_1+\mu_1+\alpha_1+\lambda_2+\mu_2+\alpha_3)}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} - \\
&-\frac{e^{\alpha_3 t}\mu_2(\lambda_1+\mu_1+\alpha_1+\lambda_2+\mu_2+\alpha_2)}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}, \\
P_{22}(t) = &\frac{e^{\alpha_1 t}[\lambda_1\mu_2+(\lambda_2+\mu_2+\alpha_2)(\lambda_2+\mu_2+\alpha_3)]}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} + \\
&+\frac{e^{\alpha_2 t}[\lambda_1\mu_2+(\lambda_2+\mu_2+\alpha_1)(\lambda_2+\mu_2+\alpha_3)]}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)} + \\
&+\frac{e^{\alpha_3 t}[\lambda_1\mu_2+(\lambda_2+\mu_2+\alpha_1)(\lambda_2+\mu_2+\alpha_2)]}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}.
\end{aligned}\right\} \tag{5.47}$$

Аналогично, решая уравнение четвертой степени, можно получить решение для $N = 3$.

Упражнение. Дайте вывод приведенного решения.

5.5. Решение уравнений процесса рождения и гибели методами спектральной теории

Приведенный ниже анализ является кратким изложением метода, в котором содержится более общий подход к решению уравнений процесса рождения и гибели. Идеи этого метода не отличаются от рассмотренных в предыдущем случае, но здесь предпринимается попытка получить решение в общем виде.

Ледерманн и Ройтер [493] исходили из конечной подматрицы n-го порядка матрицы A, рассмотренной в гл. 3 и 4. Обозначим ее $A^{(n)}$. Решение уравнения $\frac{dX(t)}{dt} = X(t)A^{(n)}$, $X(0) = I^{(n)}$, записано в спектральной форме. В пределе, т. е. при $n \to \infty$, оно стремится к решению задачи для процесса рождения и гибели в общем виде.

Положив в подматрице $A^{(n)}$ параметр $\lambda_n = 0$, получим урезанную конечную матрицу n-го порядка, которую обозначим $B^{(n)}$. Урезанная матрица $B^{(n)}$ представляет собой рассмотренную ранее конечную подматрицу B, в которой N заменено на n и $\lambda_n = 0$. Решение дифференциального уравнения

$$\frac{dY(t)}{dt} = Y(t)B^{(n)}, \quad Y(0) = I^{(n)},$$

также записано в спектральной форме; существует такая последовательность значений n, для которой имеет место сходимость к решению задачи процесса рождения и гибели. Последнее не обязано совпадать с решением, полученным ранее.

Авторы показывают, что в первом случае при $t \geqslant 0$ элементы неотрицательны и, являясь вероятностями, не могут превосходить единицу. Чтобы получить спектральное решение, вначале устанавливается соотношение между определителями $\Phi_n(\xi)$:

$$\Phi_n(\xi) - (\xi + \lambda_n + \mu_n)\Phi_{n-1}(\xi) + \lambda_{n-1}\mu_n\Phi_{n-2}(\xi) = 0. \qquad (5.48)$$

Это определители конечных подматриц, полученных для уравнений процесса рождения и гибели, преобразованных по Лапласу с параметром ξ. Ледерманн и Ройтер показали, что характеристические корни, или собственные значения, подматрицы $A^{(n)}$ при условии, что $\Phi(\xi)=0$, различны и отрицательны или равны нулю. При этом наибольшее собственное значение равно нулю в том и только в том случае, если $\lambda_0=0$. Собственные значения подматрицы $A^{(n)}$ не совпадают с собственными значениями подматрицы $A^{(n-1)}$, т. е. существует наибольшее собственное значение подматрицы $A^{(n)}$, превосходящее наибольшее собственное значение подматрицы $A^{(n-1)}$ или равное ему. Последнее превосходит собственное значение подматрицы $A^{(n)}$, превосходящее собственное значение матрицы $A^{(n-1)}$ и т. д. до наименьшего собственного значения подматрицы $A^{(n)}$.

Собственные значения урезанной конечной матрицы, расположенные в убывающем порядке, меньше соответствующих собственных значений исходной конечной подматрицы или равны им. Частные решения удовлетворяют тем же соотношениям (с соответствующими индексами).

Введем обозначения: $l_{-1}=\lambda_0$, $l_0=1$, $l_\nu=(\lambda_1\lambda_2\ldots\lambda_\nu)^{-1}$, $\nu \geqslant 1$, $m_0=1$, $m_\nu=(\mu_1\mu_2\ldots\mu_\nu)^{-1}$, $\nu \geqslant 1$, $\alpha_i^{(n)}$ $(n=0, 1, \ldots, n)$ для собственных значений; $x_r^{(n)}$, $y_r^{(n)\prime}$ $(r=0, 1, \ldots, n)$ для собственных векторов; т. е.

$$x_i^{(n)}A^{(n)} = x_i^{(n)}\alpha_i^{(n)}, \quad A^{(n)}y_i^{(n)\prime} = \alpha_i^{(n)}y_i^{(n)\prime}, \quad i=0, 1, \ldots, n,$$

$$\theta_r^{(n)} = x_r^{(n)}y_r^{(n)\prime}, \quad r=0, 1, \ldots, n,$$

$$\theta_0^{(n)} = 1, \text{ если } \lambda_0=0.$$

Запишем ортогональные идемпотенты в виде

$$D_r^{(n)} = \frac{y_r^{(n)\prime}x_r^{(n)}}{\theta_r^{(n)}} \equiv d_r^{(n)}(i, j).$$

Затем покажем, что

$$d_r^{(n)}(i, j) = \frac{l_{i-1}m_j\Phi_{j-1}\left[\alpha_r^{(n)}\right]\Phi_{j-1}\left[\alpha_r^{(n)}\right]}{\theta_r^{(n)}}, \quad r=0, 1, \ldots, n.$$

Если записать

$$y_0^{(n)\prime} = [\zeta_0^{(n)}, \zeta_1^{(n)}, \ldots, \zeta_n^{(n)}],$$

то $d_0^{(n)}(i, j) = \zeta_i^{(n)}\delta_{0j}$, если $\lambda_0=0$.

Спектральное разложение элемента подматрицы n-го порядка имеет вид

$$f_{ij}^{(n)}(t)=\sum_{r=0}^{n} e^{t\alpha_r(n)} d_r^{(n)}(i,\ j). \tag{5.49}$$

Аналогично находится разложение $g_{ij}^{(n)}(t)$ элементов урезанной конечной матрицы. Путем предельного перехода из этих результатов можно построить решения $f_{ij}(t)$ и $g_{ij}(t)$ бесконечной системы уравнений.

Для подматрицы $A^{(n)}$ при $\lambda_0>0$ спектральная функция $\rho^{(n)}(x)$ в пределах $-\infty<x<\infty$ определяется как неубывающая ступенчатая функция с величиной скачка, равной $\frac{1}{\theta_r^{(n)}}$, при $x=\alpha_r^{(n)}$ $(r=1,\ldots,n)$ и $\rho^{(n)}(0)=0$.

Таким образом,

$$\rho^{(n)}(x)=\begin{cases} 0, & x\geqslant 0,\\ -\left(\frac{1}{\theta_0^{(n)}}+\ldots+\frac{1}{\theta_r^{(n)}}\right), & \alpha_{r+1}^{(n)}\leqslant x\alpha_r^{(n)},\ 0\leqslant r\leqslant n-1,\\ -\left(\frac{1}{\theta_0^{(n)}}+\ldots+\frac{1}{\theta_n^{(n)}}\right), & x<\alpha_n^{(n)}. \end{cases} \tag{5.50}$$

Заметим, что при $x\geqslant 0$ функция $\rho^n(x)=0$, так как $0>\alpha_0^{(n)}$. Использовав полученные ранее результаты, можно записать

$$f_{ij}^{(n)}(t)=l_{i-1}m_j\int_{-\infty}^{\infty}\Phi_{i-1}(x)\,\Phi_{j-1}(x)\,e^{tx}d\rho^{(n)}(x),\quad 0\leqslant i\leqslant n,\quad 0\leqslant j\leqslant n. \tag{5.51}$$

Так как $f_{ij}^{(n)}(0)=\delta_{ij}$, то при $t=0$ это выражение равно нулю.

Если $a_{ij}^{(n)}$ — элемент подматрицы $A^{(n)}$, то имеем следующее соотношение:

$$a_{ij}^{(n)}=l_{i-1}m_j\int_{-\infty}^{\infty}x\Phi_{i-1}(x)\,\Phi_{j-1}(x)d\rho^{(n)}(x),\ 0\leqslant i\leqslant n,\ 0\leqslant j\leqslant n. \tag{5.52}$$

При $n\to\infty$ функция $\rho^{(n)}(x)\to\rho(x)$, и полученное значение $f_{ij}(t)$ удовлетворяет уравнениям процесса рождения и гибели. Аналогичные рассуждения приводятся для $g_{ij}(t)$, и задача рассматривается также при $\lambda_0=0$. Затем, положив $t\to\infty$, рассматриваются предельные значения (стационарное состояние). Потом эти результаты применяются к тем процессам, для которых

$$\frac{\lambda_{n+1}}{\lambda_n}=1+\frac{a}{n}+O\left(\frac{1}{n^2}\right),$$
$$\frac{\lambda_n}{\mu_n}=c\left[1+\frac{b}{n}+O\left(\frac{1}{n^2}\right)\right],\quad c>0, \tag{5.53}$$

где a, b и c — постоянные.

Такой процесс выражается аналитически через параметры a, b и c. Показано, что для определенных значений a, b и c две аппроксимации, т. е. решения для конечной подматрицы и для урезанной конечной подматрицы, имеют различные значения. Приводятся примеры для случая постоянных коэффициентов $\lambda_n = \lambda$ и $\mu_n = \lambda$ как при $\lambda_0 = 0$, так и при $\lambda_0 > 0$, а также предпринята попытка рассмотреть случай линейно изменяющихся параметров.

Задачи

1. Рассмотрев корни характеристического уравнения, покажите, что если

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

то

$$A^{100} = \begin{bmatrix} \dfrac{3^{100}+1}{2} & \dfrac{3^{100}-1}{2} \\ \dfrac{3^{100}-1}{2} & \dfrac{3^{100}+1}{2} \end{bmatrix}.$$

Вычислите e^A.

2. Пусть

$$\frac{dx}{dt} = x + y, \quad \frac{dy}{dt} = x - y.$$

Запишите эту систему в матричном виде и дайте ее решение с помощью матричных методов.

3. Покажите, что при $0 < p < 1$ и $0 < q < 1$

$$\lim_{n\to\infty} T^n = \frac{1}{1+p+q} \begin{bmatrix} q & 1-p \\ q & 1-p \end{bmatrix},$$

где

$$T = \begin{bmatrix} p & 1-p \\ q & 1-q \end{bmatrix}.$$

Джилтей рассмотрел следующую задачу. Пусть $a_{ij}dt$ — вероятность того, что за промежуток времени dt система перейдет из состояния j ($j = 1, \ldots, m$) в состояние i ($i = 1, \ldots, m$; $i \neq j$).

Пусть далее $P_j(t)$ — вероятность того, что в момент времени t система находится в состоянии j. Тогда

$$P_i(t+dt) = P_i(t)\left(1 - \sum_{k\neq i} a_{ki}dt\right) + \sum_{j\neq i} a_{ij}P_j(t)\,dt.$$

Запишем $\sum_{\neq i} a_{ki} = -a_{ii}$. После упрощения получим

$$P_i'(t) = \sum_{j=1}^{m} a_{ij}P_j(t), \quad i = 1, \ldots, m.$$

Кроме того, имеем

$$\sum_{j=1}^{m} P_j(t) = 1.$$

Если рассматривается матрица A, составленная из элементов a_{ij}, то можно показать, что все отличные от нуля корни характеристического уравнения $|A - zI| = 0$ (где I — единичная матрица) имеют отрицательные действительные части. Так, если

$$|a_{ii}| > \sum_{i \neq j} |a_{ij}|, \quad i = 1, \ldots, m$$

(что справедливо для данного случая), то D, определитель матрицы A, не равен нулю. Если же он равен нулю, то между строками матрицы существует линейная зависимость, т. е. существуют такие числа $b_1, \ldots, b_m$, не все равные нулю, что

$$\sum_{i=1}^{m} b_i a_{ij} = 0, \; j = 1, \ldots, m.$$

Выберем такое число b_k, что $|b_k| = \max |b_i|$, $(i = 1, \ldots, m)$, тогда равенство при $j = k$ примет вид

$$a_{kk} = -\frac{\sum\limits_{i \neq k} |a_{ik}| \, |b_i|}{|b_k|} \leqslant \sum_{i \neq k} |a_{ik}|,$$

что приводит к противоречию.

Если $z^* = x + iy$ — отличный от нуля комплексный характеристический корень, то главная диагональ элементов матрицы $[A - z^*I]$ имеет в качестве модулей

$$[(a_{kk} - x)^2 + y^2]^{\frac{1}{2}}, \quad k = 1, \ldots, m.$$

Определитель $|A - zI|$ (который должен обращаться в нуль при значении характеристического корня, равном z^*) не может равняться нулю, если хотя бы для одного значения k не выполняется следующее соотношение:

$$[(a_{kk} - x)^2 + y^2]^{\frac{1}{2}} \leqslant \sum_{k \neq j} |a_{kj}| < |a_{kk}|.$$

При этом для $|A - z^*I|$ должно выполняться первое неравенство, а для D всегда выполняется второе неравенство. Это приводит к соотношению

$$(a_{kk} - x)^2 + y^2 < a_{kk}^2.$$

Но, по определению, a_{kk} нигде не положительно, следовательно, для того чтобы это неравенство удовлетворялось, значение x должно быть отрицательным.

Предложите пример применения этой теоремы.

Часть III

НЕМАРКОВСКИЕ МОДЕЛИ МАССОВОГО ОБСЛУЖИВАНИЯ

В этой части книги рассматриваются случаи, когда распределение длительности промежутков времени между моментами поступления требований в систему или распределение времени обслуживания, или же то и другое не является экспоненциальным. В гл. 6 исследуется многоканальная система с пуассоновским входящим потоком и постоянным временем обслуживания. Затем рассматривается одноканальная система с пуассоновским входящим потоком и эрланговским временем обслуживания и система с эрланговским входящим потоком и экспоненциальным временем обслуживания. В гл. 7 вводится понятие о методе вложенных цепей Маркова; этот метод применяется к исследованию группового обслуживания. В гл. 8 изучается система с пуассоновским входящим потоком и произвольным временем обслуживания. Исследование этой системы продолжается в гл. 9, в которой главное внимание уделяется времени ожидания. Затем здесь же рассматривается система, имеющая общий входящий поток и произвольное распределение времени обслуживания. В гл. 10 рассматривается одноканальная система с экспоненциальным временем обслуживания, в которую поступает поток с ограниченным последействием.

13 Зак. 2076

Глава 6

ЧАСТНЫЕ СЛУЧАИ НЕМАРКОВСКИХ МОДЕЛЕЙ

6.1. Введение

После рассмотрения различных моделей обслуживания с пуассоновским входящим потоком и экспоненциальным временем обслуживания (будем называть их просто марковскими системами) читатель может заинтересоваться, все ли системы массового обслуживания являются марковскими. Очевидно, что могут существовать и другие случаи. Например, на практике система может иметь не экспоненциально распределенное, а постоянное время обслуживания. В различных частях книги мы будем возвращаться к марковским системам массового обслуживания при изучении смежных вопросов, связанных с рассмотренными случаями.

Здесь исследование будет проводиться главным образом для стационарного состояния, поскольку решение для переходного состояния получить трудно. Будут изучены две системы с пуассоновским входящим потоком: с постоянным и эрланговским временем обслуживания. В последнем случае каждое требование можно рассматривать проходящим через ряд обслуживающих устройств, расположенных одно за другим, каждому из которых соответствует свое распределение времени обслуживания. Будем считать, что в данном случае они имеют одно и то же экспоненциальное распределение. Это может иметь место, например, когда автомобилям, случайным образом прибывающим на пункт обслуживания, требуются различные операции, скажем, заправка бензином, водой, зарядка аккумуляторов и т. д.

Третья система — одноканальная, с эрланговским входящим потоком и экспоненциальным временем обслуживания при обслуживании требований в порядке поступления.

Все дело в том, что́ именно требуется исследовать. Одним из естественных показателей является вероятность того, что в системе находится заданное число требований. С ее помощью можно найти среднее число требований, находящихся в системе. Другим показателем является распределение времени ожидания. Ниже будут получены выражения для обоих этих показателей.

6.2. Многоканальная система с постоянным временем обслуживания

Рассмотрим пуассоновский поток с параметром λ, поступающий в систему с c параллельными каналами, имеющими одинаковое постоянное время обслуживания. Этот случай исследовал Кроммелин [164, 165]. Примем время обслуживания равным одной временно́й единице. Заметим, что для эргодичности системы необходимо, чтобы $c > \lambda$.

Пусть p_n — вероятность того, что при стационарном состоянии в данный момент времени в очереди и на обслуживании находится ровно n требований, и пусть a_n — вероятность того, что в этот же момент времени в системе находится не более n требований, т. е.

$$a_n = \sum_{i=0}^{n} p_i. \tag{6.1}$$

Поскольку время обслуживания принято равным одной временно́й единице, то вероятность того, что к концу временно́го интервала в системе будет находиться определенное число требований, можно выразить через вероятности того или иного числа требований, имеющихся в начале интервала, умноженные на вероятности поступления в этом интервале соответствующего числа требований. Так как это событие может появиться при различных комбинациях, то вероятности складываются. Заметим, что в этом временно́м интервале будет закончено обслуживание всех требований, уже обслуживающихся в момент начала данного интервала и они покинут систему.

В таком случае через интервал, равный одной временно́й единице (постоянное время обслуживания), имеем

$$\begin{aligned}
&p_0 = a_c e^{-\lambda},\\
&p_1 = a_c \lambda e^{-\lambda} + p_{c+1} e^{-\lambda},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&p_n = a_c \frac{\lambda^n}{n!} e^{-\lambda} + p_{c+1} \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} + \ldots + p_{n+c} e^{-\lambda}.
\end{aligned} \tag{6.2}$$

Первое уравнение описывает случай, когда очередь отсутствует, если в начале интервала на обслуживании находилось не более c требований и в течение интервала требования не поступали. Второе уравнение описывает случай, когда в очереди находится одно требование, если в начале интервала на обслуживании находилось не более c требований, а за это время поступило еще одно; или же в начале интервала в системе находилось $c+1$ требований, а затем не поступило ни одного требования, и т. д. Множитель $\left(\frac{\lambda^n}{n!}\right)e^{-\lambda}$ — вероятность поступления n требований в этом интервале.

1. Начнем с вычисления p_n. Введем обозначение:

$$P(z)=\sum_{n=0}^{\infty} p_n z^n.$$

Тогда

$$P(1)=1.$$

Если, кроме того, ввести обозначение

$$P_c(z)=\sum_{n=0}^{c} p_n z^n, \tag{6.3}$$

то, умножая уравнения на z в соответствующей степени и суммируя, получим

$$P(z)=\frac{P_c(z)-a_c z^c}{1-z^c e^{\lambda(1-z)}}. \tag{6.4}$$

Так как $0\leqslant p_n\leqslant 1$, то $P(z)$ — регулярная функция, ограниченная внутри единичного круга $|z|<1$. Таким образом, внутри единичного круга и на его границе числитель имеет нули, которые совпадают со всеми нулями знаменателя внутри единичного круга и на его границе. Согласно теореме Руше существует c таких нулей. Обозначим их через $z_1, \ldots, z_c$. Очевидно, что $z_c=1$ есть нуль.

Поскольку числитель есть полином степени c, то его можно представить в виде

$$K(z-1)(z-z_1)\ldots(z-z_{c-1}), \tag{6.5}$$

где постоянная K определяется из соотношения

$$\lim_{z\to 1} P(z)=1.$$

В этом случае имеем

$$P(z)=-\frac{c-\lambda}{(1-z_1)\ldots(1-z_{c-1})}\,\frac{(z-1)(z-z_1)\ldots(z-z_{c-1})}{1-z^c e^{\lambda(1-z)}}. \tag{6.6}$$

Теперь можно найти вероятность p_n как коэффициент при z^n, разложив функцию $P(z)$ в степенной ряд. Например, при $c=1$, дифференцируя функцию

$$P(z)=\frac{(1-\lambda)(1-z)}{1-z e^{\lambda(1-z)}} \tag{6.7}$$

n раз по z, разделив на $n!$ и положив $z=0$, получим

$$p_0=(1-\rho),\qquad p_1=(1-\rho)(e^{\rho}-1),$$

$$p_n = (1-\rho)\sum_{k=1}^{n}(-1)^{n-k}e^{k\rho}\left[\frac{(k\rho)^{n-k}}{(n-k)!}+\frac{(k\rho)^{n-k-1}}{(n-k-1)!}\right],\ n \geqslant 2. \quad (6.8)$$

Здесь $\rho = \frac{\lambda}{\mu}$ (заметим, что рассматривается интенсивность обслуживания $\mu = 1$, но выражение справедливо при любом μ).

2. Вычислим теперь $P(t=0)$ — вероятность отсутствия ожидания. Она, очевидно, равна сумме вероятностей того, что занято $i<c$ каналов. Таким образом,

$$P(t=0) = a_{c-1} = \sum_{i=0}^{c-1} p_i. \quad (6.9)$$

Пусть

$$Q(z) = \sum_{k=0}^{\infty} a_k z^k, \quad (6.10)$$

тогда, поскольку $a_k - a_{k-1} = p_k$, имеем

$$Q(z)(1-z) = P(z) \quad \text{или} \quad Q(z) = \frac{P(z)}{1-z}. \quad (6.11)$$

Так как $P(z)$ известно, то легко вычислить a_{c-1} как коэффициент при z^{c-1} в выражении для $Q(z)$. Нетрудно показать, что в числителе z^{c-1} появляется с коэффициентом K, следовательно,

$$P(t=0) = \frac{c-\lambda}{(1-z_1)\ldots(1-z_{c-1})}$$

и

$$\log P(t=0) = \log(c-\lambda) - \sum_{k=1}^{c-1}\log(1-z_k). \quad (6.12)$$

Вычислим теперь это выражение в явном виде.

Использовав результат из книги Уиттекера и Ватсона [877][1], имеем

$$\frac{1}{2\pi i}\int_C f(z)\,\frac{g'(z)}{g(z)}\,dz = \sum r_i f(a_i) - \sum s_j f(b_j), \quad (6.13)$$

где C — контур в плоскости z, а функции $f(z)$ и $g(z)$ — аналитические внутри контура C и на его границе; кроме того, функция $g(z)$ имеет внутри контура C конечное число полюсов b_j порядка s_j соответственно. Функция $g(z)$ также имеет внутри контура C нули a_i с соответствующей кратностью r_i.

[1] Имеется русский перевод: Уиттекер и Ватсон. Курс современного анализа, т. I—II, М., 1963. — *Прим. ред.*

Запишем

$$g(z)=1-\frac{e^{\lambda(z-1)}}{z^c}=-\frac{e^{\lambda(z-1)}}{z^c}\left[1-z^c e^{\lambda(1-z)}\right], \tag{6.14}$$

где $z_1, \ldots, z_{c-1}$ — простые нули, а $z=0$ — полюс порядка c.

Итак,

$$\frac{1}{2\pi i}\int_\Gamma \log(z-1)\frac{g'}{g}\,dz=\sum_{k=1}^{c-1}\log(z_k-1)-c\log(-1)=$$

$$=(c-1)\pi i+\sum_{k=1}^{c-1}\log(1-z_k)-c\pi i=-\pi i+\sum_{k=1}^{c-1}\log(1-z_k)=$$

$$=-\pi i+\log(c-\lambda)-\log P(t=0). \tag{6.15}$$

Таким образом, чтобы получить выражение для $P(t=0)$ в явном виде, вычислим интеграл, записанный слева.

Разделим Γ [контур, который исключает особую точку $z=1$ и внешние нули функции $g(z)$] на две части: 1) вдоль контура C и 2) вдоль остальной части, как показано на рис. 6.1. Вначале вычислим интеграл по контуру C. Путем интегрирования по частям получим

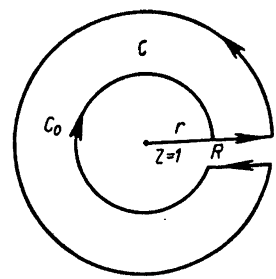

Рис. 6.1.

$$\frac{1}{2\pi i}\int_C \log(z-1)\frac{g'}{g}\,dz=$$

$$=\frac{1}{2\pi i}\left[\log(z-1)\log g(z)\right]_C-$$

$$-\frac{1}{2\pi i}\int_C \frac{\log g(z)}{z-1}\,dz,$$

где первая величина в правой части должна быть вычислена в граничных точках контура C, т. е. при $\theta=0$ и при $\theta=2\pi$, а значение при $\theta=0$ вычитается из значения при $\theta=2\pi$. Чтобы показать это, положим

$$z=1+Re^{i\theta},$$

логарифмируя, получим

$$\log(z-1)=\log R+i\theta$$

и

$$\log g(z)=\log\left[1-\frac{e^{\lambda Re^{i\theta}}}{(1+Re^{i\theta})^c}\right]=$$

$$=\log\left[1-\frac{e^{\lambda R(\cos\theta+i\sin\theta)}}{(1+R\cos\theta+iR\sin\theta)^c}\right]. \tag{6.16}$$

Можно показать, что при изменении θ от нуля до 2π аргумент $\log g(z)$ возвращается к начальному нулевому значению. Таким образом,

$$[\log(z-1)\log g(z)]_0^{2\pi} = (\log R + 2\pi i)\log\left[1 - \frac{e^{\lambda R}}{(1+R)^c}\right] - \\ - \log R \log\left[1 - \frac{e^{\lambda R}}{(1+R)^c}\right] = 2\pi i g(1+R). \quad (6.17)$$

Вдоль прямолинейных отрезков имеем

$$\int \log(z-1)\frac{g'(z)}{g(z)}\,dz = \int_{1+r}^{1+R} \log(z-1)\frac{g'(z)}{g(z)}\,dz + \\ + \int_{1+R}^{1+r} \log(z-1)\frac{g'(z)}{g(z)}\,dz. \quad (6.18)$$

Здесь для первого выражения правой части $\theta=0$, а для второго $\theta=2\pi$. Когда $\log(z-1)$ разлагается на действительную и мнимую части, то после сокращения это равенство будет иметь вид

$$\int_{1+r}^{1+R} -2\pi i\frac{g'(z)}{g(z)}\,dz = -2\pi i\,[\log g]_{1+r}^{1+R} = \\ = -2\pi i\,[\log g(1+R) - \log g(1+r)]. \quad (6.19)$$

На меньшей окружности имеем $z=1+re^{i\theta}$, $dz=ire^{i\theta}d\theta$. При $r\to 0$, или, что то же самое, при $z\to 1$, из разложения в ряд в окрестности точки $z=1$ имеем, что функция $g(z)$ ведет себя как первый член, так как нулевой член при $z=1$ обращается в нуль, т. е.

$$\lim_{z\to 1} g(z) \approx (z-1)\,g'(z) \approx (c-\lambda)(z-1) \quad (6.20)$$

и

$$\lim_{z\to 1}\frac{g'(z)}{g(z)} \approx \frac{1}{z-1} = \frac{e^{i\theta}}{r}. \quad (6.21)$$

Итак,

$$\int_{C_0} \log(z-1)\frac{g'}{g}\,dz = \int_{2\pi}^{0} (\log r + i\theta)\,i\,d\theta\,[1+o(r)] = 2\pi^2 - 2\pi i\log r.$$

Так как $\lim_{r\to 0} r\log r = 0$, то

$$\int_{C_0} \log(z-1)\frac{g'}{g}\,dz = 2\pi^2 - 2\pi i\,[g(1+r) - \log(c-\lambda)], \quad (6.22)$$

потому что

$$g(z) \approx (c-\lambda)(z-1) = (c-\lambda) r e^{i\theta}, \qquad (6.23)$$

а при $z = 1 + r, \quad \theta = 0$

$$\log g(1+r) \approx \log (c-\lambda) r = \log (c-\lambda) + \log r$$

или

$$\log r \approx \log(1+r) - \log(c-\lambda). \qquad (6.24)$$

Объединив полученные ранее результаты, получим

$$\frac{1}{2\pi i}\int_{\Gamma} \log(z-1)\frac{g'}{g}\,dz = -\frac{1}{2\pi i}\int_{C}\frac{\log g}{z-1}\,dz - \pi i + \log(c-\lambda). \qquad (6.25)$$

Поэтому

$$\log P(t=0) = \frac{1}{2\pi i}\int_{C}\frac{\log g}{z-1}\,dz. \qquad (6.26)$$

Но

$$\log g(z) = \log\left[1 - \frac{e^{\lambda(z-1)}}{z^{c}}\right] = -\sum_{n=1}^{\infty}\frac{e^{n\lambda(z-1)}}{nz^{nc}}, \qquad (6.27)$$

что справедливо, так как на контуре C

$$\left|\frac{e^{\lambda(z-1)}}{z^{c}}\right| < 1.$$

Поэтому

$$\log P(t=0) = -\frac{1}{2\pi i}\int_{C}\frac{dz}{z-1}\sum_{n=1}^{\infty}\frac{e^{n\lambda(z-1)}}{nz^{nc}} =$$

$$= -\sum_{n=1}^{\infty}\frac{1}{n}\,\frac{1}{2\pi i}\int_{C}\frac{e^{n\lambda(z-1)}}{z^{nc}}\,\frac{dz}{z-1}. \qquad (6.28)$$

В данном случае при $z=0$ существует полюс порядка nc, а при $z=1$ существует другой полюс первого порядка. Произведем почленное интегрирование. Согласно теории вычетов получим, что если $z=z_0$ — полюс функции $h(z)$ порядка m, то вычет равен произведению выражения $\frac{1}{(m-1)!}$ на $(m-1)$-ю производную функции $(z-z_0)^m\, h(z)$, вычисленной при z_0. Затем все вычеты складываются. Окончательно получим логарифм искомого выражения

$$\log P(t=0) = -\sum_{n=1}^{\infty}\frac{1}{n}\left[1 - \sum_{i=0}^{nc-1}\frac{(n\lambda)^i}{i!}e^{-n\lambda}\right] =$$

$$= -\sum_{n=1}^{\infty}\frac{1}{n}\sum_{i=nc}^{\infty}\frac{(n\lambda)^i}{i!}\,e^{-n\lambda}, \qquad (6.29)$$

так как

$$\lim_{z\to 1}(z-1)\frac{e^{n\lambda(z-1)}}{z^{nc}(z-1)}=1. \tag{6.30}$$

Согласно теореме Лейбница

$$\lim_{z\to 0}\frac{1}{(nc-1)!}\frac{d^{nc-1}}{dz^{nc-1}}z^{nc}\frac{e^{n\lambda(z-1)}}{z^{nc}(z-1)}=-\sum_{j=0}^{nc-1}\frac{(n\lambda)^j}{j!}e^{-n\lambda}. \tag{6.31}$$

3. Полячек получил для больших c следующее приближенное выражение:

$$P(t>0)=1-P(t=0)=\frac{1}{1-\rho}\frac{\rho^c e^{(1-\rho)c}}{\rho\sqrt{2\pi c}}, \tag{6.32}$$

где $\rho=\left(\frac{\lambda}{c}\right)<1$.

Это еще один важный показатель — вероятность ожидания положительной длительности. Кроме того, для больших c Полячек получил выражение

$$W_q=\frac{1}{c\sqrt{2\pi c}}\frac{\rho^c e^{(1-\rho)c}}{(1-\rho)^2}\left[1+O\left(\frac{1}{c}\right)\right].$$

4. Вычислим теперь вероятность того, что время ожидания будет меньше t, т. е. $P(<t)=1-P(>t)$. Пусть $t=T+s$, где T — целое число временны́х единиц, а s — дробная часть. Пусть b_n — вероятность того, что не более n требований из числа находящихся на обслуживании в данный момент времени будут обслуживаться в момент времени s, а b_{Tc+c-1} — вероятность того, что после момента времени s на обслуживании будет находиться не более $Tc+c-1$ требований, и, следовательно, после момента времени $s+T=t$ будет обслуживаться не более $c-1$ требований, так как обслуживание c требований заканчивается за одну единицу времени, и, значит, обслуживание Tc требований закончится за T единиц.

Тогда

$$P(<t)=b_{Tc+c-1}. \tag{6.33}$$

Заметим, что

$$a_n=e^{-\lambda s}\sum_{k=0}^{n}b_k\frac{(\lambda s)^{n-k}}{(n-k)!}. \tag{6.34}$$

Пусть

$$R(z)=\sum_{k=0}^{\infty}b_k z^k \quad \text{при} \quad |z|<1 \tag{6.35}$$

и

$$Q(z)=\sum_{n=0}^{\infty}a_n z^n=e^{-\lambda z}\sum_{k=0}^{\infty}b_k z^k e^{\lambda z s}. \tag{6.36}$$

Тогда

$$R(z)=Q(z)\,e^{\lambda s(1-z)}=$$
$$=a_{c-1}(z-z_1)\dots(z-z_{c-1})\sum_{j=0}^{\infty}z^{jc}e^{j\lambda}e^{-\lambda z(j+s)}. \qquad (6.37)$$

Коэффициент при z^{Tc+c-1} равен

$$b_{Tc+c-1}=\sum_{m=0}^{T}\sum_{k=0}^{c-1}a_k\frac{[-\lambda(m+s)]^{T-mc+c-1-k}}{(T-mc+c-1-k)!}e^{\lambda(m+s)}. \qquad (6.38)$$

Более удобным способом для получения b_{Tc+c-1} является использование функции

$$R(z)=\frac{a_{c-1}(z-z_1)\dots(z-z_{c-1})e^{\lambda s(1-z)}}{1-z^c e^{\lambda(1-z)}}, \qquad (6.39)$$

которая имеет полюс в точке $z=1$ [несмотря на то, что функция $Q(z)$ не имела такого полюса, так как ее числитель имел нуль в точке $z=1$]. Разложим функцию $R(z)$ в ряд Лорана до следующего полюса, который является действительным. Чтобы показать, что следующий полюс является действительным, рассмотрим выражение $z^c e^{\lambda(1-z)}$.

При возрастании z от значения $z=1$ производная данного выражения положительна. Но при $z\to\infty$ это выражение обращается в нуль. Следовательно, существует такое значение r_0, при котором это выражение принимает значение, равное единице, и, таким образом, знаменатель функции $R(z)$ обращается в нуль. То, что r_0 — первый полюс, следует из соотношений

$$z^c e^{\lambda(1-z)}=r^c e^{\lambda(1-r\cos\theta)}e^{i(c\theta-\lambda r\sin\theta)}, \quad z=re^{i\theta},$$
$$|z^c e^{\lambda(1-z)}|>r^c e^{\lambda(1-r)}>1 \quad \text{при} \quad 1\leqslant r\leqslant r_0.$$

Так как $z=1$ есть особая точка, то можно записать

$$R(z)=-\frac{1}{z-1}+H(z),$$

где функция $H(z)$ аналитическая внутри круга $|z|<r_0$.

Пусть $H(z)=\sum_{n=0}^{\infty}h_n z^n$. При $1<|z|<r_0$ функцию $R(z)$ можно разложить в ряд Лорана, записав ее знаменатель в виде

$$-z^c e^{\lambda(1-z)}\left[1-\frac{1}{z^c e^{\lambda(1-z)}}\right].$$

Таким образом,

$$R(z)=-a_{c-1}(z-z_1)\dots$$
$$\dots(z-z_{c-1})\sum_{n=0}^{\infty}z^{-(n+1)c}e^{-(n+1-s)\lambda}e^{z\lambda(n+1-s)}. \qquad (6.40)$$

Затем получим

$$H(z)=R(z)+\frac{1}{z-1}=R(z)+\sum_{n=1}^{\infty}\frac{1}{z^n} \quad \text{при} \quad 1<|z|<r_0.$$

Заметим, что при $|z|>1$

$$\frac{1}{z-1}=\frac{1}{z\left(1-\frac{1}{z}\right)}=\frac{1}{z}\left(1+\frac{1}{z}+\frac{1}{z^2}+\ldots\right)=\sum_{n=1}^{\infty}\frac{1}{z^n}.$$

Это позволяет определить $H(z)$. Но функция $H(z)$ аналитическая также и при $|z|<1$, и функция $(1-z)^{-1}=\sum_{n=0}^{\infty} z^n$ аналитическая при $|z|<1$; следовательно, можно получить разложение функции $R(z)$ при $|z|<1$, откуда находим искомые коэффициенты.

5. Используя выражение $\sum_{n=1}^{\infty} np_n$ и проведя аналогичный анализ, можно показать, что среднее время ожидания равно

$$W_q=\frac{1}{\lambda}\sum_{k=1}^{c-1}\frac{1}{1-z_k}+\frac{\lambda^2-c^2+c}{2\lambda(c-\lambda)}=$$

$$=\sum_{i=1}^{\infty}e^{-i\lambda}\left[\sum_{j=ic}^{\infty}\frac{(i\lambda)^j}{j!}-\frac{c}{\lambda}\sum_{j=ic+1}^{\infty}\frac{(i\lambda)^j}{j!}\right]. \tag{6.41}$$

Если желательно пользоваться постоянной $\rho=\frac{\lambda}{\mu c}$, где $\frac{1}{\mu}$ — постоянное время обслуживания, принятое равным одной временнóй единице, то можно заменить λ на ρc. Читатель может показать справедливость этого, произведя анализ размерности W_q.

6. При $c=1$ имеем (см. гл. 3)

$$P(<t)=(1-\rho)\sum_{i=0}^{k}e^{\rho(\mu t-i)}\frac{[-\rho(\mu t-i)]^i}{i!}, \tag{6.42}$$

где k — наибольшее целое число, не превосходящее μt, и

$$W_q=\int_0^{\infty} t\,dP(<t)=\frac{\rho}{2\mu(1-\rho)}. \tag{6.43}$$

Укажем на то важное обстоятельство, что для системы с экспоненциальным временем обслуживания время ожидания при регулярном входящем потоке составляет половину среднего времени ожидания при пуассоновском входящем потоке при одинаковой интенсивности поступления требований. Таким образом, более выгодно иметь регулярный входящий поток.

Среднее время ожидания для требований, которые ожидают положительное время, равно

$$\frac{W_q}{P(t>0)}=\frac{1}{2\mu(1-\rho)}. \tag{6.44}$$

7. Для общего случая Молина получил приближенное выражение

$$W_q=P(t>0)\frac{1}{\mu c(1-\rho)}\frac{c}{c+1}\frac{1-\rho^{c-1}}{1-\rho^c}, \tag{6.45}$$

где

$$P(t>0)=\frac{\dfrac{\left(\dfrac{\lambda^c e^{-\lambda}}{c!}\right)c}{c-\lambda}}{1-e^{-\lambda}\sum\limits_{n=c}^{\infty}\left(\dfrac{\lambda_n}{n!}\right)+\dfrac{\left(\dfrac{\lambda^c e^{-\lambda}}{c!}\right)c}{c-\lambda}}. \tag{6.46}$$

В последнем выражении было принято $\mu=1$. Чтобы получить выражение, содержащее μ, нужно заменить λ на $\frac{\lambda}{\mu}$.

8. В качестве примера проведем вычисление среднего времени ожидания W_q при $c=1$ с помощью альтернативного метода, предложенного Фраем [271]. Известно, что $W_q=\frac{L_q}{\lambda}$, где L_q — среднее число требований, находящихся в очереди. Но $L_q=L-\rho$, где ρ — загрузка системы (если $\mu=1$, то $\rho=\lambda$) и $L=\sum\limits_{n=0}^{\infty} np_n$. Значение этого выражения можно получить, вычислив $\sum\limits_{n=0}^{\infty} n^2p_n$, что позволяет найти L.

Итак,

$$S=\sum_{n=0}^{\infty} n^2p_n=\sum_{n=0}^{\infty} n^2\left(\sum_{k=0}^{n} p_{n+1-k}\frac{\lambda^k e^{-\lambda}}{k!}+p_0\frac{\lambda^n e^{-\lambda}}{n!}\right). \tag{6.47}$$

Заметим, что

$$\sum_{n=0}^{\infty}\sum_{k=0}^{n}=\sum_{k=0}^{\infty}\sum_{n=k}^{\infty}. \tag{6.48}$$

Значение $D\equiv\sum\limits_{n=0}^{\infty} n^2\left(\frac{\lambda^n e^{-\lambda}}{n!}\right)$ находим следующим образом:

$$\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}=e^{\lambda},$$

$$\frac{d}{d\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}=\sum_{n=1}^{\infty} n\frac{\lambda^{n-1}}{n!}=e^{\lambda},$$

$$\frac{d}{d\lambda}\sum_{n=1}^{\infty} n\frac{\lambda^{n-1}}{n!}=\sum_{n=2}^{\infty} n(n-1)\frac{\lambda^{n-2}}{n!}=e^{\lambda}. \tag{6.49}$$

После упрощения получим, что среднее выражение составляет $\frac{D}{\lambda^2} - \frac{e^\lambda}{\lambda}$. Таким образом, $D = \lambda e^\lambda (1+\lambda)$. Следовательно,

$$p_0 \sum_{n=0}^{\infty} n^2 \frac{\lambda^n e^{-\lambda}}{n!} = p_0 \lambda (1+\lambda). \tag{6.50}$$

При двойном суммировании положим $m = n - k + 1$, тогда

$$n^2 = m^2 + 2m(k-1) + (k-1)^2$$

и

$$\sum_{k=0}^{\infty} n^2 \sum_{n=k}^{\infty} p_{n+1-k} \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \sum_{m=1}^{\infty} m^2 p_m +$$

$$+ 2 \sum_{k=0}^{\infty} (k-1) \frac{\lambda^k e^{-\lambda}}{k!} \sum_{m=1}^{\infty} m p_m + \sum_{k=0}^{\infty} (k-1)^2 \frac{\lambda^k e^{-\lambda}}{k!} \sum_{m=1}^{\infty} p_m. \tag{6.51}$$

Окончательно имеем

$$S = S + 2\lambda L - 2L + (1-p_0)\lambda(1+\lambda) - 2(1-p_0)\lambda + \\ + (1-p_0) + p_0\lambda(1+\lambda). \tag{6.52}$$

Но

$$p_0 = 1 - \lambda, \tag{6.53}$$

поэтому

$$L = \frac{\lambda^2 - 2\lambda}{2(\lambda - 1)} \tag{6.54}$$

и

$$W_q = \frac{\frac{(\lambda^2 - 2\lambda)}{2(\lambda - 1)} - \lambda}{\lambda} = \frac{\lambda}{2(1-\lambda)}. \tag{6.55}$$

Если необходимо получить результат для общего значения времени обслуживания μ, то, заменив λ на $\frac{\lambda}{\mu} \equiv \rho$, получим $W_q = \frac{\rho}{2\mu(1-\rho)}$.

9. Эверетт [228] приводит метод вычисления p_n для больших значений n, используя то обстоятельство, что для больших n справедливо соотношение $p_{n+1} \approx p_n K$ или, в более общем виде, $p_{n+c} \approx K^c p_n$, где K — постоянная, зависящая от числа каналов c и от λ (время обслуживания снова принято равным одной временнóй единице). После подстановки этих значений в систему уравнений и упрощения можно просуммировать полученные ряды, которые для больших n можно заменить бесконечными рядами. Это приводит к следующему соотношению, которое дает возможность получить K:

$$K = e^{\left(\frac{\lambda}{c}\right)\left(1 - \frac{1}{K}\right)}.$$

Показано, что постоянная K аналогична загрузке ρ для многоканальной системы с экспоненциальным временем обслуживания. Заменив в предыдущей задаче ρ на K, получим оценку для системы с постоянным временем обслуживания, которая подставляется в систему уравнений для нахождения вероятности p_n. В данном случае p_n есть новая оценка. Затем значение p_n используется в этих уравнениях и т. д. до тех пор, пока не достигается стабильное состояние. Этот итерационный метод был успешно проверен Эвереттом.

6.3. Пуассоновский входящий поток, эрланговское время обслуживания

В этом параграфе рассматривается система $M/E_k/1$, а в следующем — система $E_k/M/1$. При эрланговском распределении времени обслуживания можно считать, что требование, поступившее на обслуживание, проходит k фаз обслуживания. Эти фазы имеют одинаковое экспоненциальное распределение с параметром μk. Как уже было показано в гл. 3, плотность распределения суммы k случайных величин, которые являются взаимно независимыми и имеют одинаковое экспоненциальное распределение с параметром μk, равна E_k.

В этом параграфе плотность распределения E_k будет употребляться также и для входящего потока с математическим ожиданием, равным $\frac{\lambda}{k}$, что, по существу, означает, что используется только $\left(\frac{1}{k}\right)$-я часть требований, поступающих по закону Пуассона. В обоих случаях результат, полученный для распределения Эрланга, говорит о том, что можно изучать задачу в более широких предпосылках; это уже пояснялось в гл. 3.

В гл. 4 было рассмотрено решение задачи для многофазового обслуживания в переходном режиме. По соображениям, связанным с историей вопроса, рассмотрим указанный случай при некотором изменении распределения времени обслуживания для стационарного состояния одноканальной системы с пуассоновским входящим потоком с параметром λ при обслуживании требований в порядке поступления. Общее время обслуживания требований имеет эрланговское распределение с плотностью

$$\frac{(\mu k)^k}{\Gamma(k)} e^{-\mu k t} t^{k-1}, \quad 0 \leqslant t < \infty.$$

Напомним, что эта функция есть распределение хи-квадрат с измененным масштабом и $2k$ степенями свободы. Математическое ожидание этой случайной величины равно $\frac{1}{\mu}$ (оно в k раз меньше математического ожидания длительности каждой фазы).

Пусть $P_n(t)$ — вероятность того, что с момента времени t для окончания обслуживания тех требований, которые поступили в систему раньше t, необходимо выполнить n фаз обслуживания. Последнее число мы будем для краткости называть «фазой системы». Каждое поступающее требование увеличивает фазу системы на k, а в каждый момент времени, когда завершается одна фаза обслуживания, фаза системы уменьшается на единицу. Таким образом,

$$\begin{aligned} P_n(t+\Delta t) = P_n(t)\,[1-(\lambda+k\mu)\,\Delta t] + P_{n+1}(t)\,k\mu\Delta t + \\ + P_{n-k}(t)\,\lambda\Delta t, \quad n \geqslant 1. \end{aligned} \tag{6.56}$$

Аналогично находим выражение для $n = 0$. При $j < 0$ имеем $P_j(t) \equiv 0$. Обычным способом эта система приводится к следующей системе уравнений для стационарного состояния:

$$\begin{gathered} \lambda p_0 = k\mu p_1, \\ (\lambda + k\mu)\,p_n = k\mu p_{n+1} + \lambda p_{n-k}, \quad n \geqslant 1. \end{gathered} \tag{6.57}$$

Загрузка системы составляет $\rho = \frac{\lambda}{k\mu}$. При $j < 0$ имеем $p_j \equiv 0$. Если разделить каждый член на μ и ввести производящую функцию

$$P(z) = \sum_{n=0}^{\infty} p_n z^n,$$

то, следуя Кроу, обычным способом получим

$$\begin{aligned} P(z) = \frac{p_0}{1-\rho z\left[\frac{(1-z^k)}{(1-z)}\right]} = p_0 \sum_{n=0}^{\infty} (\rho z)^n \left(\frac{1-z^k}{1-z}\right)^n = \\ = p_0 \sum_{n=0}^{\infty} \rho^n (z + z^2 + \ldots + z^k)^n. \end{aligned} \tag{6.58}$$

Так как $P(1) = 1$, то $p_0 = 1 - k\rho$.

Заметим, что при $k = 1$ задача вычисления p_n нам уже знакома. При произвольном k столь же легко найдем:

$$(1-z^k)^n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} z^{ik},$$

$$(1-z)^{-n} = \sum_{j=0}^{\infty} (-1)^j \binom{-n}{j} z^j = \sum_{j=0}^{\infty} (-1)^{2j} \binom{n+j-1}{j} z^j,$$

так как

$$\binom{-n}{j} = (-1)^j \binom{n+j-1}{j}.$$

Следовательно,

$$z^n (1 + z + \ldots + z^{k-1})^n = \sum_{j=0}^{\infty} \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{n+i-1}{j} z^{j+ik+n},$$

и

$$P(z)=(1-k\rho)\sum_{m=0}^{\infty}\rho^{m}\times$$

$$\times\sum_{j=0}^{\infty}\sum_{i=0}^{m}(-1)^{i}\binom{m}{i}\binom{m+j-i}{j}z^{j+ik+m}. \qquad (6.59)$$

Коэффициент при z^n, где $n=j+ik+m$, равен

$$p_n=(1-k\rho)\sum\rho^{m}(-1)^{i}\binom{m}{i}\binom{m+j-1}{j}. \qquad (6.60)$$

Здесь суммирование производится согласно указанному разбиению числа n. Например, при $k=2$, $n=7$ имеем

$$p_7=(1-2\rho)\ (4\rho^4+10\rho^5+6\rho^6+\rho^7).$$

Среднюю величину фазы $L_{\text{фаз}}$ можно получить непосредственно из промежуточных уравнений путем вычисления $\sum\limits_{n=1}^{\infty}n^2p_n$, как это было выполнено для случая постоянного времени обслуживания. Таким образом, имеем

$$(1+\rho)\sum_{n=1}^{\infty}n^2p_n=\sum_{n=1}^{\infty}n^2p_{n+1}+\rho\sum_{n=k}^{\infty}n^2p_{n-k},$$

или, в другой форме записи,

$$(1+\rho)\sum_{n=1}^{\infty}n^2p_n=\sum_{n=1}^{\infty}(n-1)^2p_n+\rho\sum_{n=0}^{\infty}(n+k)^2p_n.$$

Произведя вычисления, получим

$$2(1-k\rho)L_{\text{фаз}}=(1-p_0)+\rho k^2. \qquad (6.61)$$

Средняя величина фазы системы равна

$$L_{\text{фаз}}=\frac{k\rho(k+1)}{2(1-k\rho)}, \qquad (6.62)$$

а среднее время ожидания для одной фазы —

$$W_{\text{фаз}}=\frac{L}{\mu}=\frac{k\rho(k+1)}{2\mu(1-k\rho)}, \qquad (6.63)$$

откуда путем деления на k получим среднее число требований, находящихся в системе, и среднее время пребывания в очереди:

$$L=\frac{\rho(k+1)}{2(1-k\rho)}=\frac{\left(\frac{\lambda}{\mu}\right)(k+1)}{2k\left(1-\frac{\lambda}{\mu}\right)}, \qquad (6.64)$$

$$W_q=\frac{\rho(k+1)}{2\mu(1-k\rho)}=\frac{\left(\frac{\lambda}{\mu}\right)(k+1)}{2k\mu\left(1-\frac{\lambda}{\mu}\right)}. \qquad (6.65)$$

6.4. Одноканальная система с эрланговским входящим потоком и экспоненциальным временем обслуживания

Предположим, что входящий поток имеет эрланговское распределение

$$dA(t)=\frac{\lambda^k}{\Gamma(k)}\,e^{-\lambda t}t^{k-1}\,dt \tag{6.66}$$

с математическим ожиданием, равным $\frac{k}{\lambda}$. Поток с таким распределением эквивалентен пуассоновскому потоку с параметром λ, из которого в систему направляется каждое k-е требование.

Допустим, что время обслуживания имеет экспоненциальное распределение с параметром μ. Р. Джексон и Николс [371] начинают исследование с некоторых упрощающих приемов. Представим состояния системы в виде (n_1, n_2), где n_1 — число требований, поступивших из исходного пуассоновского потока с момента поступления последнего требования на обслуживание, а n_2 — число требований, фактически находящихся в системе. Таким образом, $0\leqslant n_1<k$, а переход $(k-1,\ n_2)\to(0,\ n_2+1)$ характеризует схему входящего потока. Если принять, что $n=kn_2+n_1$, то пара чисел (n_1, n_2) однозначно определяется с помощью целой части величины, заключенной в квадратных скобках,

$$n_2=\left[\frac{n}{k}\right],$$

$$n_1=n-\left[\frac{n}{k}\right]k.$$

Обозначим через $p_n(t)$ вероятность того, что в момент времени t система находится в состоянии n. Тогда уравнения для стационарного состояния будут иметь вид

$$\begin{aligned}
&\lambda p_0-\mu p_k=0, \quad n=0,\\
&\lambda p_n-\mu p_{n+k}-\lambda p_{n-1}=0, \quad 1\leqslant n\leqslant k-1,\\
&(\mu+\lambda)p_n-\mu p_{n+k}-\lambda p_{n-1}=0, \quad n\geqslant k.
\end{aligned} \tag{6.67}$$

Пусть

$$P(z)=\sum_{n=0}^{\infty}p_nz^n \quad \text{и} \quad \rho=\frac{\lambda}{\mu}.$$

Чтобы получить $P(z)$, умножим приведенные выше уравнения на z^n, просуммируем по указанным значениям n и произведем сложение найденных результатов. Получим следующее выражение:

$$(\lambda p_0 - \mu p_k) + \sum_{n=1}^{k-1} (\lambda p_n - \mu p_{n+k} - \lambda p_{n-1}) z^n +$$

$$+ \sum_{n=k}^{\infty} (\mu p_n + \lambda p_n - \mu p_{n+k} - \lambda p_{n-1}) z^n = 0. \qquad (6.68)$$

После упрощения получим

$$\lambda \sum_{n=0}^{\infty} p_n z^n - \mu \sum_{n=0}^{\infty} p_{n+k} z^n - \lambda \sum_{n=1}^{\infty} p_{n-1} z^n + \sum_{n=k}^{\infty} \mu p_n z^n = 0$$

или

$$\lambda P(z) - \frac{\mu}{z^k} \sum_{n=0}^{\infty} p_{n+k} z^{n+k} - \lambda z P(z) + \mu \sum_{n=k}^{\infty} p_n z^n = 0. \qquad (6.69)$$

Дальнейшее упрощение приводит к выражению

$$\lambda P(z) - \lambda z P(z) - \left(\frac{\mu}{z^k} - \mu\right) \sum_{n=k}^{\infty} p_n z^n = 0$$

или

$$\lambda P(z) - \lambda z P(z) - \left(\frac{\mu}{z^k} - \mu\right) \left[- \sum_{n=0}^{k-1} p_n z^n + P(z)\right] = 0, \qquad (6.70)$$

откуда находим

$$P(z) = \frac{\left(\frac{\mu}{z^k} - \mu\right) \sum_{n=0}^{k-1} p_n z^n}{\lambda z - \lambda + \frac{\mu}{z^k} - \mu}.$$

Окончательно имеем

$$P(z) = \frac{\mu (1 - z^k) \sum_{n=0}^{k-1} p_n z^n}{\lambda z^{k+1} - \lambda z^k + \mu - \mu z^k} = \frac{(1 - z^k) \sum_{n=0}^{k-1} p_n z^n}{\rho z^{k+1} - (\rho + 1) z^k + 1}. \qquad (6.71)$$

Нужно вычислить неизвестные вероятности, появившиеся в правой части выражения. Функция $P(z)$ должна сходиться, по крайней мере, внутри единичного круга. В данном случае знаменатель имеет $k+1$ нулей. Применяя к знаменателю теорему Руше, получим, что k этих нулей лежат внутри единичного круга или на его границе.

Таким образом, если $f(z) = -z^k(\rho+1)$, $g(z) = \rho z^{k+1} + 1$, абсолютные значения этих функций на окружности $|z| = 1 + \delta$ удовлетворяют теореме Руше, и, поскольку функция $f(z)$ имеет k нулей внутри этого круга, то сумма функций f и g, равная знаменателю, также имеет там k нулей. Один нуль знаменателя равен

$z=1$, а $k-1$ нулей должны совпадать с нулями $\sum_{n=0}^{k-1} p_n z^n$. В противном случае $P(z)$ не будет сходиться для значений z внутри единичного круга, вследствие того, что какой-нибудь из нулей знаменателя не совпадает с нулем числителя. Следовательно, один нуль знаменателя находится вне единичного круга; обозначим его z_0. Затем, приравняв общие множители числителя и знаменателя, можно записать

$$(z-1)(z-z_0)\sum_{n=0}^{k-1} p_n z^n = A\left[\rho z^{k+1}-(1+\rho)z^k+1\right], \qquad (6.72)$$

где A — постоянная.

Имеем

$$P(z)=\frac{A\sum_{n=0}^{k-1} z^n}{z_0-z}.$$

Определив A из условия $P(1)=1$, получим

$$P(z)=\frac{(z_0-1)\sum_{n=0}^{k-1} z^n}{(z_0-z)\,k}. \qquad (6.73)$$

Дифференцируя эту производящую функцию, можно показать, что

$$p_n=\frac{1-z_0^{-(n-1)}}{k}, \quad n\leqslant k-1,$$

$$p_n=\frac{z_0^{-(n-1)}\left(z_0^k-1\right)}{k}, \quad n\geqslant k. \qquad (6.74)$$

Вероятность того, что в системе находится n требований, равна

$$P_n=\sum_{j=nk}^{nk+k-1} p_j, \qquad (6.75)$$

откуда

$$P_0=1-\rho, \quad P_n=(1-z_0^{-k})\,z_0^{-(n-1)k}. \qquad (6.76)$$

Для пуассоновского входящего потока, т. е. при $k=1$, вероятность Q_n того, что в момент поступления требования в системе находится n требований, равна P_n; при $k>1$

$$Q_n=kp_{kn+k-1}=(1-z_0^k)\,z_0^{-nk},$$

$$Q_0=1-z_0^{-k}. \qquad (6.77)$$

Распределение времени ожидания имеет вид

$$w(t)=\sum_{n}^{\infty} Q_n\frac{\mu_n t^{n-1}}{(n-1)!}e^{-\mu t}=z_0^{-k}\lambda(z_0-1)\exp\left[-\lambda(z_0-1)t\right]. \qquad (6.78)$$

Распределение времени ожидания для тех требований, для которых оно положительно, имеет вид $z_0^k w(t)$ и, как показал Смит (см. гл. 9), является экспоненциальным.

Теперь найдем решение для регулярного входящего потока

$$A(x)=\begin{cases}0, & x<1,\\ 1, & x\geqslant 1.\end{cases}$$

Допустим, что $k\to 0$ и $\lambda\to 0$ таким образом, что $\frac{k}{\lambda}=1$.

Пусть $z_0=1+\frac{y}{k}$ и $\rho y=1-e^{-y}+O(k^{-1})$, откуда

$$z_0=1+\frac{y_0}{k}+O(k^{-2}),$$

где y_0 — действительный положительный корень уравнения $\rho=\frac{(1-e^{-y})}{y}$.

После подстановки в выражение для P_n получим

$$P_0=1-\rho,$$

$$P_n=\rho(1-e^{-y_0})e^{-(n-1)y_0}, \qquad n\geqslant 1,$$

$$Q_n=\rho y_0(1-\rho y_0)^n=(1-e^{-y_0})e^{-ny_0},$$

$$w(t)=e^{-y_0}\mu(1-e^{-y_0})\exp[-\mu(1-e^{y_0})t]. \tag{6.79}$$

6.5. Эрланговский входящий поток, постоянное время обслуживания

Приведенная в гл. 3 методика вычисления распределения времени ожидания для системы с пуассоновским входящим потоком и постоянным временем обслуживания, равным одной временной единице, может быть применена для общего случая, когда

$$dA(x)=\lambda\frac{(\lambda x)^{c-1}}{(c-1)!}e^{-\lambda x}dx.$$

Равенство $c=1$ соответствует частному случаю. Введем оператор

$$\Omega=\frac{(-1)^{c-1}}{(c-1)!}\lambda^c\frac{\partial^{c-1}}{\partial\lambda^{c-1}},$$

с помощью которого после преобразования переменных получим выражение

$$P(w)=\begin{cases}\int\limits_0^\infty P(u)\,\Omega e^{-\lambda(u+1-w)}du, & w<1,\\ \int\limits_{w-1}^\infty P(u)\,\Omega e^{-\lambda(u+1-w)}du, & w\geqslant 1.\end{cases}$$

Преобразование Лапласа—Стилтьеса имеет вид

$$\gamma(s)=\frac{\lambda^c e^{-s}}{(\lambda-s)^c}\gamma(s)-s\int_0^\infty P(u)\,\Omega e^{-\lambda(u+1)}du=\frac{sf_{c-1}(s)}{\lambda^c e^{-s}-(\lambda-s)^c},$$

где $f_{c-1}(s)$ — многочлен $(c-1)$-й степени относительно s.

Можно показать, что при $\left(\frac{\lambda}{c}\right)<1$ знаменатель имеет ровно c нулей (в полуплоскости $R(s)\geqslant 0$); обозначим эти нули

$$s_0=0,\quad s_1,\ \ldots,\ s_{c-1}.$$

Имеем

$$R(s_i)>0,\quad i=1,\ \ldots,\ c-1.$$

Поэтому в данных точках числитель также обращается в нуль. Окончательно имеем

$$\gamma(s)=\frac{\lambda^{c-1}(c-\lambda)}{\lambda^c e^{-s}-(\lambda-s)^c}s\prod_{k-1}^{c=1}\frac{s_k-s}{s_k}.$$

Положив $s\to\infty$, получим (см. определение)

$$P(=0)=\frac{\lambda^{c-1}(c-\lambda)}{\prod\limits_{k=1}^{c-1}s_k}.$$

Разложив $\gamma(s)$ в ряд по степеням s, можно вычислить моменты распределения. Так, среднее время ожидания равно

$$W_q=\sum_{k=1}^{c-1}\frac{1}{s_k}+\frac{1}{2}\,\frac{\lambda^2-c^2+c}{\lambda(c-\lambda)}.$$

Глава 7

ГРУППОВОЕ ПОСТУПЛЕНИЕ ТРЕБОВАНИЙ И ГРУППОВОЕ ОБСЛУЖИВАНИЕ

7.1. Введение

До сих пор мы рассматривали системы массового обслуживания, в которых клиенты прибывали и обслуживались поодиночке. Однако существуют такие ситуации, когда клиенты могут прибывать группами и обслуживание их также должно быть групповым. К примеру, несколько человек могут пойти в ресторан одной компанией и будут обслужены все вместе. Телефонистка может получить несколько междугородных телефонных вызовов одновременно. В одно и то же время лифт обслуживает несколько человек. Даже в том случае, когда самолеты прибывают в аэропорт поодиночке, при рассмотрении процесса обслуживания прибывших пассажиров выясняется, что в некотором смысле можно применять допущения о групповом поступлении требований, как, например, при проведении таможенного досмотра групп пассажиров. Заметим кстати, что пассажиры прибывают на обслуживание не все одновременно, так как идут обычно небольшими группами.

В данном случае мы имеем дело с более общим типом задачи массового обслуживания, частным случаем которой является ординарный входящий поток и обслуживание одиночных требований. Для многих практических приложений единственно реальными являются допущения о групповом поступлении требований и групповом обслуживании.

Мы выведем зависимости для среднего числа ожидающих требований и среднего времени ожидания в случае стационарного состояния одноканальной системы с пуассоновским входящим потоком и произвольным временем обслуживания, в которой требования обслуживаются группами, не превосходящими определенного числа s. Так, всякий раз, когда обслуживающее устройство освобождается, на обслуживание поступает s требований, если число требований, находящихся в очереди, не меньше s. В противном случае на обслуживание поступают все требования, находящиеся в очереди. Будем исходить из работ Бейли и Даунтона [21, 188, 189]. Для наших выкладок потребуется простое понятие о вложенных цепях Маркова, введенное Кендаллом [428].

Кендалл использовал то обстоятельство, что можно исследовать одноканальную систему с пуассоновским входящим потоком в моменты времени, когда требования покидают систему. Эти моменты являются моментами (точками) регенерации (понятие, введенное Пальмом). Согласно методу Кендалла момент времени является точкой регенерации, если информация о поведении процесса до этого момента времени не влияет на прогноз дальнейшего поведения процесса; это марковское свойство, но требуется, чтобы оно было выполнено лишь для отдельных моментов времени. Марковский процесс является именно тем процессом, для которого каждый момент времени является точкой регенерации. Эффект метода Кендалла заключается в том, что задача приводится к задаче для марковской цепи с дискретным временем, несмотря на то, что сам процесс массового обслуживания может и не быть марковским.

Понятие о точках регенерации может быть применено и к моментам поступления требований, если время обслуживания имеет экспоненциальное распределение.

Для произвольного входящего потока и произвольного времени обслуживания единственно возможными точками регенерации являются такие моменты времени, в которые одновременно поступает новое требование и обслуженное требование уходит из системы (что случается исключительно редко), или моменты времени, в которые поступающее требование находит систему свободной (что является случайным событием). Однако в последнем случае весьма сложно установить соотношение между точками регенерации.

Заметим, что необходимо предполагать пуассоновский характер входящего потока, для которого число требований, находящихся в очереди в момент времени, когда обслуженное требование покидает систему, и число требований, поступающих за время следующего периода обслуживания, взаимно независимы (читатель может убедиться в этом на примере вывода формулы Полячека—Хинчина). Для входящих потоков более общего типа можно ожидать наличие зависимости между указанными двумя величинами.

7.2. Пуассоновский входящий поток, групповое обслуживание

Рассмотрим систему с пуассоновским входящим потоком и произвольным распределением времени обслуживания $B(t)$.

Введем вероятности перехода r_{ij}. Каждая из этих постоянных обозначает вероятность того, что следующим состоянием будет E_j (т. е. в очереди будет находиться j требований) при условии, что предыдущим состоянием было E_i. Длина очереди измеряется в моменты времени (точки регенерации), непосредственно предшествующие окончанию обслуживания группы требований, или, что то же самое, в моменты времени, следующие непосредственно за окончанием обслуживания группы требований. Если обозначить через π_j ($j=0, 1, \ldots$) вероятность того, что система в момент времени, предшествующий началу обслуживания, находится в состоянии E_j, то очевидно, что π_j получим из π_i путем умножения π_i на r_{ij} и суммирования по i, чтобы учесть все возможные пути перехода из состояния E_i в состояние E_j. Если система находится в состоянии статистического равновесия, то

$$\pi_j = \sum_{i=0}^{\infty} \pi_i r_{ij}. \tag{7.1}$$

(См. § 7.6, где определяются вероятности p_j.) Заметим, что для неустановившегося состояния в левой части имеем π_j^{n+1}, а в правой — π_i^n.

Так как необходимо получить π_j для всех j, то вначале нужно найти способ определения r_{ij}. Заметим, что мы будем иметь дело с величиной π_j, которая не обязательно будет выражать в любой момент времени вероятности для действительно стационарного состояния.

Вероятность поступления за время t требований пуассоновского потока равна $\frac{(\lambda t)^n e^{-\lambda t}}{n!}$ ($n=0, 1, 2, \ldots$). Пусть $B(t)$ — функция распределения времени обслуживания $\left[\text{ее производная, если она существует, равна } \frac{dB(t)}{dt} = b(t)\right]$. Предположим, что длительности обслуживания, которые одновременно являются промежутками времени между двумя последовательными уходами требований из системы, имеют взаимно независимые распределения. Тогда вероятность поступления n требований за произвольно выбранный промежуток времени обслуживания равна

$$k_n = \frac{1}{n!}\int_0^{\infty} e^{-\lambda t}(\lambda t)^n dB(t), \quad n=0, 1, 2, \ldots, \tag{7.2}$$

где интеграл взят по всем интервалам t, имеющим распределение $B(t)$. Введем обозначение

$$K(z) = \sum_{n=0}^{\infty} k_n z^n = \int_0^{\infty} e^{-(1-z)\lambda t} dB(t) \equiv \beta[\lambda(1-z)]. \tag{7.3}$$

Существует простая связь между $K(z)$ и преобразованием Лапласа—Стилтьеса $\beta(z)$ функции распределения времени обслуживания.

Умножив обе части равенства (7.1) на z^j и просуммировав по j, получим

$$P(z) \equiv \sum_{j=0}^{\infty} \pi_j z^j = \sum_{j=0}^{\infty} z^j \sum_{i=0}^{\infty} \pi_i r_{ij}. \tag{7.4}$$

Так как обслуживание производится группами, содержащими не более s требований, то изменение числа требований, находящихся в системе, от i к j получим, когда все i требований обслуживаются в составе одной группы (при $i \leqslant s-1$) и поступает j требований или когда (при $i \geqslant s$) обслуживается s требований, $i-s$ требований остается в очереди, поступает еще $j-(i-s)$ новых требований и общее число их становится равным j. Наконец, при $i > j+s$ имеем $r_{ij}=0$, так как даже в том случае, если обслуживается s требований, число их не может уменьшиться до j. Таким образом,

$$r_{ij} = \begin{cases} k_j & \text{при } 0 \leqslant i \leqslant s-1, \\ k_{j-(i-s)} & \text{при } j+s \geqslant i \geqslant s, \\ 0 & \text{при } i > j+s \end{cases} \tag{7.5}$$

и $k_j = 0$ при $j < 0$.

Итак, имеем

$$P(z) = \sum_{j=0}^{\infty} z^j \left(\sum_{i=0}^{s-1} \pi_i k_j + \sum_{i=s}^{j+s} \pi_i k_{j-i+s} \right) =$$
$$= K(z) \sum_{i=0}^{s-1} \pi_i + \sum_{j=0}^{\infty} z^j (\pi_s k_j + \pi_{s+1} k_{j-1} + \ldots + \pi_{s+j} k_0). \tag{7.6}$$

Заметим, что $\pi_s k_j + \ldots + \pi_{s+j} k_0$ — свертка двух последовательностей. Мы воспользуемся тем обстоятельством, что производящая функция свертки равна произведению производящих функций исходных последовательностей. Производящая функция последовательности $\{k_j\}$ равна $K(z)$, а последовательности $\{\pi_{s+j}\}$ —

$$\sum_{j=0}^{\infty} \pi_{j+s} z^j = \sum_{i=s}^{\infty} \pi_i z^{i-s}.$$

Далее,

$$z^{-s} \sum_{i=s}^{\infty} \pi_i z^i = \left[P(z) - \sum_{i=0}^{s-1} \pi_i z^i \right] z^{-s}. \tag{7.7}$$

Окончательно получаем соотношение

$$P(z) = K(z) z^{-s} \left[P(z) + \sum_{i=1}^{s-1} \pi_i (z^s - z^i) \right]. \tag{7.8}$$

Разрешив его относительно $P(z)$, находим

$$P(z) = \frac{\sum_{i=0}^{s-1} \pi_i (z^s - z^i)}{\frac{z^s}{K(z)} - 1}. \tag{7.9}$$

Здесь неизвестные вероятности $\pi_0, \ldots, \pi_{s-1}$ определяются с учетом того обстоятельства, что для сходимости функции $P(z)$ внутри единичного круга и на его границе нули числителя должны совпадать с нулями знаменателя внутри круга $|z| = 1 + \delta$ для малых положительных значений δ. Эти условия выполняются для примеров, рассматриваемых ниже.

В этом параграфе будем считать, что длительность обслуживания распределена по закону хи-квадрат с $2k$ степенями свободы, т. е. что она имеет эрланговское распределение

$$b(t) = \frac{\mu^k}{\Gamma(k)} t^{k-1} e^{-\mu t}, \quad 0 \leqslant t < \infty, \tag{7.10}$$

где среднее время обслуживания $\frac{k}{\mu}$ связано с загрузкой соотношением

$$\rho \equiv \frac{\lambda k}{\mu s} = \frac{\lambda}{s}\left(\sum_{j=0}^{s} j\pi_j + s\sum_{j=s+1}^{\infty}\pi_j\right).$$

Среднее число требований, поступающих за время обслуживания, равно среднему числу фактически обслуженных требований в группе. Заметим, что в такой системе стационарное состояние может наступить при $\rho < 1$, и будем полагать, что это условие выполняется.

Теперь, используя (7.2) и (7.3), получим

$$K(z) = \left[1 + \frac{\rho s (1-z)}{k}\right]^{-k}. \tag{7.11}$$

Если рассмотреть знаменатель выражения (7.9), то можно показать, что внутри круга $|z| \leqslant 1$ он имеет $s-1$ простых нулей z, отличных от $z=1$.

Замечание. Для определения числа нулей к функциям z^s и $z^s - K(z)$ внутри круга $|z| = 1+\delta$ и на его границе применяется теорема Руше. Внутри единичного круга и на его границе не может быть кратных нулей: в противном случае как знаменатель, так и его производная будут обращаться в нуль, и, разделив оба эти выражения, мы получили бы $\rho > 1$, что невозможно.

При $z \to 1$ имеем $P(z) \to 1$. Таким образом, применяя к (7.9) правило Лопиталя при $z \to 1$, получим

$$\sum_{i=0}^{s-1} (s-i)\pi_i = s - \rho s. \tag{7.12}$$

Эти уравнения совместно с $s-1$ уравнениями, которые получаются при подстановке в числитель соответствующих нулей знаменателя, образуют невырожденную систему уравнений (т. е. определитель матрицы коэффициентов не обращается в нуль), с помощью которой можно определить вероятности π_j, а затем нормировать их. Если коэффициенты, соответствующие данному значению π_j, записаны в столбец, то вычтя из каждого столбца соседний с ним левый столбец, получим новый определитель, и, наконец, из каждой строки вынесем общие множители.

Получим:

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ 1 & z_1 & \dots & z_1^{s-1} \\ \dots & \dots & \dots & \dots \\ 1 & z_{s-1} & \dots & z_{s-1}^{s-1} \end{vmatrix} \prod_{i=1}^{s-1} (z_i - 1). \tag{7.13}$$

Этот определитель не обращается в нуль, так как все значения z_i различны и ни одно из них не равно единице.

Рассмотрев (7.11), заключаем, что знаменатель в формуле (7.9) имеет k нулей z_j, лежащих вне единичного круга, так как всего имеется $s+k$ нулей. Следовательно, после сокращения s общих множителей числителя и знаменателя, т. е. $z-1$ и $z-z_i$ ($i = 1, \dots, s-1$) [в противном случае для этих значений не существовало бы значение $P(z)$ внутри единичного круга], имеем

$$P(z) = \frac{A}{\prod_{j=s}^{s+k-1} (z_j - z)}, \tag{7.14}$$

где A — коэффициент пропорциональности.

Положив $z=1$ и используя условие $P(1)=1$, можно также записать

$$P(z) = \prod_{j=s}^{s+k-1} \frac{z_j - 1}{z_j - z}, \tag{7.15}$$

откуда путем разложения на простые дроби можно получить π_j.

Замечание. Краткое изложение методики нахождения нулей функции см. в работе Даунтона [189].

Подставив в выражение для $P(z)$ значение $z = e^{\theta}$, получим моментную производящую функцию $M(\theta)$, а взяв логарифм от последней, получим производящую функцию семиинвариантов $\psi(\theta)$. Среднее число требований, находящихся в очереди, найдем, взяв первую производную от производящей функции семиинвариантов по θ при $\theta = 0$. В результате получим

$$\sum_{j=s}^{s+k-1} (z_j - 1)^{-1}. \tag{7.16}$$

Вторая производная дает дисперсию

$$\sum_{j=s}^{s+k-1} z_j (z_j - 1)^{-2}. \tag{7.17}$$

Упражнение. Проверьте справедливость последних двух формул.

Для экспоненциального времени обслуживания в формуле (7.15) следует положить $k=1$; получим

$$P(z)=\frac{z_s-1}{z_s-z},$$

где z_s — единственный нуль, расположенный вне единичного круга.

Вероятности образуют геометрическую прогрессию.

Для постоянного времени обслуживания принимаем $k\to\infty$ и $s\to\infty$, их отношение сохраняется равным единице; полагаем $\rho s=\lambda$. Это приводит к бесконечным произведениям. Положив $k\to\infty$ в выражении для $K(z)$ при непосредственном анализе, получим выражение

$$P(z)=\frac{\sum_{j=0}^{s-1}\pi_j(z^s-z^j)}{z^s e^{\rho s(1-z)}-1}. \qquad (7.18)$$

Так как знаменатель в формуле (7.18) имеет s нулей z_i внутри единичного круга и на его границе, то числитель можно записать в следующем виде:

$$\sum_{j=0}^{s-1}\pi_j(z^s-z^j)=\left(\sum_{j=0}^{s-1}\pi_j\right)(z-1)\prod_{i=1}^{s-1}(z-z_i). \qquad (7.19)$$

Совместно с s уравнениями, применявшимися ранее для определения π_j, это соотношение используется при получении выражения для $\sum_{j=0}^{s-1}\pi_j$. Здесь снова употребляется определитель матрицы коэффициентов, где $\sum_{j=0}^{s-1}\pi_j$ считается постоянной величиной.

Так как для определения s неизвестных имеется $s+1$ уравнений, то определитель расширенной матрицы должен обращаться в нуль. Вычитаем из каждого столбца соседний с ним левый столбец; оставляем предпоследний, т. е. s-й, столбец без изменений; выносим за скобки общий множитель из каждой строки и приравниваем нулю определитель, отбросив вынесенные множители (так как их произведение не равно нулю); затем производим разложение по элементам первой строки и сокращаем общие множители.

Имеем выражение

$$\sum_{j=0}^{s-1}\pi_j=(s-\rho s)\prod_{i=1}^{s-1}(1-z_i)^{-1}. \qquad (7.20)$$

Подставив значение равенства (7.20) в (7.18), получим

$$P(z)=\frac{(s-\rho s)(z-1)\prod_{i=1}^{s-1}\frac{(z-z_i)}{(1-z_i)}}{z^s e^{\rho s(1-z)}-1}. \tag{7.21}$$

Аналогичным способом найдем выражения для среднего числа требований, находящихся в очереди,

$$L_q=\frac{s-(s-\rho s)^2}{2(s-\rho s)}+\sum_{i=1}^{s-1}(1-z_i)^{-1} \tag{7.22}$$

и дисперсию

$$\frac{s(s+2\rho s)+6\rho s(s-\rho s)^2-(s-\rho s)^4}{12(s-\rho s)^2}-\sum_{i=1}^{s-1} z_i(1-z_i)^2. \tag{7.23}$$

7.3. Преобразование Лапласа функции распределения времени ожидания

Обозначим через $C_r(w)$ вероятность того, что если к началу обслуживания требование окажется r-м в группе, то его время ожидания будет меньше или равно w. Вероятность того, что требование, которое является r-м в группе, поступает в промежутке времени $(t, t+dt)$, равна $w_r dt$, где w_r — среднее число r-х требований, поступающих за время $\frac{1}{\lambda}$ (кроме того, это — среднее число r-х требований, обслуженных за время $\frac{1}{\lambda}$, так как система находится в статистическом равновесии).

Если $E(t)$ — среднее время обслуживания, то

$$w_r=\sum_{j=r}^{\infty}\frac{\pi_j}{\lambda E(t)}, \tag{7.24}$$

так как вычисление произведено для r требований, и вероятность того, что в очереди находится не меньше r требований (а также среднее число r-х требований, обслуженных в каждом промежутке времени обслуживания), равна $\sum_{j=r}^{\infty}\pi_j$, и при статистическом равновесии среднее число периодов обслуживания за время $\frac{1}{\lambda}$ равно $\frac{1}{\lambda E(t)}$. Заметим, что $\sum_{r=1}^{s} w_r=1$.

Теперь можно записать распределение времени ожидания в виде

$$C(w)=\sum_{r=1}^{s} w_r C_r(w). \tag{7.25}$$

Вероятность того, что за время ω поступит j требований, равна $\frac{(\lambda w)^j e^{-\lambda w}}{j!}$, и, следовательно, выражение

$$\frac{1}{j!}\int_0^\infty (\lambda w)^j e^{-\lambda w} dC_r(w) = \frac{\pi_{j+r}}{\sum_{i=r}^{\infty} \pi_i} \tag{7.26}$$

есть условная вероятность того, что в момент времени w в очереди находится $j+r$ требований, при условии, что имеется не менее r требований. Следовательно,

$$P(z) = \sum_{j=0}^{r-1} \pi_j z^j + \left(\sum_{i=r}^{\infty} \pi_i\right) z^r \int_0^\infty e^{-\lambda w} \sum_{j=0}^{\infty} \frac{(\lambda w z)^j}{j!} dC_r(w) =$$

$$= \sum_{j=0}^{r-1} \pi_j z^j + \left(\sum_{i=r}^{\infty} \pi_i\right) z^r M_r(z). \tag{7.27}$$

В данном случае

$$M_r(z) = \int_0^\infty e^{-(1-z)\lambda w} dC_r(w). \tag{7.28}$$

Умножая обе части равенства (7.28) на w_r и суммируя по r $(1 \leqslant r \leqslant s)$, получим выражение

$$M(z) \equiv \int_0^\infty e^{-(1-z)\lambda w} dC(w) \equiv \gamma[\lambda(1-z)], \tag{7.29}$$

которое связано с преобразованием Лапласа—Стилтьеса функции распределения времени ожидания. Заметим, что

$$M(z) = \sum_{r=1}^{s} w_r M_r(z). \tag{7.30}$$

Умножим (7.27) на z^{s-r} и просуммируем по r $(1 \leqslant r \leqslant s)$. Воспользовавшись (7.24) и (7.9), получим

$$M(z) = \frac{\left[\frac{1}{K(z)}\right] - 1}{E(t)(1-z)} P(z). \tag{7.31}$$

Из (7.29) и (7.31) можно найти преобразование Лапласа—Стилтьеса (с параметром p) функции распределения времени ожидания

$$\gamma(p) = \frac{P\left(1 - \frac{p}{\lambda}\right)}{pE(t)} \left[\frac{1}{\beta(p)} - 1\right], \tag{7.32}$$

где $\beta(p)$ — преобразование Лапласа—Стилтьеса функции распределения времени обслуживания.

Из (7.10), используя то обстоятельство, что $M(1)=1$, находим

$$M(z)=\frac{\mu}{k}\frac{\left[1+\frac{\lambda(1-z)}{\mu}\right]^k-1}{1-z}\prod_{i=s}^{s+k-1}\frac{z_i-1}{z_i-z}, \tag{7.33}$$

где z_i — нули, лежащие вне единичного круга. Искомое преобразование Лапласа—Стилтьеса имеет вид

$$\gamma(p)=\frac{\mu}{kp}\left[\left(1+\frac{p}{\mu}\right)^k-1\right]\prod_{i=s}^{s+k-1}\frac{\lambda(z_i-1)}{\lambda(z_i-1)+p}. \tag{7.34}$$

Пусть $\lambda=1$, тогда из (7.34) найдем среднее время ожидания

$$E(w)=\sum_{i=s}^{s+k-1}\frac{1}{z_i-1}-\frac{k-1}{2\mu} \tag{7.35}$$

и

$$\sigma^2(w)=\sum_{i=s}^{s+k-1}\frac{1}{(z_i-1)^2}+\frac{(k-1)(k-5)}{12\mu^2}. \tag{7.36}$$

Для постоянного времени обслуживания, если $\mu\to\infty$ и $k\to\infty$ и если при этом их отношение $\frac{k}{\mu}=b$ остается постоянным,

$$M(z)=\frac{s-b}{-b}\frac{e^{b(1-z)}-1}{z^se^{b(1-z)}-1}\prod_{i=1}^{s-1}\frac{z-z_i}{1-z_i}. \tag{7.37}$$

Преобразование Лапласа—Стилтьеса имеет вид

$$\frac{s-b}{-b}\frac{e^{bp}-1}{(1-p)^se^{bp}-1}\prod_{i=1}^{s-1}\frac{1-p-z_i}{1-z_i} \tag{7.38}$$

и

$$E(w)=\sum_{i=1}^{s-1}\frac{1}{1-z_i}+\frac{s(b+1-s)}{2(s-b)}, \tag{7.39}$$

$$\sigma^2(w)=\frac{s(6-s+2b)}{12}-\frac{s(5s-8b)}{12(s-b)^2}-\sum_{i=1}^{s-1}\frac{1}{(1-z_i)^2}. \tag{7.40}$$

Для случайной величины q, определяющей число требований, находящихся в очереди, при распределении времени обслуживания по закону хи-квадрат с $2k$ степенями свободы и $\rho=\frac{1}{\mu s}$ при $\lambda=1$ Даунтон [189] получил формулы:

$$\lim_{s\to\infty}\frac{E(q)}{s}=\sum_{j=1}^{k}\frac{1}{s(z_j-1)},$$

$$\lim_{s\to\infty}\frac{\sigma^2(q)}{s^2}=\sum_{j=1}^{k}\frac{1}{[s(z_j-1)]^2},$$

$$\lim_{s\to\infty}\frac{E(w)}{s}=\sum_{j=1}^{k}\frac{1}{s(z_j-1)}-\frac{(k-1)\rho}{2k},$$

$$\lim_{s\to\infty}\frac{\sigma^2(w)}{s^2}=\sum_{j=1}^{k}\frac{1}{[s(z_j-1)]^2}+\frac{(k-1)(k-5)\rho^2}{12k^2}.$$

Первые две величины можно найти с помощью производящей функции моментов распределения длины очереди $\frac{E(q)}{s}$, $\frac{E(q^2)}{s}$ и т. д. В пределе эта функция превращается в преобразование Лапласа функции распределения $\lim\limits_{s\to\infty}\left(\frac{q}{s}\right)$, которое имеет вид

$$P_s(z)=\sum_{j=0}^{\infty}\pi_j\left(1-\frac{z}{s}\right)^j=\prod_{j=1}^{k}\frac{s(z_j-1)}{s(z_j-1)+z}. \qquad (7.41)$$

Вторые две величины находим из преобразования Лапласа—Стилтьеса функции распределения отношения времени ожидания к числу требований в группе s. Это преобразование получим из (7.34), заменив z на $\frac{l}{s}$.

Для случая постоянного времени обслуживания с помощью аналогичных, но более сложных выкладок, находя вычеты (большей частью в простых полюсах), Даунтон получил преобразование Лапласа функции распределения времени ожидания при $s\to\infty$ в виде

$$\frac{1-e^{-\rho z}}{\rho z}. \qquad (7.42)$$

Отсюда

$$\lim_{s\to\infty}\frac{E(w)}{s}=\frac{\rho}{2},$$

$$\lim_{s\to\infty}\frac{\sigma^2(w)}{s^2}=\frac{\rho^2}{12},$$

$$\lim_{s\to\infty}\frac{E(q)}{s}=\rho=\lim_{s\to\infty}\frac{\sigma^2(q)}{s}.$$

Само собой разумеется, $\lim\limits_{s\to\infty}\frac{\sigma^2(q)}{s^2}=0$.

Как показал Бейли, предельные свойства распределения длины очереди применимы и к распределению длины очереди в многоканальной системе со случайным входящим потоком и постоянным временем обслуживания. В данном случае этот способ можно применить и для того, чтобы получить аналогичные результаты для распределения времени ожидания.

Замечание 1. Групповое обслуживание и групповое поступление требований. Недавно Миллер-младший [568] рассмотрел этот вопрос по-новому и исследовал случай группового поступления требований в одноканальную систему при групповом обслуживании по принципу «первым пришел — первым обслужен». Требование, поступающее последним, может присоединиться к группе требований, уже находящихся на обслуживании, если число требований в группе не превышает заданного, или же оно не присоединяется к ней и, следовательно, будет ожидать начала обслуживания в составе следующей группы. Для того, чтобы обслуживание началось, не обязательно, чтобы число требований в группе достигало заданного. Для второго случая получить решение трудно.

Для первого случая, когда требование, поступившее последним, присоединяется к группе, находящейся на обслуживании, Миллер-младший получил

$$K(z)=\sum_{j=0}^{\infty} z^j \int_0^{\infty} a_j(t)\,dB(t)=\int_0^{\infty} \mathrm{e}^{-\lambda t\,[1-Q(z)]}dB(t), \tag{7.43}$$

где $a_j(t)$ — вероятность того, что в промежутке времени $[0, t]$ поступит j требований, и $Q(z)=\sum_{j=1}^{\infty} q_j z^j$, где $q_j=P(N=j)=0$ при $j>n_0$ для некоторых n_0, а N — число требований в группе, поступившей в промежутке времени $[0, t]$.

Из выражения

$$K(z)=\frac{1}{1+\left(\frac{\lambda}{\mu}\right)\left(\frac{1-z}{1-\rho z}\right)} \tag{7.44}$$

он получил производящую функцию вероятностей π_j для стационарного состояния системы, в которой входящий поток имеет геометрическое распределение $(1-p)p^{v-1}$, $0<p<1$, а время обслуживания одного требования имеет экспоненциальное распределение $B(t)=1-\mathrm{e}^{-\lambda t}$. Искомая производящая функция имеет вид

$$\sum_{j=0}^{\infty} \pi_j z^j=\frac{1-p-\frac{\lambda}{\mu}}{1-\left(p+\frac{\lambda}{\mu}\right)z}. \tag{7.45}$$

Следовательно, распределение для стационарного состояния также является геометрическим с параметром $p+\frac{\lambda}{\mu}$.

7.4. Биномиальный входящий поток, произвольное время обслуживания

В этом параграфе используется метод вложенных цепей Маркова.

Излагаемый здесь материал поможет читателю получить лучшее представление об этом методе.

Начнем рассмотрение с упрощенного варианта одной важной задачи, которая не решается аналитическим путем. Затем будет дано аналитическое решение простой задачи массового обслуживания, которая может оказаться полезной при решении первой задачи.

Пусть на борту авианосца находится N самолетов. В течение промежутка времени T в воздухе должно непрерывно патрулировать $m<N$ самолетов. Во время патрулирования самолетов авианосец облегчен.

На авианосце имеется c ремонтно-технических пунктов. Продолжительность обслуживания каждого самолета имеет экспоненциальное распределение, одинаковое для всех самолетов во всех пунктах. Самолеты поступают на обслуживание независимо друг от друга с вероятностью p. Распределение числа самолетов, требующих обслуживания, подчиняется биномиальному закону. Самолет, совершающий посадку в конце полетного времени T, поступает на обслуживание. Сколько самолетов должно быть на борту авианосца, чтобы с заданной вероятностью π в любой момент времени в течение всего периода операции t патрулировало m самолетов?

Этот пример можно использовать для рассмотрения биномиального входящего потока. Для углубленного изучения биномиального потока читатель отсылается к многочисленным работам Полячека. Здесь приводится лишь упрощенный анализ.

Мейслинг [561] рассмотрел одноканальную систему при обслуживании требований в порядке поступления. Этот автор предположил, что ось времени разбита на интервалы определенной длины, в каждом из них может поступить одиночное требование с вероятностью p; поступление требований в различных интервалах времени — события независимые. Вероятность отсутствия требования в течение интервала равна $q=1-p$. Требования имеют взаимно независимые одинаково распределенные длительности обслуживания.

Получив среднее число требований, находящихся в очереди, и среднее время ожидания, Мейслинг переходит к пределу и находит решение для непрерывного времени при пуассоновском входящем потоке. Если промежуток времени t_n рассматривается как состоящий из n интервалов длительности Δt, то вероятность того, что за время t_n поступит m требований, равна

$$A_{m,n}=\binom{n}{m}p^m q^{n-m},\quad 0\leqslant m\leqslant n. \tag{7.46}$$

При $m > n$ эта вероятность равна нулю. Нетрудно показать, что $E[m] = np$ и

$$E[m(m-1)] = n(n-1)p^2.$$

Математическое ожидание числа требований составляет

$$\lambda = \frac{E[m]}{n\Delta t} = \frac{p}{\Delta t}. \qquad (7.47)$$

Длительность обслуживания t имеет распределение, при котором время принимает значения, кратные Δt. Она задается вероятностями b_n ($n=0, 1, \ldots$) того, что время обслуживания равно $n\Delta t$ (точнее, принадлежит соответствующему полуинтервалу с исключенным левым и включенным правым концом). Заметим, что

$$\sum_{n=0}^{\infty} b_n = 1, \quad E(t) = \Delta t \sum_{n=0}^{\infty} nb_n, \quad E[t(t-\Delta t)] =$$

$$= \overline{\Delta t}^2 \sum_{k=0}^{\infty} n(n-1) b_n,$$

где $E(t)$ — математическое ожидание, а $E[t(t-\Delta t)]$ — дисперсия времени обслуживания.

Загрузка системы (отношение интенсивности поступления требований к интенсивности обслуживания) $\rho = \lambda E(t)$; в этом случае предполагается, что $\rho < 1$.

Используя изложенную ранее в общих чертах теорию вложенных цепей Маркова, получим вероятность того, что за время обслуживания одного требования поступит m требований:

$$k_m = \sum_{n=0}^{\infty} A_{m,n} b_n. \qquad (7.48)$$

Вероятности r_{ij} те же, что и полученные ранее, с тем лишь отличием, что в данном случае число требований в группе равно единице. Из рассмотрения систем с групповым поступлением требований и групповым обслуживанием имеем

$$P(z) = \frac{\pi_0 K(z)(z-1)}{z - K(z)}, \qquad (7.49)$$

где

$$K(z) = \sum_{m=0}^{\infty} k_m z^m = \sum_{m=0}^{\infty} z^m \sum_{n=m}^{\infty} \binom{n}{m} p^m q^{n-m} b_n =$$

$$= \sum_{n=0}^{\infty} b_n \sum_{m=0}^{n} \binom{n}{m} (zp)^m q^{n-m} = \sum_{n=0}^{\infty} b_n (zp+q)^n. \qquad (7.50)$$

Так как $P(1) = \sum_{j=0}^{\infty} \pi_j = 1$ и $K(1) = 1$, $K'(1) = \rho$, то, применяя к формуле (7.49) правило Лопиталя (при $z \to 1$), получим $\pi_0 = 1 - \rho$.

Среднее число требований, находящихся в очереди, получим, вычислив $P'(1)$. Это потребует повторного применения правила Лопиталя. В выражении

$$K''(1) = \sum_{n=0}^{\infty} b_n p^2 n(n-1)(p+q)^n = \lambda^2 E[t(t-\Delta t)]$$

используются два полученных ранее выражения, определяющие λ и $E[t(t-\Delta t)]$. Получим

$$L = P'(1) = \rho + \frac{\lambda^2 E[t(t-\Delta t)]}{2(1-\rho)}. \tag{7.51}$$

При $\Delta t \to 0$ биномиальное распределение превращается в пуассоновское, распределение времени обслуживания становится непрерывным. В этом случае имеем

$$L = \rho + \frac{\lambda^2 E(t^2)}{2(1-\rho)}. \tag{7.52}$$

Среднее число требований, поступающих за среднее время ожидания и среднее время обслуживания, равно математическому ожиданию числа требований, находящихся в очереди. Это обстоятельство используется в данном случае для получения среднего времени ожидания

$$W_q = \frac{\lambda E[t(t-\Delta t)]}{2(1-\rho)}. \tag{7.53}$$

Если время обслуживания имеет геометрическое распределение

$$b_k = (1-d)d^k,$$

то

$$E(t) = \Delta t \frac{d}{1-d},$$

$$E[t(t-\Delta t)] = \overline{\Delta t}^2 \frac{2d^2}{(1-d)^2} = 2[E(t)]^2;$$

и, следовательно, при $\Delta t \to 0$

$$L = \frac{\rho}{1-\rho},$$

$$W_q = E(t)\frac{\rho}{1-\rho}. \tag{7.54}$$

Это выражение совпадает с формулой, полученной для пуассоновского входящего потока и экспоненциального времени обслуживания.

При постоянном времени обслуживания T_0

$$E[t(t-\Delta t)] = E(t^2) - \Delta t E(t) = T_0^2 - \Delta t T_0,$$

$$L = \rho + \frac{\rho^2\left(1-\frac{\Delta t}{T_0}\right)}{2(1-\rho)}, \tag{7.55}$$

$$W_q = T_0 \frac{\rho\left(1-\frac{\Delta t}{T_0}\right)}{2(1-\rho)}. \tag{7.56}$$

7.5. Групповое поступление требований в систему с произвольным временем обслуживания (случай переходного состояния)

Используя метод вложенных цепей Маркова, Гейвер [284] рассмотрел задачу группового поступления требований в одноканальную систему, находящуюся в нестационарном режиме при произвольном распределении времени обслуживания $B(t)$. Поступающим требованиям приписываются определенные номера для обслуживания в установленном порядке. В этом параграфе дается краткое изложение полученных результатов.

Пусть $\{\Delta_k\}$ — последовательность положительных независимых в совокупности случайных величин, причем

$$P(\Delta_k \leqslant x) = 1 - e^{-\lambda x}, \quad \lambda > 0, \tag{7.57}$$

где Δ_k — промежуток времени между моментами поступления двух последовательных групп.

Время поступления k-й группы записывается в виде

$$\alpha_k = \sum_{l=1}^{k} \Delta_l. \tag{7.58}$$

Число требований в группе равно C_k, где

$$P(C_k = j) = c_j. \tag{7.59}$$

Число требований, поступающих в промежутке времени $(0, t)$, равно

$$A(0, t) = \sum_{0 < \alpha_l < t} C_l. \tag{7.60}$$

Имеем

$$P[A(0, t) = k] = a_k(t) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} c_k^{n*}, \tag{7.61}$$

где c_k^{n*} — n-кратная свертка $\{c_k\}$.

Кроме того, можно ввести производящую функцию

$$a(z, t) = \sum_{k=0}^{\infty} a_k(t) z^k = \exp\{-\lambda t\,[1 - c(z)]\}, \tag{7.62}$$

где

$$c(z) = \sum_{j=0}^{\infty} c_j z^j. \tag{7.63}$$

Состояние системы задается двумя величинами. Первая из них — число требований, находящихся в системе в момент ухода n-го требования, считая с начального момента времени t_0, а вторая — время ухода n-го требования. Последовательность таких состояний образует цепь Маркова. Вначале определяется период занятости. Если $P_{ij}^{(n)}(t)$ — условная вероятность того, что в системе в момент времени, непосредственно следующий за уходом n-го требования (который происходит до момента времени $t+t_0$), находится $j>0$ требований при условии, что в момент времени t_0 в системе находилось $i>0$ требований и число требований за это время не стало равным нулю, то $P_{ij}^{(n)}(t)$ удовлетворяет приведенному выше уравнению Чепмена—Колмогорова:

$$P_{ij}^{(n+1)}(t)=\sum_{h=-1}^{j-1}\int_0^t P_{i,j-h}^{(n)}(t-x)\,a_{h+1}(x)\,dB(x). \tag{7.64}$$

Используя преобразование Лапласа—Стилтьеса и производящую функцию, найдем решение этого уравнения в общем виде. Затем получим совместное распределение числа требований, находящихся в системе в момент времени t после начала периода занятости, и числа требований, обслуженных за это время.

Математическое ожидание и дисперсия длительности периода занятости соответственно равны

$$\frac{\rho}{\lambda\delta_1(1-\rho)} \tag{7.65}$$

и

$$\frac{\sigma^2+\rho\mu_2\left(\frac{\delta_2}{\delta_1}\right)}{1-\rho}, \tag{7.66}$$

где $\rho=\lambda\delta_1\mu_1$ — загрузка системы, μ_1 — математическое ожидание времени обслуживания, μ_2 — его второй момент; $\sigma^2=\mu_2-\mu_1^2$, δ_i $(i=1, 2)$ — соответственно первый и второй моменты распределения $\{c_k\}$.

Математическое ожидание числа требований, обслуженных за период занятости, и его дисперсия равны соответственно

$$\frac{1}{1-\rho} \tag{7.67}$$

и

$$\frac{\rho\left(\frac{\delta_2}{\delta_1}\right)+\rho^2\left(\frac{\sigma^2}{\mu_1^2}\right)}{(1-\rho)^3}. \tag{7.68}$$

Наконец, на основании результатов, полученных для периода занятости, рассматривается весь процесс, включая периоды занятости и периоды простоя, и с помощью теории восстановления исследуются его эргодические свойства. Производящая функция вероятностей p_j стационарного состояния (при $\rho<1$), соответствующих $P_{ij}(t)$ (вероятностям того, что в момент времени t

в системе находится j требований при условии, что в момент времени $t=0$ в системе находилось i требований), равна

$$P(z)=\sum_{j=0}^{\infty} p_j z^j=\frac{(1-\rho)(1-z)\,\beta[\lambda\{1-c(z)\}]}{\beta[\lambda\{1-c(z)\}]-z}, \tag{7.69}$$

где β — преобразование Лапласа—Стилтьеса функции $B(t)$.

Математическое ожидание числа требований, находящихся в системе, равно

$$\rho+\frac{\rho}{2(1-\rho)}\left[\left(\frac{\delta_2}{\delta_1}-1\right)+\rho\left(1+\frac{\sigma^2}{\mu_1^2}\right)\right]. \tag{7.70}$$

Распределение времени ожидания имеет преобразование Лапласа—Стилтьеса

$$(1-p)\left(1+\lambda\frac{1-c\{\beta(z)\}}{z-\lambda[1-c\{\beta(z)\}]}\right). \tag{7.71}$$

Если вся группа будет покидать систему одновременно, то это выражение примет вид

$$\frac{1-\rho}{1-\lambda\left(\frac{[1-c\{\beta(z)\}]}{z}\right)}. \tag{7.72}$$

Это выражение — формула Полячека—Хинчина.

Среднее время ожидания в этом случае равно

$$\frac{\rho}{2(1-\rho)}\left(\frac{\delta_2}{\delta_1}+\frac{\sigma^2}{\mu_1^2}\right)\mu_1, \tag{7.73}$$

а дисперсия составляет

$$\frac{\rho}{(1-\rho)^2}\,[\rho C_1^2+(1-\rho)\,C_2], \tag{7.74}$$

где C_1 и C_2 — первый и второй начальные моменты распределения, преобразование которого имеет вид

$$\frac{1}{\mu_1\delta_1}\,\frac{1-c\{\beta(z)\}}{z}. \tag{7.75}$$

7.6. Доказательство А. Я. Хинчина

Рассмотрим одноканальную систему с пуассоновским входящим потоком и произвольным распределением времени обслуживания. Пусть π_n ($n=0, 1, 2, \ldots$) — вероятность того, что в момент времени, следующий непосредственно после окончания обслуживания, в очереди находится n требований. Вероятность того, что в момент поступления требования в системе находится n требований, также обозначается через π_n, так как ее значение равно предыдущему. Это справедливо для стационарного процесса с одношаговыми переходами (заметим, что групповое обслуживание к этой категории не относится), в котором состояние $n=0$ есть рекуррентное ненулевое состояние.

Чтобы показать это, заметим, что при длительном протекании процесса среднее число поступающих требований стремится к тому же значению, что и среднее число требований, покидающих систему, и, таким образом, спустя некоторое время переходы из состояния E_n в состояние E_{n+1} сопровождаются переходами из состояния E_{n+1} в состояние E_n. Тождество этих двух распределений определяется законом больших чисел.

Следующее далее доказательство есть видоизмененный вариант приведенного А. Я. Хинчиным доказательства того положения, что для стационарного состояния вероятность π_n равна вероятности p_n. Заметим, что вероятность π_n рассматривается для тех моментов времени, когда требования покидают систему, в то время как p_n берется для произвольного момента времени. Вначале установим связь между значениями π_n.

Пусть π_n — вероятность того, что после окончания обслуживания требования в системе находится n требований. В этот момент времени на обслуживание поступает новое требование, и, таким образом, число требований, находящихся в очереди, сокращается с n до $n-1$. Вероятность того, что случайно выбранное требование поступит на обслуживание без ожидания, равна π_0. Необходимым и достаточным условием этого является следующее: в момент поступления предыдущего требования в системе нет ожидающих требований (вероятность этого события равна $\pi_0+\pi_1$), и за время его обслуживания новые требования не поступают.

Вероятность того, что за время обслуживания поступит n требований, равна

$$k_n = \frac{1}{n!}\int\limits_0^\infty (\lambda t)^n e^{-\lambda t}\,dB(t), \quad n \geqslant 0. \tag{7.76}$$

Следовательно,

$$\pi_0 = (\pi_0 + \pi_1)\,k_0. \tag{7.77}$$

В общем виде

$$\pi_n = \pi_{n+1}k_0 + \pi_n k_1 + \ldots + \pi_2 k_{n-1} + (\pi_1 + \pi_0)\,k_n, \quad n \geqslant 0,$$

$$\sum_{n=0}^{\infty} \pi_n = 1. \tag{7.78}$$

Таким образом, вероятность того, что после ухода данного требования в системе останется n требований, равна вероятности того, что после ухода предыдущего требования в системе останется $n+1$ требований и за время обслуживания данного требования не поступят новые требования, плюс вероятность того, что после ухода предыдущего требования в системе останется n требований и за время обслуживания данного требования поступит одно требование.

Пусть p_n — вероятность того, что в любой момент времени в установившемся состоянии в системе находится n требований. А. Я. Хинчин приводит следующее доказательство того, что $p_n = \pi_n$.

Можно показать (см. также гл. 9), что вероятность того, что через промежуток времени y обслуживающее устройство останется занятым тем же самым требованием, равна

$$\mu \int_y^\infty B_c(x)\,dx, \tag{7.79}$$

где $B_c(x) = 1 - B(x)$ и $\frac{1}{\mu}$ — среднее время обслуживания. Так, для длительности пребывания, заключенной между моментами времени y и $y+dy$, эта вероятность равна $\mu B_c(y)\,dy$. В данном случае $\rho = \frac{\lambda}{\mu}$ — вероятность того, что канал занят. Таким образом, если h_n — вероятность того, что за время, пока канал занят, поступит n требований, то

$$h_n = \frac{\rho\mu}{n!} \int_0^\infty (\lambda y)^n e^{-\lambda y} B_c(y)\,dy. \tag{7.80}$$

Так как вероятность p_n рассматривается для произвольного момента времени, а вероятность π_n — для моментов, когда требование покидает систему, то необходимо рассматривать и те требования, которые могут поступить в течение периода занятости канала, имеющего любую возможную длительность.

Таким образом,

$$p_n = \pi_n h_0 + \pi_{n-1} h_1 + \ldots + (\pi_1 + \pi_0) h_{n-1},$$

или просто

$$p_n = \lambda \int_0^\infty B_c(x)\, u_{n-1}(x)\,dx, \tag{7.81}$$

где

$$u_{n-1}(x) = \pi_n e^{-\lambda x} + \pi_{n-1} (\lambda x) e^{-\lambda x} + \ldots + (\pi_1 + \pi_0) \frac{(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}. \tag{7.82}$$

Можно показать, что это выражение удовлетворяет дифференциально-разностному уравнению

$$\frac{du_n(x)}{dx} = \lambda\,[u_{n-1}(x) - u_n(x)]. \tag{7.83}$$

Из соотношения (7.78) имеем

$$\pi_n = \int_0^\infty u_n(x)\,dB(x), \tag{7.84}$$

после интегрирования по частям получим

$$\pi_n = u_n(0) + \int_0^\infty u'_n(x)\,[1 - B(x)]\,dx =$$

$$= \pi_{n+1} + \lambda \int_0^\infty [u_{n-1}(x) - u_n(x)]\,[1 - B(x)]\,dx =$$

$$= \pi_{n+1} + p_n - p_{n+1}. \qquad (7.85)$$

Отсюда имеем

$$\pi_n - p_n = \pi_{n+1} - p_{n+1}. \qquad (7.86)$$

Это соотношение показывает, что разность $\pi_n - p_n$ должна быть постоянной и одинаковой для всех n. Но π_n и p_n — члены сходящихся рядов, поэтому если просуммировать $\pi_n - p_n = \text{const}$ по n, то в левой части равенства (7.86) получим конечную величину, в действительности равную нулю, так как π_n и p_n — вероятности. Следовательно, и для правой части эта величина будет конечной в том и только в том случае, если эта постоянная равна нулю. Поэтому мы заключаем, что $\pi_n = p_n$.

Для многих практических целей представляет интерес выражение

$$p_n = \int_0^\infty u_n(x)\,dB(x). \qquad (7.87)$$

Замечание 2. Подход к системам с групповым поступлением требований и групповым обслуживанием с точки зрения полумарковских процессов. Фабенс [230] исследовал систему $E_s/G/1$ и одноканальную систему с пуассоновским входящим потоком и групповым обслуживанием, когда в группе находится ровно s требований. Пусть для системы $E_s/G/1$, в которой входящий поток имеет гамма-распределение или эрланговское распределение порядка s (s — целое число) с параметром λ, $B(t)$ — распределение времени обслуживания одиночного требования. Применяя метод вложенных цепей Маркова, Фабенс нашел, что π_n (вероятность того, что в момент окончания обслуживания данного требования в системе остается еще n требований) имеет ту же самую величину, что и π_n в системе с групповым обслуживанием. Таким образом, даже в том случае, когда постановки задач различны, вложенные цепи Маркова могут быть одними и теми же.

Имея π_n можно определить p_n — вероятность того, что в любой момент времени в системе находится n требований. Именно в этом случае используется понятие о полумарковском процессе. Пусть $X(t)$ — случайный процесс с непрерывным или дискретным временем, который принимает в любой момент времени одно из счетного числа состояний. Пусть последовательность состояний, или значений $X(t)$ есть цепь Маркова X_n с вероятностями перехода

$$P\{X_{n+1}=j \mid X_n=i\}=q_{ij}. \tag{7.88}$$

Пусть длительность T времени пребывания в состоянии i накануне следующего перехода при условии, что следующий переход происходит в состояние j, есть случайная величина с распределением $W_{ij}(t)$. Будем считать, что первый переход происходит в момент $t=0$. Пусть значение $X(t)$ непрерывно справа, за исключением тех состояний, время пребывания в которых равно нулю. Такой процесс называется полумарковским.

Имеем

$$p_n=\frac{c}{\lambda}\sum_{j=0}^{n}\pi_j \tag{7.89}$$

при $n<s$ и

$$p_n=c\sum_{j=s}^{n}\int_0^\infty \frac{(\lambda t)^{n-j}}{(n-j)!}e^{-\lambda t}[1-B(t)]\,dt+$$

$$+c\pi\int_0^\infty \frac{(\lambda t)^{n-s}e^{-\lambda t}}{(n-s)!}[1-B(t)]\,dt \tag{7.90}$$

при $n\geqslant s$, где

$$\pi=\sum_{j=0}^{s-1}\pi_j;\quad c=ab;\quad a=\frac{1}{1+\pi}; \tag{7.91}$$

$$b=\frac{1}{\sum_j \mu_i r_i};\quad r_i=\begin{cases} a\pi_i & i<s,\\ a(\pi+\pi_s), & i=s,\\ a\pi_i, & i>s,\end{cases} \tag{7.92}$$

$\mu_i=\sum_j q_{ij}\mu_{ij}$, где μ_{ij} — математическое ожидание $W_{ij}(T)$.

Вероятность того, что в системе $E_s/G/1$ время ожидания меньше t, равна

$$P(\leqslant t)=s\sum_{j=0}^{\infty}p_{sj+s-1}\tilde{B}^j(t), \tag{7.93}$$

где

$$\tilde{B}^1(t)=\int_0^t \frac{1-B(x)}{\mu_2}\,dx,$$

а $\tilde{B}^{n+1}(t)$ — свертка функций B и $\tilde{B}^n$ при $n\geqslant 1$.

Для системы с групповым обслуживанием и пуассоновским входящим потоком имеем

$$P(\leqslant t)=\sum_{i=0}^{\infty}\sum_{j=1}^{s} V(t-w\mid W,\ i,\ j)\,\tilde{d}B^{i}(w)\,p_{is+j+1},\qquad(7.94)$$

где

$$V(u\mid w,\ i,\ j)=\sum_{m=j}^{s-1}\Gamma_{s-m}(u)\frac{(\lambda w)^{m-j}e^{-\lambda w}}{(m-j)!}+$$

$$+H(u)\sum_{n=s-j}^{\infty}\frac{(\lambda w)^{n}e^{-\lambda w}}{n!}.\qquad(7.95)$$

Здесь

$$\Gamma_n(u)=\int_{0}^{n}\frac{\lambda^{n}x^{n-1}e^{-\lambda x}}{(n-1)!}\qquad(7.96)$$

и $H(u)$ — функция, равная 0 при $u<0$ и равная 1 при $u\geqslant 0$.

Для системы с групповым обслуживанием и входящим потоком, имеющим гамма-распределение, время ожидания можно определить, комбинируя те два метода, которые использовались для вывода полученных выше величин. Средняя длительность периода занятости для системы $E_s/G/1$ и для системы с групповым обслуживанием одинакова и равна

$$\frac{\sum_{i=0}^{s-1}\pi_i(s-i)}{\lambda\sum_{j=0}^{s-1}\pi_j}\left(\frac{1}{\sum_{i=0}^{s-1}p_i}-1\right).\qquad(7.97)$$

Если система с пуассоновским входящим потоком и групповым обслуживанием имеет конечное число состояний N, то матрица вероятностей перехода вложенной цепи является усеченной. В последнем, т. е. в N-м, столбце усеченной матрицы каждая вероятность перехода Q_i равна сумме вероятностей, записанных в ее строке и опущенных при переходе от бесконечной матрицы к усеченной, включая и элемент, который занимает то же самое положение, что и в бесконечной матрице. Таким образом, обозначая звездочкой вероятности для конечного числа состояний, имеем

$$\pi_n^{*}=\pi_n,\qquad(7.98)$$

$$\pi_N^{*}=\frac{\sum_{j=0}^{N-s-1}Q_{N-j}\pi_{s+j}^{*}+Q_N\pi^{*}}{1-Q_s}\qquad(7.99)$$

и

$$p_N^* = c\sum_{j=s}^{N}\pi_j^*\int_0^\infty \sum_{m=N}^{\infty}\frac{(\lambda t)^{m-j}e^{-\lambda t}}{(m-j)!}[1-B(t)]\,dt +$$

$$+ c\pi^*\int_0^\infty \sum_{m=N}^{\infty}\frac{(\lambda t)^{m-s}e^{-\lambda t}}{(m-s)!}[1-B(t)]\,dt. \qquad (7.100)$$

Задачи.

1. Пусть в системе с пуассоновским входящим потоком при обслуживании группами, содержащими по s требований, время обслуживания постоянно:

$$B(t) = \begin{cases} 0, & 0 \leqslant t < s, \\ 1, & s \leqslant t < \infty \end{cases}.$$

Покажите, что $K(z) = e^{-\rho s(1-s)}$.

2. Джейсуол [382] определяет p_{nr} как вероятность того, что в стационарном состоянии в очереди ожидает n требований, а обслуживание находится в r-й фазе. Он исследовал систему, в которой на обслуживание поступают группы по s требований; если же в очереди находится менее s требований, то очередь также поступает на обслуживание в момент освобождения обслуживающего устройства. Джейсуол предполагает, что обслуживание состоит из j фиктивных фаз. Группа может по выбору поступить в любую фазу. Если с вероятностью C_r будет выбрана r-я фаза, то после окончания обслуживания группа должна перейти в $(r-1)$-ю фазу и т. д. до первой фазы. Входящий поток — пуассоновский с параметром λ. Время обслуживания в каждой фазе — экспоненциальное с параметром μ, обслуживание групп в одноканальной системе производится по принципу «первым прибыл — первым обслужен». Покажите, что уравнения имеют вид

$$-(\lambda+\mu)p_{nr} + \lambda p_{n-1,\,r} + \mu C_r p_{n+s,\,1} = 0, \quad n > 0, \quad 1 < r < j,$$

$$-(\lambda+\mu)p_{nj} + \lambda p_{n-1,\,j} + \mu C_j p_{n+s,\,1} = 0, \quad n > 0,$$

$$-(\lambda+\mu)p_{0r} + \mu p_{0,\,r+1} + \mu C_r \sum_{m=1}^{s} p_{m1} + \lambda C_r p_0 = 0, \quad 1 \leqslant r < j,$$

$$-(\lambda+\mu)p_{0j} + \mu C_j \sum_{m=1}^{s} p_{m1} + \lambda C_1 p_0 = 0,$$

$$-\lambda p_0 + \mu p_{01} = 0,$$

где p_0 — вероятность того, что обслуживающее устройство свободно.

Джейсоул получил уравнение для производящей функции.

Совсем недавно он рассмотрел решение этой задачи для неустановившегося состояния [383]. Уравнения для переходного состояния читателю предлагается вывести самостоятельно.

Пусть преобразование Лапласа (с параметром α) производящей функции $P_{nr}(t)$ имеет вид

$$Q_n^*(x, \alpha) = \sum_{r=1}^{j} x^r P_{nr}^*(\alpha),$$

и пусть

$$F(x, y, \alpha) = \sum_{n=0}^{\infty} y^n Q_n^*(x, \alpha),$$

$$x = \frac{\mu}{\mu+\alpha+\lambda-\lambda y}.$$

Тогда, умножая эти равенства на x и y в соответствующих степенях и суммируя по r и n $(1 \leqslant r \leqslant j,\ 0 \leqslant n < \infty)$, после упрощения получим

$$G(y, \alpha) \equiv \sum_{n=0}^{\infty} y^n P_{n1}^*(\alpha) =$$

$$= \frac{\mu \sum\limits_{n=1}^{s-1} P_{n1}^*(\alpha)(y^s - y^n) + P_0^*(\alpha)(\lambda y^s - \alpha - \lambda) + 1}{\mu \left\{ \dfrac{y^s}{\sum\limits_{r=1}^{j} C_r \left[\dfrac{\mu}{(\mu + \alpha + \lambda - \lambda y)} \right]^r - 1} \right\}}.$$

Применив к знаменателю теорему Руше, найдем, что знаменатель имеет s различных нулей внутри единичного круга. Окончательно имеем

$$\sum_{n=0}^{\infty} y^n Q_n^*(1, \alpha) = \frac{\mu \left\{ \dfrac{1}{\sum\limits_{r=1}^{j} C_r \left[\dfrac{\mu}{(\mu + \alpha + \lambda - \lambda y)} \right]^r - 1} \right\}}{\alpha + \lambda - \lambda y} G(y, \alpha).$$

Заметим, что $Q_n^*(1, \alpha)$ — преобразование Лапласа вероятности того, что в момент времени t имеется n требований, ожидающих в очереди.

Джейсуол предпринял попытку приспособить это решение к эрланговскому распределению времени обслуживания; однако оказалось невозможным получить решение в явном виде. Он воспроизвел известную формулу, полученную Кларком, приведенную в гл. 4, для одноканальной системы с экспоненциальным временем обслуживания, где $C_r = 1$ при $r = 1$ и $C_r = 0$ при $r \neq 1$.

Получите эту формулу с помощью предыдущих выражений.

Глава 8

ЧИСЛО ТРЕБОВАНИЙ, НАХОДЯЩИХСЯ В СИСТЕМЕ; ДЛИТЕЛЬНОСТЬ ПЕРИОДА ЗАНЯТОСТИ; ЧИСЛО ТРЕБОВАНИЙ, ОБСЛУЖЕННЫХ В ТЕЧЕНИЕ ПЕРИОДА ЗАНЯТОСТИ В СИСТЕМЕ С ПУАССОНОВСКИМ ВХОДЯЩИМ ПОТОКОМ И ПРОИЗВОЛЬНЫМ РАСПРЕДЕЛЕНИЕМ ВРЕМЕНИ ОБСЛУЖИВАНИЯ

8.1. Введение

Рассмотрим для одноканальной системы с пуассоновским входящим потоком и произвольным распределением времени обслуживания среднее число требований, находящихся в системе, распределение длительности периода занятости и вероятность того, что в этом периоде будет обслужено данное число требований.

8.2. Число требований в системе, находящейся в стационарном состоянии, при обслуживании требований в порядке поступления

Предположим, что p_j — вероятность того, что в произвольный момент времени стационарного режима система находится в состоянии E_j ($j=0, 1, 2, \ldots$), т. е. имеется j требований, ожидающих в очереди и находящихся на обслуживании.

Приведенный здесь вывод основан на доказательстве А. Я. Хинчина, изложенном в гл. 7, где было показано, что в стационарном состоянии вероятность p_j совпадает с π_j — вероятностью того, что в момент окончания обслуживания система переходит в состояние E_j. При рассмотрении систем с групповым поступлением требований и групповым обслуживанием мы использовали производящую функцию

$$P(z)=\frac{\sum\limits_{i=0}^{s-1}\pi_i\,(z^s-z^i)}{\dfrac{z^s}{K(z)}-1}. \tag{8.1}$$

Если положить $s=1$, $\pi_i=p_i$ и $K(z)=\beta\,[\lambda(1-z)]$, то получим производящую функцию вероятностей $\{p_j\}$:

$$P(z)=\frac{p_0(z-1)\,\beta\,[\lambda(1-z)]}{z-\beta\,[\lambda(1-z)]}. \tag{8.2}$$

Можно показать, что при $z \to 1$

$$\frac{1-K(z)}{1-z} \to \frac{\lambda}{\mu},$$

где $\frac{1}{\mu}$ — математическое ожидание времени обслуживания. Следовательно,

$$1 = P(1) = \frac{p_0}{1-\frac{\lambda}{\mu}}.$$

Итак, окончательно имеем выражение

$$P(z) = \frac{\left(1-\frac{\lambda}{\mu}\right)\beta[\lambda(1-z)]}{1-\frac{\{1-\beta[\lambda(1-z)]\}}{(1-z)}}, \tag{8.3}$$

которое справедливо при $\frac{\lambda}{\mu}<1$.

Если $\frac{\lambda}{\mu} \geqslant 1$, то все вероятности p_j равны нулю, так как каждое состояние является нулевым рекуррентным состоянием. Это можно видеть из того, что рассматриваемая цепь является неприводимой; следовательно, все ее состояния относятся к одному и тому же классу. (В частности, E_0 — нулевое рекуррентное состояние.)

8.3. Распределение длительности периода занятости

Нам уже встречалась задача определения функции распределения длительности периодов занятости для одноканальной системы. Эти длительности, описываемые случайными величинами X_1, X_2, ..., имеют взаимно независимые одинаковые распределения, поскольку в то время, когда канал свободен, не происходит ничего такого, что могло бы изменить распределение следующего периода. Период занятости начинается, когда требование поступает на обслуживание, и заканчивается, когда завершается обслуживание последнего требования из очереди, образовавшейся во время этого процесса, если непосредственно после этого ни одно требование не поступает, и, следовательно, канал становится свободным.

Пусть $G(x) = P(X_n \leqslant x)$ — распределение длительности периода занятости одноканальной системы с пуассоновским входящим потоком с параметром λ и произвольным распределением времени обслуживания $B(t)$. Функция $G(x)$ не зависит от того, как производится выбор на обслуживание.

Заметим, что распределение суммы n периодов занятости есть n-кратная свертка функции распределения $G(x)$. Из гл. 3

известно, что n-кратная свертка функции распределения $G(x)$ получена сверткой функции $G(x)$ с равной ей функцией, сверткой результирующей функции опять с функцией $G(x)$ и т. д. Так, если написать $G_1(x)=G(x)$, то получим двухкратную свертку функции $G(x)$ в виде

$$G_2(x)=\int_0^x G_1(x-y)\,dG(y) \tag{8.4}$$

и вообще, n-кратную свертку

$$G_n(x)=\int_0^x G_{n-1}(x-y)\,dG(y),\ n\geqslant 2. \tag{8.5}$$

Вычислим теперь $G(x)$. Длительность периода занятости можно разделить на длительность обслуживания первого требования и длительность периода занятости для требований, поступивших впоследствии. Таким образом, если длительность обслуживания первого требования равна y и за это время поступило n требований (вероятность этого события определяется законом Пуассона), то, как указывалось в гл. 2 для детерминированной модели, эти n требований можно рассматривать отдельно. Первое требование поступает на обслуживание непосредственно после окончания обслуживания начального требования; если за время обслуживания его поступят новые требования, то все они будут обслужены раньше следующего из оставшихся $n-1$ требований. Аналогично рассуждаем и при поступлении на обслуживание рассматриваемого требования.

Таким образом, n требований и те требования, которые поступают за время их обслуживания, обеспечивают длительность периода занятости от момента времени $x-y$ до момента времени x. Но распределение длительности периода занятости для каждого из этих n требований равно $G(x-y)$, а взяв n-кратную свертку, найдем $G_n(x-y)$. Это выражение должно быть умножено на вероятность поступления n требований за время y и просуммировано по всем n.

Чтобы получить $G(x)$, этот результат нужно умножить на $dB(y)$ и проинтегрировать по всем возможным длительностям y. Следовательно, окончательно имеем

$$G(x)=\int_0^x \sum_{n=0}^{\infty} e^{-\lambda y}\frac{(\lambda y)^n}{n!}\,G_n(x-y)\,dB(y), \tag{8.6}$$

где

$$G_0(x)=\begin{cases}1, & \text{если } x\geqslant 0,\\ 0, & \text{если } x<0.\end{cases} \tag{8.7}$$

Применив преобразование Лапласа—Стилтьеса к последнему уравнению и введя обозначение

$$\Gamma(s) \equiv \int_0^\infty e^{-sx} dG(x), \tag{8.8}$$

имеем (см. статью Такача [800])

$$\Gamma(s) = \sum_{n=0}^{\infty} \frac{[\Gamma(s)]^n}{n!} \int_0^\infty e^{(s+\lambda)x} (\lambda x)^n \, dB(x) =$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \lambda^n [\Gamma(s)]^n \beta^{(n)}(s+\lambda) = \beta[s+\lambda-\lambda\Gamma(s)], \tag{8.9}$$

где β — преобразование Лапласа—Стилтьеса функции распределения времени обслуживания. Это функциональное уравнение, которому должно удовлетворять преобразование Лапласа—Стилтьеса функции $G(x)$ для $\mathrm{Re}(s) \geqslant 0$. Такач приводит доказательство того, что это уравнение однозначно определяет функцию $\Gamma(s)$ [при $\Gamma(\infty)=0$], а из этой функции также однозначно определяется $G(x)$. Он показал, что выражение для $G(x)$ справедливо только в том случае, если $\frac{\lambda}{\mu} \leqslant 1$. Время обслуживания с вероятностью, равной $[1 - \lim_{x \to \infty} G(x)]$, может иметь бесконечную длительность. Математическое ожидание времени обслуживания равно $\frac{1}{\mu}$.

В данном случае $\beta(s)$ — убывающая функция при $0 \leqslant s < \infty$, при этом $\beta(0) = 1$ и $\beta(\infty) = 0$. Эта функция дифференцируема при $0 \leqslant s < \infty$, и $\beta'(s)$ возрастает монотонно. Имеем $\beta'(0) = -\frac{1}{\mu}$ и $\beta'(\infty) = 0$. Так как $\beta(s)$ — монотонная функция, то она имеет единственную обратную функцию $\beta^{-1}(x)$ при $0 \leqslant x \leqslant 1$.

8.4. Число требований, обслуженных за период занятости

Пусть f_j — вероятность того, что за период занятости будет обслужено j требований, и $F(u) = \sum_{j=1}^{\infty} f_j u^j$ — ее производящая функция. Умножим вероятность f_j на распределение длительности обслуживания j требований, которое является n-кратной сверткой распределения времени обслуживания, и просуммируем от единицы до бесконечности. Получим $G(x)$. Если снова обозначить через $F(x)$ преобразование Лапласа—Стилтьеса функции $G(x)$, то получим $\Gamma(s)=\sum_{j=1}^{\infty} f_j [\beta(s)]^j$, так как преобразование n-кратной свертки функции есть n-я степень интегрального преобразования этой функции.

Так как можно записать $\beta(s)=u$, т. е. $s=\beta^{-1}(u)$, то эта величина равна $F(u)$; как уже указывалось, она существует при $0 \leqslant u \leqslant 1$. Из формулы для распределения длительности периода занятости имеем

$$F(u)=\Gamma[\beta^{-1}(u)]=\beta\{\beta^{-1}(u)+\lambda-\lambda\Gamma[\beta^{-1}(u)]\}. \qquad (8.10)$$

Это преобразование является линейной функцией своего аргумента, оно равно

$$u\beta[\lambda-\lambda F(u)]. \qquad (8.11)$$

Из определения β имеем

$$F(u)=u\int_0^{\infty} e^{-\lambda t[1-F(u)]} dB(t). \qquad (8.12)$$

Такач приводит доказательство существования и единственности аналитического решения $F(u)$ для $|u| \leqslant 1$ при $F(0)=0$, где $\lim_{u \to 1} F(u)$ равен наименьшему положительному корню уравнения

$$\beta[\lambda(1-x)]=x.$$

Задачи

1. Покажите, что если распределение времени обслуживания равно $1-e^{-\mu x}$, то

$$p_j=\left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^j.$$

Если же время обслуживания равно постоянной величине α, то

$$p_j=(1-\lambda\alpha)\sum_{k=1}^{j}(-1)^{j-k} e^{k\lambda\alpha}\left[\frac{(k\lambda\alpha)^{j-k}}{(j-k)!}+\frac{(k\lambda\alpha)^{j-k-1}}{(j-k-1)!}\right].$$

При $k=1$ множитель в квадратных скобках опускается.

2. Получите формулу Полячека—Хинчина, взяв производную от $P(z)$ по z при $z=0$.

3. Покажите, что если $B(x)=1-e^{-\mu x}$ $(x \geqslant 0)$, то $\beta(s)=\dfrac{1}{\left(1+\dfrac{s}{\mu}\right)}$ и

$$\Gamma(s)=\frac{\lambda+\mu+s-\sqrt{(\lambda+\mu+s)^2-4\lambda\mu}}{2\lambda}.$$

Перед радикалом берется знак минус, так как $\Gamma(\infty)=0$. Покажите, что

$$G'(x)=\frac{1}{x}\sqrt{\frac{\mu}{x}}\exp[-(\lambda+\mu)x]\,I_1(2\sqrt{\lambda\mu}\,x).$$

Покажите также, что при $\frac{\lambda}{\mu} < 1$ средняя длительность периода занятости равна $\frac{1}{\mu - \lambda}$, а в противном случае она равна бесконечности.

4. Если время обслуживания — постоянная величина α, то, положив $y = \lambda\alpha F(u)$, $x = u e^{-\lambda\alpha}\lambda\alpha$, получим $y e^{-y} = x$.

Записав

$$y(x) = \sum_{j=1}^{\infty} \frac{j^{j-1}}{j!} x^j,$$

покажите, что производящая функция $F(u)$ для f_j (вероятностей того, что за период занятости будет обслужено j требований) имеет вид

$$F(u) = \sum_{j=1}^{\infty} e^{-\lambda\alpha j} (\lambda\alpha j)^{j-1} \frac{u^j}{j}$$

(см. приложение теории массового обслуживания к движению автомобильного транспорта). Покажите, что

$$G(x) = \sum_{j=1}^{\left[\frac{x}{\alpha}\right]} \frac{e^{-\lambda\alpha j}}{j!} (\lambda\alpha j)^{j-1}$$

и что

$$f_j = \frac{e^{-\lambda\alpha j}}{j!} (\lambda\alpha j)^{j-1}.$$

Если время обслуживания имеет экспоненциальное распределение с математическим ожиданием, равным α, то покажите, что

$$F(u) = \frac{(1+\lambda\alpha) - \sqrt{(1+\lambda\alpha)^2 - 4\lambda\alpha u}}{2\lambda\alpha},$$

а следовательно,

$$f_j = \frac{1}{2} \binom{2j}{j} \frac{(\lambda\alpha)^{j-1}}{(1+\lambda\alpha)^{2j-1}} \cdot \frac{1}{2j-1}.$$

Глава 9

ВРЕМЯ ОЖИДАНИЯ В ОДНОКАНАЛЬНЫХ И МНОГОКАНАЛЬНЫХ СИСТЕМАХ С ПУАССОНОВСКИМ И ПРОИЗВОЛЬНЫМ ВХОДЯЩИМ ПОТОКОМ И ПРОИЗВОЛЬНЫМ ВРЕМЕНЕМ ОБСЛУЖИВАНИЯ

9.1. Введение

В этой главе мы рассмотрим два типа уравнений, решения которых дают распределение времени ожидания в очереди (исключая время обслуживания) для одноканальной и многоканальной систем при обслуживании требований в порядке поступления. Уравнение первого типа — интегро-дифференциальное уравнение, которое первоначально получил Такач [800]; независимо от него его вывел также Декан [172]. Уравнение Такача дает распределение времени ожидания в системе с пуассоновским входящим потоком и произвольным распределением времени обслуживания для переходного состояния. Вывод Декана распространяется на c-канальную систему, находящуюся в стационарном состоянии.

Уравнение второго типа — это интегральное уравнение Винера—Хопфа, полученное Линдли. С его помощью можно получить распределение времени ожидания в системе, в которую поступает поток с ограниченным последействием (т. е. промежутки времени между требованиями взаимно независимы и имеют одно и то же произвольное распределение), а время обслуживания имеет произвольное распределение. Кифер и Вольфовиц [442] исследовали многоканальную систему с произвольным входящим потоком и произвольным временем обслуживания. Эти результаты применимы и к марковским системам.

Решение уравнения первого типа получается с учетом временной зависимости, а затем находится решение для стационарного случая. Общее решение, о котором упоминалось ранее, рассматривается для случая одноканальной системы, для других случаев его нахождение связано с большими трудностями.

9.2. Интегро-дифференциальное уравнение Такача

1. Одноканальная система

Рассмотрим одноканальную систему с пуассоновским входящим потоком с параметром λ и произвольным распределением времени обслуживания (одинаковым для всех требований) при обслуживании требований в порядке поступления. При выводе уравнения для этого процесса будем следовать интуитивному рассуждению Декана. Вначале сделаем важное замечание о том, что функция $W(t)$, которая описывает время ожидания в очереди в момент времени t, в моменты поступления новых требований (если время обслуживания не равно нулю) имеет положительные скачки на величину, равную времени обслуживания вновь поступившего требования. Если новые требования не поступают, то функция $W(t)$ стремится к нулю с угловым коэффициентом, равным минус единице. Имеем марковский процесс с непрерывным параметром t. Заметим, что функция $W(t)$ — это не вероятность ожидания в течение точно установленного промежутка времени, а промежуток времени, в течение которого требование, поступающее в момент времени t, должно ожидать перед поступлением на обслуживание при обслуживании требований в порядке поступления. Подобные замечания относятся и к случаю многоканальной системы.

Выведем уравнение для одного канала. Пусть $P(w, t)$ — вероятность того, что требование ожидает в очереди в течение времени $W(t) \leqslant w$ при условии, что оно поступило в момент времени t.

Рассмотрим большое число N одноканальных систем с одинаковым распределением входящего потока и одинаковым распределением времени обслуживания, действующих в одно и то же время. Необходимо выразить $P(w, t+\Delta t)$ через $P(w, t)$ и $P(w+\Delta t, t)$. В момент времени t эти N систем можно разделить на две группы:

1. Системы, для которых время ожидания $W(t) \leqslant w$. Число их равно $NP(w, t)$ [т. е. N умножается на долю тех требований, которые ожидают в течение времени $W(t) \leqslant w$].

2. Системы, для которых время ожидания $W(t) > w$. Число их равно $N - NP(w, t)$.

В момент времени $t+\Delta t$ число систем первой группы становится равным $NP(w+\Delta t, t)$ минус число тех систем, которые имели время ожидания $W(t) \leqslant w$, но вследствие поступления требований в интервале Δt время ожидания в них становится равным $W(t) > w$. Найдем эту величину.

Вначале определим число систем, для которых в момент времени t время ожидания находится в промежутке $(x, x+dx)$. Так как $P(x, t)$ — функция распределения, то после дифференцирования при $x>0$ получим плотность распределения. Число этих систем равно $N\left[\frac{\partial P(x, t)}{\partial x}\right]dx$, если $x > 0$, или $NP(0, t)$, если $x = 0$.

Предполагается, что: 1) в промежутке времени $(t, t+\Delta t)$ время ожидания превзойдет величину w, если за время Δt поступит одно требование, и 2) время обслуживания y этого требования, сложенное с x, превзойдет w (т. е. $x+y>w$, следовательно, $y>w-x$). Поэтому нужно умножить число систем, для которых время ожидания равно x, на вероятность поступления одного требования за время Δt, т. е. на $\lambda\Delta t$, и на вероятность того, что время обслуживания этого требования превзойдет $w-x$. Если плотность распределения y равна $b(y)$, то вероятность последнего события равна $\int\limits_{w-x}^{\infty} b(y)\,dy$.

Поэтому для фиксированного значения времени ожидания $w>0$ число систем, которые перейдут из первой группы во вторую, определяется выражением

$$\left[N\lambda\Delta t\,\frac{\partial P(x,t)}{\partial x}\,dx\right]\left[\int\limits_{w-x}^{\infty} b(y)\,dy\right]=N\lambda\Delta t\,\frac{\partial P(x,t)}{\partial x}\,dx B_c(w-x), \quad (9.1)$$

которое должно быть просуммировано по всем x, $0<x\leqslant w$. Здесь

$$B_c(y)\equiv 1-\int\limits_0^y b(u)\,du\equiv 1-B(y).$$

Если $x=0$, то выражение для числа систем, которые переходят во вторую группу [с временем ожидания $W(t)>w$], имеет вид

$$N\lambda\Delta t P(0,t)\int\limits_w^{\infty} b(y)\,dy=N\lambda\Delta t P(0,t)\,B_c(w). \quad (9.2)$$

Следовательно,

$$NP(w,t+\Delta t)=NP(w+\Delta t,t)-N\lambda\Delta t\int\limits_{+0}^{w}\frac{\partial P(x,t)}{\partial x}B_c(w-x)\,dx-$$
$$-N\lambda\Delta t P(0,t)B_c(w). \quad (9.3)$$

Мы рассмотрели большое число систем N только для того, чтобы получить уравнение. Разделим обе части равенства (9.3) на N, вычтем $P(w,t)$ из обеих частей уравнения, разделим на Δt и перейдем к пределу; тогда получим

$$\frac{\partial P(w,t)}{\partial t}=\frac{\partial P(w,t)}{\partial w}-\lambda\int\limits_{+0}^{w}\frac{\partial P(x,t)}{\partial x}B_c(w-x)\,dx-\lambda P(0,t)B_c(w). \quad (9.4)$$

Для малых положительных Δt применено разложение в ряд

$$P(w+\Delta t,t)=P(w,t)+\frac{\partial P(w,t)}{\partial w}\Delta t+o(\Delta t).$$

Решение данного интегро-дифференциального уравнения должно удовлетворять следующим условиям: $P(w,0)=1$ для всех w, $P(\infty,t)=1$ для всех t.

Интегрируя по частям, получим

$$\frac{\partial P(w, t)}{\partial t}=\frac{\partial P(w, t)}{\partial w}-\lambda P(w, t)+\lambda \int_0^w P(w-v, t)\, dB_c(v). \quad (9.5)$$

Заметим, что если в (9.4) записать $B(t)=1-B_c(t)$ и взять в качестве нижнего предела 0 вместо $+0$ (включая, таким образом, значение интеграла в начальной точке), то получим

$$\frac{\partial P(w, t)}{\partial t}=\frac{\partial P(w, t)}{\partial w}-\lambda P(w, t)+\lambda \int_0^w B(w-x)\, d_x P(x, t). \quad (9.6)$$

Это интегро-дифференциальное уравнение Такача справедливо и при замене λ на $\lambda(t)$. Данное уравнение было выведено также Бенешем.

Обращая внимание на наличие свертки, возьмем от уравнения (9.4) преобразование Лапласа по w (там, где оно существует). Получим

$$\frac{\partial P^*(s, t)}{\partial t}=sP^*(s, t)-P(0, t)-\lambda\left[sP^*(s, t)-P(0, t)\right] B_c^*(s)-$$
$$-\lambda P(0, t) B_c^*(s)=sP^*(s, t)\left[1-\lambda B_c^*(s)\right]-P(0, t),$$

где

$$P^*(s, t)=\int_0^\infty e^{-sw} P(w, t)\, dw;$$

$$\int_0^\infty e^{-sw} \frac{\partial P(w, t)}{\partial w}\, dw=-P(0, t)+sP^*(s, t).$$

Приравняем левую часть уравнения нулю, опустим t, тогда получим уравнение для стационарного состояния. Решив его относительно $P^*(s)$, найдем

$$sP^*(s)=\frac{P(0)}{1-\lambda B_c^*(s)}=\frac{P(0)}{1-\lambda\left[\frac{1}{s}-\frac{b^*(s)}{s}\right]}. \quad (9.7)$$

Но $\lim\limits_{w\to\infty} P(w)=1$, и на основании свойства преобразования Лапласа

$$\lim_{w\to\infty} P(w)=\lim_{s\to 0} sP^*(s).$$

Замечание. В данном случае $P(w)$ играет ту же самую роль, что и $P(<t)$, — вероятность того, что время ожидания меньше t. Так как $\lim\limits_{s\to 0} sP^*(s)=1$, то

$$P(0)=\lim_{s\to 0} \frac{s-\lambda\left[1-b^*(s)\right]}{s}.$$

Это неопределенное выражение; к нему можно применить правило Лопиталя. Напомним, что

$$\lim_{s\to 0} b^*(s) = \lim_{s\to 0}\int_0^\infty e^{-sy} b(y)\,dy = \int_0^\infty b(y)\,dy = 1.$$

Дифференцируя по s отдельно числитель и знаменатель, получим

$$P(0) = 1 - \frac{\lambda}{\mu} \equiv 1 - \rho,$$

где

$$\frac{db^*(s)}{ds} = \int_0^\infty \frac{d}{ds} e^{-sy} b(y)\,dy = -\int_0^\infty y e^{-sy} b(y)\,dy,$$

и при $s\to 0$ найдем среднее значение $b(y)$, которое обозначим через $\frac{1}{\mu}$. Окончательно решение нашей задачи получим в виде

$$sP^*(s) = \frac{1-\rho}{1-\lambda\left\{\frac{[1-b^*(s)]}{s}\right\}}. \tag{9.8}$$

Взяв обратное преобразование, найдем искомое выражение для $P(w)$ — вероятности того, что в стационарном состоянии время ожидания в очереди не больше w.

Пример. Пусть $b(y) = \mu e^{-\mu y}$. Тогда

$$B_c(y) = 1 - \int_0^y \mu e^{-\mu t}\,dt = e^{-\mu y}$$

и

$$B_c^*(s) = \int_0^\infty e^{-sy} e^{-\mu y}\,dy = \frac{1}{s+\mu}.$$

Следовательно, из формулы (9.7) находим

$$P^*(s) = P(0)\left[\frac{1}{s+\mu-\lambda} + \frac{\mu}{s(s+\mu-\lambda)}\right].$$

Заметим, что второй член этого выражения есть преобразование Лапласа свертки показательной и единичной функций. Получим

$$P(w) = P(0)\left[e^{-(\mu-\lambda)w} + \mu\int_0^w e^{-(\mu-\lambda)t}\,dt\right].$$

Легко проверить, что $P(0)=1-\frac{\lambda}{\mu}$, так как $\lim\limits_{w\to\infty} P(w)=1$. Кроме того,

$$W_q \equiv \int_0^\infty w \frac{dP(w)}{dw}\, dw = \frac{\frac{\lambda}{\mu}}{\mu\left(1-\frac{\lambda}{\mu}\right)},$$

что можно показать, вычислив производную $\frac{dP}{dw}$ и проинтегрировав ее. Это знакомое уже нам выражение для среднего времени ожидания, которое можно получить также с помощью формулы Полячека—Хинчина.

Если функция плотности $b(x)$ не существует, то можно использовать преобразование Лапласа—Стилтьеса, т. е.

$$\gamma(s,\ t)=\int_0^\infty e^{-sw} dP(w,\ t),$$

$$\beta(s)=\int_0^\infty e^{-sx}\, dB(x).$$

Здесь $B(x)$ — функция распределения времени обслуживания.

Из определения этого преобразования и преобразования Лапласа известно, что при $t\to\infty$ имеем $sP^*(s)=\gamma(s)$ и $b^*(s)=\beta(s)$, если $b(t)$ существует. Таким образом, мы должны иметь выражение

$$\gamma(s)=\frac{1-\rho}{1-\lambda\left\{\frac{[1-\beta(s)]}{s}\right\}}, \tag{9.9}$$

которое было получено Такачем непосредственно из уравнения (9.6).

Для переходного состояния Такач получил

$$\frac{\partial\gamma(s,\ t)}{\partial t}=\gamma(s,\ t)\,[s-\lambda+\lambda\beta(s)]-sP(0,\ t). \tag{9.10}$$

Он приводит следующее решение этого дифференциального уравнения, содержащего $\lambda(t)$ вместо λ и $\Lambda(t)=\int_0^t \lambda(u)\, du$:

$$\gamma(s,\ t)=\exp\{st-[1-\beta(s)]\,\Lambda(t)\}\left(1-s\int_0^t \exp\left\{-sx+\right.\right.$$
$$\left.\left.+[1-\beta(s)]\,\Lambda(x)\right\}P(0,\ x)\,dx\right). \tag{9.11}$$

Такач показал, что это выражение действительно является преобразованием Лапласа—Стилтьеса исходного дифференциального уравнения.

Кроме того, он показал, что если $\frac{1}{\mu}$ — среднее время ожидания, $\lim\limits_{t\to\infty}\lambda(t)=\lambda$ и $\frac{\lambda}{\mu}<1$, то $\lim\limits_{t\to\infty}P(w,\ t)=P(w)$ существует, не зависит от распределения $P(w,\ 0)$ в начальный момент времени, однозначно определяется преобразованием Лапласа—Стилтьеса (9.9) и, таким образом, является решением поставленной задачи. Если $\frac{\lambda}{\mu}\geqslant 1$, то $P(w)$ не существует, и для всех w имеем $\lim\limits_{t\to\infty}P(w,\ t)=0$.

Вместо использования преобразований Декан предложил численный метод решения. Для стационарного состояния можно записать

$$p(w)=\frac{dP(w)}{dw}.$$

Это приводит к равенству

$$p(w)-\lambda\int_0^w p(x)B_c(w-x)\,du=\lambda\left(1-\frac{\lambda}{\mu}\right)B_c(w), \qquad (9.12)$$

и если $w\to 0$, то

$$p(0)=\lambda\left(1-\frac{\lambda}{\mu}\right).$$

Введем обозначение $q(x)=p(x)B_c(w-x)=p(x)B_c(n\Delta-x)$; разобьем интервал интегрирования на n малых подынтервалов длины Δ и решим полученное интегральное уравнение Вольтерра, применяя метод трапеций

$$\int_0^{n\Delta} q(x)\,dx=\Delta\left\{\frac{q(0)}{2}+q(\Delta)+q(2\Delta)+\ldots+q[(n-1)\Delta]\,\frac{q(n)}{2}\right\}.$$

Так как $B_c(w)$ известно, то, используя уравнение для $p(w)$, можно последовательно находить $p(0)$, $p(\Delta)$, $p(2\Delta)$ и т. д.

Можно допустить, что для переходного состояния $P(w,\ t)=$ $=P_1(w)P_2(w)$. Тогда, подставив это выражение в уравнение, получим дифференциальное уравнение, левая часть которого зависит только от t, а правая часть зависит от w. Это означает, что каждая из этих частей равна постоянной величине. В результате получим два уравнения, каждое с одной переменной.

2. Время ожидания в многоканальной системе для стационарного состояния

Теперь мы снова рассмотрим систему с пуассоновским входящим потоком и произвольным распределением времени обслуживания, одинаковым для всех каналов, при обслуживании требований в порядке поступления. Число каналов равно c. Анализ аналогичен проведенному ранее, с тем лишь отличием, что в данном случае приращение величины $W(t)$ не равно времени обслуживания одиночного требования, поступившего за время Δt. В данном случае это промежуток времени между началом обслуживания этого требования (требование поступило в момент времени t, когда все остальные каналы заняты) и моментом времени, когда канал впервые стал свободен, так как для того, чтобы требование ожидало, нужно, чтобы все каналы были заняты.

Вероятность того, что занятый канал (исключая канал, занятый требованием, поступившим в течение времени Δt) не освободится через промежуток времени y после начала обслуживания этого требования, равна

$$\mu \int_y^\infty B_c(x)\,dx, \tag{9.13}$$

где $\frac{1}{\mu}$ — среднее время, в течение которого канал занят.

В этом выражении значение интеграла равно среднему времени, в течение которого канал будет оставаться занятым после момента времени y. Отношение этих величин и является искомой вероятностью. Производя дальнейшие выкладки и применив правило Лопиталя, получим

$$\int_0^\infty B_c(y)\,dy = \int_0^\infty \left[1 - \int_0^y b(x)\,dx\right] dy = \lim_{s\to 0} \int_0^\infty e^{-sy}\left[1 - \int_0^y b(x)\,dx\right] dy =$$
$$= \lim_{s\to 0}\left[\frac{1}{s} - \int_0^\infty b(x)\,dx \int_x^\infty e^{-sy}\,dy\right] = \lim_{s\to 0}\frac{1 - b^*(s)}{s} = \frac{1}{\mu}.$$

Таким образом, вероятность $H(y)$ того, что все каналы заняты, равна вероятности $B_c(y)$ того, что канал, в который поступило данное требование, остается занятым в течение времени, превосходящего y, умноженной на вероятность того, что остальные $c-1$ каналов также заняты в течение времени, превосходящего y,

$$H(y) \equiv \left[\mu \int_y^\infty B_c(x)\,dx\right]^{c-1} B_c(y). \tag{9.14}$$

В данном случае функция $H(y)$ играет ту же роль, что и функция $B_c(y)$ в случае одноканальной системы.

Как и ранее, рассмотрим для N систем, находящихся в стационарном состоянии, изменение времени ожидания за время Δt. Число систем, в которых время ожидания находится между w и $w+\Delta t$, равно

$$N \frac{dP(w)}{dw} \Delta t.$$

Из этого числа нужно вычесть число систем, в которых время ожидания превосходит w. В данном случае это число равно

$$N\lambda\Delta t \int\limits_{+0}^{w} \frac{dP(x)}{dx} H(w-x)\, dx \tag{9.15}$$

при $x > 0$.

Если $x=0$, то число систем, которые за время Δt переходят во вторую группу, т. е. в группу систем, в которых время ожидания становится больше w, вычисляется следующим образом.

Поступающее требование имеет нулевое время ожидания в тех системах, где находится $(0, 1, \ldots, c-1)$ требований. Число таких систем равно $NP(0)$. Только те из этих систем, в которых находится $c-1$ требований, могут за время Δt перейти в группу систем, имеющих положительное время ожидания (т. е. когда поступающее требование занимает один канал, а в остальных каналах находится $c-1$ других требований). Следовательно, за это время окажутся занятыми все каналы, и вновь поступающее требование будет иметь положительное время ожидания. Число систем этой группы, в которых в промежутке времени $(t, t+\Delta t)$ время ожидания станет больше w, равно

$$Np_{c-1}\lambda\Delta t H(w), \tag{9.16}$$

где p_{c-1} — вероятность того, что в стационарном состоянии в системе находится $c-1$ требований.

Следовательно, уравнение для стационарного состояния c-канальной системы имеет вид

$$0 = \frac{dP(w)}{dw} - \lambda \int\limits_{+0}^{w} \frac{dP(x)}{dx} H(w-x)\, dx - \lambda p_{c-1} H(w). \tag{9.17}$$

Преобразование Лапласа функции $P(w)$, которая входит в это уравнение, имеет вид

$$P^*(s) = \frac{P(0)}{s} + p_{c-1} \frac{\lambda H^*(s)}{s\,[1-\lambda H^*(s)]}. \tag{9.18}$$

Как $P(0)$, так и p_{c-1} неизвестны.

Совместно с соотношением $\lim\limits_{s\to 0} sP^*(s)=1$ рассмотрим следующее разложение в ряд для малых значений s:

$$H^*(s)=\int\limits_0^\infty e^{-us}H(u)\,du=\sum_{n=0}^{\infty}\left[\int\limits_0^\infty \frac{(-u)^n}{n!}H(u)\,du\right]s^n\equiv$$
$$\equiv a_0+a_1 s+\ldots, \tag{9.19}$$

где

$$a_0=\int\limits_0^\infty H(u)\,du. \tag{9.20}$$

Получим выражение

$$P(0)=1-\frac{\lambda a_0}{1-\lambda a_0}p_{c-1}, \tag{9.21}$$

в котором остается найти p_{c-1}. Данный метод не позволяет определить p_{c-1}. Эту постоянную приходится находить по методу Монте-Карло или из каких-либо других соображений.

Замечание. Если $Q(w)=1-P(w)$, то

$$Q^*(s)=\frac{p_{c-1}}{s}\left[\frac{\lambda a_0}{1-\lambda a_0}-\frac{\lambda H^*(s)}{1-\lambda H^*(s)}\right]. \tag{9.22}$$

Взяв обратное преобразование, получим

$$Q(w)=p_{c-1}\left[\frac{\lambda a_0}{1-\lambda a_0}-R(w)\right], \tag{9.23}$$

где $R(w)$ — обратное преобразование второго члена, стоящего в квадратных скобках выражения (9.22), умноженного на $\frac{1}{s}$. Производная $r(w)=\frac{dR(w)}{dw}$ имеет вид

$$r(w)-\lambda\int\limits_0^w r(x)H(w-x)\,dx=\lambda H(w). \tag{9.24}$$

Как указывалось ранее, она может быть найдена численными методами. Заметим, что

$$\frac{dP(w)}{dw}=p_{c-1}r(w), \tag{9.25}$$

т. е. имеем решение с точностью до постоянного коэффициента пропорциональности.

3. Обобщение времени ожидания, принадлежащее Б. В. Гнеденко

Б. В. Гнеденко [981] дал общую формулировку задачи определения времени ожидания. Требование, поступающее в одноканальную систему с обслуживанием в порядке поступления, может либо получить отказ, либо уйти из очереди через некоторое время, либо дождаться начала обслуживания и затем покинуть систему до окончания обслуживания, либо остаться до конца обслуживания. Пусть для такого режима $D(\omega)$ — распределение вероятностей того, что время ожидания в очереди равно w; $G_x(w)$ — распределение вероятностей того, что, прождав в очереди в течение времени x, требование остается на обслуживании в течение промежутка времени, равного w (т. е. обслуживание не заканчивается, и, таким образом, требование ожидает в течение дополнительного промежутка времени w), $G_x(+0)=0$; $B(y)$ — функция распределения времени обслуживания.

Пусть случайная величина $W(t)$ — время ожидания начала обслуживания для требования, поступившего в момент времени t. Запишем

$$P(w,\ t)=P\left[W(t)\leqslant w\right].$$

Допустим, что входящий поток имеет пуассоновское распределение с параметром λ. Рассмотрим условия, при которых справедливо соотношение $W(t+\Delta t)\leqslant w$:

1. $W(t)<w+\Delta t$ и за промежуток времени $(t,\ t+\Delta t)$ не поступает ни одного требования.

2. $W(t)=x<w+\Delta t$; за промежуток времени $t+\varepsilon$ $(0<\varepsilon<\Delta t)$ поступает одно требование, которое покидает систему необслуженным, т. е. оно ожидает в течение времени, меньшего $x-\varepsilon$. (Это определяется с помощью D.)

3. То же, что и 2, с тем лишь отличием, что требование обслуживается неполностью и покидает систему до момента времени $w-x+\Delta t$. Заметим, что до начала обслуживания оно ожидает в течение времени, большего $x-\varepsilon$. (Это определяется с помощью D и G.)

4. То же, что и 3, с тем лишь отличием, что требование остается для завершения обслуживания и, следовательно, находится на обслуживании в течение времени, большего $w-x+\Delta t$. Ясно, что время ожидания начала обслуживания больше $x-\varepsilon$. (Это определяется с помощью D, G и B.)

Переходя к вероятностям, получим

$$\begin{aligned}P(w,\ t+\Delta t)&=(1-\lambda\Delta t)P(w+\Delta t,\ t)+\\&+\int_0^{\Delta t}d\varepsilon\int_0^{w+\Delta t}\{D(x-\varepsilon)+[1-D(x-\varepsilon)]\,G_{x-\varepsilon}(w-x+\Delta t)+\\&+[1-D(x-\varepsilon)]\,[1-G_{x-\varepsilon}(w-x+\Delta t)]\,B(w-x+\\&+\Delta t)\}\,d_xP(x,\ t)+o(\Delta t).\end{aligned}\tag{9.26a}$$

Произведем упрощения, разделим на Δt и затем устремим Δt к нулю. Получим уравнение

$$\frac{\partial P(w,\ t)}{\partial t}=\frac{\partial P(w,\ t)}{\partial w}-\lambda P(w,\ t)+\lambda\int_0^w\{D(x)+[1-D(x)]\,G_x(w-x)+$$
$$+[1-D(x)]\,[1-G_x(w-x)]\,B(w-x)\}\,d_xP(x,\ t),\quad (9.26б)$$

для которого уравнение Такача является частным случаем.

При $t\to\infty$ это уравнение превращается в уравнение для стационарной функции распределения $P(w)$, существование которой можно показать. Это же уравнение получим, приравняв левую часть нулю и опустив аргумент t в правой части. В результате после упрощений имеем

$$\frac{dP(w)}{dw}=\lambda\int_0^w[1-D(x)]\,[1-G_x(w-x)]\,[1-B(w-x)]\,dP(x).\quad (9.26в)$$

Это уравнение имеет единственное решение с разрывом в точке $w=0$ и является абсолютно непрерывным при $w\neq 0$.

Введем некоторые важные функции:

1. Функцию распределения эффективного времени ожидания

$$P_1(w)=P(w)+[1-P(w)]\,D(w),\quad (9.26г)$$

которая является распределением вероятностей совместного появления двух событий: ожидания в очереди и ухода из системы до начала обслуживания.

2. Функцию распределения общего времени пребывания в системе

$$P_2(w)=P(w)-\frac{P'(w)}{\lambda}+[1-P(w)]\,D(w).\quad (9.26д)$$

3. Функцию распределения отношения эффективного времени обслуживания ко времени, которое необходимо для завершения обслуживания,

$$P_3(w)=1-\int_0^\infty\int_0^\infty[1-G_x(w,\ y)]\,dP(x)\,dB(y),\ 0<w<1.\quad (9.26е)$$

4. Распределение вероятностей полного удовлетворения потребности в обслуживании (т. е. требование остается до окончания обслуживания)

$$P_4(w)=\int_0^\infty[1-G_x(w)]\,dP(x).\quad (9.26ж)$$

5. Вероятность потери требования

$$\alpha_1=\int_0^\infty D(x)\,dP(x). \tag{9.26з}$$

6. Вероятность того, что обслуживание не будет выполнено до конца,

$$\alpha_2=\int_0^\infty [1-D(x)]\int_0^\infty G_x(y)\,dB(y)\,dP(x).^{1} \tag{9.26и}$$

Упражнение 1. Плотность распределения времени ожидания требований, получающих отказ, есть дельта-функция, и, следовательно,

$$D(w)=\begin{cases}0, & w\leqslant 0,\\ 1, & w>0.\end{cases}$$

Если время ожидания в очереди имеет фиксированную длительность τ, после чего клиент уходит, т. е. покидает очередь, то, перенеся начало отсчета в точку τ, получим, как и в случае отказов,

$$D(w)=\begin{cases}0, & w\leqslant\tau,\\ 1, & w>\tau,\end{cases}\quad G_x(w)\equiv 0.$$

Если τ — случайная величина, то $G_x(w)=0$. Если время пребывания в системе ограничено величиной τ, то $G_x(w)=1$ при $w+x>\tau$; если τ — случайная величина, то

$$G_x(w)=\frac{D(w+x)-D(x)}{1-D(x)}. \tag{9.26к}$$

Покажите, что для системы с экспоненциальным временем обслуживания с параметром μ в случае, если требования уходят из очереди до начала обслуживания, имеем (заметим, что если требование поступает на обслуживание, то оно остается до его завершения)

$$P(w)=\begin{cases}\dfrac{P(0)}{1-\rho}\left(1-\rho e^{-(\mu-\lambda)w}\right) & \text{при } w<\tau,\\ 1-\rho P(0)\,e^{\lambda\tau-\mu w} & \text{при } w>\tau,\end{cases} \tag{9.26л}$$

где $\rho=\dfrac{\lambda}{\mu}$.

Далее мы приведем результат Баррера, состоящий в том, что

$$P(0)=p_0=\frac{1-\rho}{1-\rho^2 e^{-\tau(\mu-\lambda)}}. \tag{9.26м}$$

---

[1] Изложенное здесь аналитическое решение задачи Б. В. Гнеденко получено в работе И. Н. Коваленко [1012]. — *Прим. ред.*

Покажите также, что

$$P_1(w)=\begin{cases}\dfrac{(1-\rho e^{-(\mu-\lambda)w})}{(1-\rho^2 e^{-(\mu-\lambda)\tau})} & \text{при } w<\tau,\\ 1 & \text{при } w>\tau,\end{cases} \tag{9.26н}$$

$$\alpha_1=\frac{\rho(1-\rho)}{e^{(\mu-\lambda)\tau}-\rho^2}. \tag{9.26о}$$

Здесь α_1 — величина скачка функции $P(w)$ в точке τ. В этой системе массового обслуживания отсутствуют потери требований вследствие неполного обслуживания (т. е. если требование поступило на обслуживание, то обслуживание его производится до конца), следовательно, $\alpha_2=0$;

$$P_2(w)=\begin{cases}\dfrac{(1-e^{-(\mu-\lambda)w})}{(1-\rho^2 e^{-(\mu-\lambda)\tau})}, & 0<w<\tau,\\ 1-\rho\left[1-\dfrac{P(0)}{1-\rho}(1-\rho e^{-(\mu-\lambda)\tau})\right]e^{-\mu(w-\tau)}, & w>\tau.\end{cases} \tag{9.26п}$$

Упражнение 2. Исследуйте разрешимость этого уравнения для стационарного состояния.

9.3. Поток с ограниченным последействием, произвольное время обслуживания

1. Интегральное уравнение Линдли для одноканальной системы

Найдем распределение времени ожидания в очереди для одноканальной системы, в которую поступает поток с ограниченным последействием, а промежутки времени обслуживания взаимно независимы и имеют произвольное распределение; обслуживание требований происходит в порядке поступления.

Обозначим через t_n промежуток времени между моментами поступления n-го и $(n+1)$-го требований, а через s_n — время обслуживания n-го требования.

Предположим, что как t_n, так и s_n — взаимно независимые случайные величины с одинаковыми распределениями и конечными средними значениями. Последовательности $\{t_n\}$, $\{s_n\}$, $n=1, 2, \ldots$, не обязательно должны быть полностью независимыми, но нередко для многих практических приложений предполагается, что они полностью независимы.

Пусть w_n — длительность ожидания n-го требования. Напомним, что справедливо следующее соотношение:

$$w_{n+1}=\begin{cases} w_n+u_n, & \text{если } w_n+u_n>0,\\ 0, & \text{если } w_n+u_n\leqslant 0,\end{cases} \tag{9.27а}$$

где $u_n=s_n-t_n$.

Так как w_n всегда имеет положительное значение, то из приведенного соотношения имеем

$$P_{n+1}(w) \equiv P(w_{n+1} \leqslant w) = P(w_{n+1} = 0) + P(0 < w_{n+1} \leqslant w) =$$
$$= P(w_n + u_n \leqslant 0) + P(0 < w_n + u_n \leqslant w) = P(w_n + u_n \leqslant w) =$$
$$= \int\limits_{w_n + u_n \leqslant w} dP_n(w_n)\, dU(u_n) = \int\limits_{u_n \leqslant w} P_n(w - u_n)\, dU(u_n), \qquad (9.27б)$$

где $P_{n+1}(w)$ — функция распределения времени ожидания $(n+1)$-го требования, а $U(u_n)$ — функция распределения случайной величины u_n.

Это распределение одинаково для всех n.

При $w \geqslant 0$ имеем $P_1(w) = 1$, так как первое требование не ожидает в очереди. Кроме того, имеем

$$P_2(w) = \int\limits_{u_1 \leqslant w} dU(u_1) = P(u_1 \leqslant w),$$

$$P_3(w) = \int\limits_{u_2 \leqslant w} \int\limits_{u_1 \leqslant w - u_2} dU(u_1)\, dU(u_2) = P(u_1 + u_2 \leqslant w,\ u_2 \leqslant w),$$

вообще

$$P_{n+1}(w) = P(u_1 + \ldots + u_n \leqslant w,\ u_2 + \ldots + u_n \leqslant w\ \ldots,\ u_n \leqslant w) =$$
$$= P(u_1 + \ldots + u_n \leqslant w,\ u_1 + \ldots + u_{n-1} \leqslant w,\ \ldots,\ u_1 \leqslant w). \qquad (9.28)$$

Второе уравнение справедливо, если случайные величины u_n независимы в совокупности и имеют одинаковые распределения, следовательно, вероятность останется неизменной при изменении индексов случайных величин u_n.

Запишем

$$U_n = \sum_{k=1}^{n} u_k, \qquad (9.29)$$

тогда

$$P_{n+1}(w) = P(U_k \leqslant w,\ k = 1, 2, \ldots, n) =$$
$$= P(U_n \leqslant w) P(U_{n-1} \leqslant w) \ldots P(U_1 < w). \qquad (9.30)$$

Таким образом, $P_{n+1}(w)$ — монотонно убывающая функция n, значения которой лежат между нулем и единицей. Следовательно, она имеет предел

$$\lim_{n \to \infty} P_{n+1}(w) = \lim_{n \to \infty} P_n(w) = P(w). \qquad (9.31)$$

Можно показать, что этот предел удовлетворяет интегральному уравнению Линдли, которое зависит от распределения u_n:

$$P(w) = \int\limits_{u < w} P(w - u)\, dU(u) = \int\limits_0^\infty P(y)\, dU(w - y). \qquad (9.32)$$

Бесконечный верхний предел означает, что u изменяется до минус бесконечности, а y — до плюс бесконечности после подстановки $y = w - u$. Это замечание справедливо и для переменной $u = s - t$. Можно показать, что $P(w)$ — неотрицательная, неубывающая и непрерывная справа функция, она равна нулю при $w < 0$, а при $w \to \infty$ стремится к единице, и, следовательно, является функцией распределения.

Если $U(w - y)$ — абсолютно непрерывная функция и, следовательно, имеет производную, то получим интегральное уравнение Винера—Хопфа

$$P(w) = \int_0^\infty P(y)\, U'(w - y)\, dy \tag{9.33}$$

где

$$U'(w) = \int_0^\infty b(y)\, a(y - w)\, dy = \int_0^\infty b(y + w)\, a(y)\, dy. \tag{9.34}$$

Заметим, что при отрицательных значениях аргумента функция обращается в нуль. Кроме того, имеем

$$U(w) = \int_0^\infty B(y + w)\, a(y)\, dy = \int_0^\infty B(y)\, a(y - w)\, dy =$$

$$= 1 - \int_0^\infty b(y)\, A(y - w)\, dy = 1 - \int_0^\infty b(y + w)\, A(y)\, dy. \tag{9.35}$$

Здесь $a(t)$ и $A(t)$ — соответственно плотность распределения и функция распределения длительности промежутков времени между моментами поступления требований, а $b(y)$ и $B(y)$ — соответственно плотность распределения и функция распределения времени обслуживания.

Пример. Рассмотрим случай поступления требований через постоянные промежутки времени T. Функция распределения промежутков времени между моментами поступления требований в этом случае имеет вид

$$A(\tau) = \begin{cases} 0 & \text{при } \tau < T, \\ 1 & \text{при } \tau \geqslant T. \end{cases}$$

Допустим, что время обслуживания имеет экспоненциальное распределение

$$b(\tau) = \begin{cases} \mu e^{-\mu\tau} & \text{при } \tau \geqslant 0, \\ 0 & \text{при } \tau < 0. \end{cases}$$

Тогда

$$U(\tau) = 1 - \int_T^\infty b(y + \tau)\, dy = B(T + \tau).$$

Таким образом,

$$U'(\tau) = b(T+\tau).$$

Отсюда

$$P(\tau) = \int_0^{T+\tau} P(x)\, b(T+\tau-x)\, dx.$$

(Заметим, что при отрицательных значениях аргумента функция $b(T+\tau-x)$ обращается в нуль.) Заменив τ на $\tau'-T$, положив $y=\tau'-x$ и, наконец, приняв новое обозначение τ вместо τ', можно записать это выражение в виде

$$P(\tau - T) = \int_0^{\tau} P(\tau - y)\, b(y)\, dy.$$

Подставляя значение $b(y)$, будем искать решение в виде

$$P(\tau) = 1 + c_1 e^{c_2 \tau};$$

для этого решения должны выполняться следующие условия:

$$\frac{1}{\mu} + \frac{c_1}{\mu + c_2} = 0,$$

$$\frac{\mu}{\mu + c_2} = e^{-c_2 T}.$$

Отсюда

$$c_1 = -[1 - P(0)] = -(1 - p_0),$$

$$c_2 = -\mu p_0.$$

Здесь p_0 — ненулевой корень уравнения $1 - p_0 = e^{-\mu p_0 T}$. Таким образом, вероятность $P(\leqslant \tau) \equiv P(\tau)$ того, что время ожидания в очереди меньше τ, равна

$$P(\leqslant \tau) = 1 - (1 - p_0)\, e^{-\mu p_0 \tau}. \tag{9.36}$$

Наконец, среднее время ожидания

$$W_q = \int_0^{\infty} \tau\, dP(\leqslant \tau) = \frac{1 - p_0}{\mu p_0}. \tag{9.37}$$

Для одноканальной системы с входящим потоком, распределенным по закону хи-квадрат с параметрами λ и n, и временем обслуживания, распределенным по закону хи-квадрат с параметрами μ и m, Сиски в своей книге приводит решение уравнения Линдли в виде следующего выражения для распределения времени ожидания [795]:

$$P(w) = 1 + \sum_{k=1}^{m} \alpha_k e^{-\beta_k w}.$$

Это решение является единственным. Здесь β_k — различные корни характеристического уравнения

$$\left(\frac{\mu}{\mu-\beta}\right)^m\left(\frac{\lambda}{\lambda+\beta}\right)^n=1$$

и

$$\alpha_k=-\prod_{\substack{j=1\\ j\neq k}}^{m}\frac{\beta_j}{\beta_j-\beta_k}\left(\frac{\mu-\beta_k}{\mu}\right)^m.$$

Среднее время ожидания равно

$$W_q=-\sum_{k=1}^{m}\frac{\alpha_k}{\beta_k}.$$

Смит [771] исследовал характер распределения времени ожидания в одноканальной системе, находящейся в стационарном состоянии. Мы уже видели, что в случае пуассоновского и регулярного входящих потоков время ожидания имеет экспоненциальное распределение. Оказалось, что экспоненциальный характер распределения времени ожидания появляется и при менее ограничительных условиях.

Если производящая функция распределения является обратной величиной полинома n-й степени, то говорят, что функция распределения принадлежит классу K_n.

Если $U(s)$ — преобразование Лапласа—Стилтьеса функции распределения $U(u)$, где u — разность между временем обслуживания и промежутком времени между требованиями, и $s=\sigma+i\tau$, то имеем следующую теорему Смита. Доказательство ее не приводится, так как оно весьма громоздко.

Теорема 9.1. Если функция распределения времени обслуживания принадлежит классу K_n, то и распределение времени ожидания (времени пребывания в системе) также принадлежит K_n при любых распределениях входящего потока, удовлетворяющих достаточно общим условиям.

Плотность распределения времени ожидания имеет вид

$$\sum_{l=1}^{n}A_l e^{a_l w},$$

где

$$A_l^{-1}=\frac{1}{a_l}\prod_{j\neq l}\left(1-\frac{a_l}{a_j}\right)$$

и a_j — n нулей (предполагается, что они различны) функции $1-\widetilde{U}(s)$ в открытой полуплоскости $\sigma<0$.

Приведем еще одно важное положение, которое является непосредственным следствием этой теоремы.

Теорема 9.2. Если время обслуживания имеет экспоненциальное распределение, то и время ожидания в очереди для тех требований, которым приходится ожидать, также имеет экспоненциальное распределение при любом распределении входящего потока.

Упражнение. Используя теорему 9.1, докажите теорему 9.2.

2. Решение задачи в общем виде с помощью криволинейного интеграла

С помощью остроумного способа Вантюра [861] нашел преобразование Лапласа функции распределения времени ожидания n-го требования. Используя уравнение (9.27a) и обозначив через $\gamma_n(s)$ преобразование Лапласа—Стилтьеса функции распределения времени ожидания n-го требования, получим

$$\gamma_n(s) = E\left[e^{-sw_n}\right]. \tag{9.38}$$

Для положительных действительных значений s можно представить функцию e^{-sw_n} с помощью криволинейного интеграла в виде

$$e^{-sw_n} = -\frac{1}{2\pi i}\oint\limits_C \exp\left[-z(w_{n-1}+u_{n-1})\right]\frac{dz}{z-s} + \varepsilon_n, \tag{9.39}$$

где

$$\varepsilon_n = \begin{cases} 0 & \text{при } w_{n-1}+u_{n-1} > 0, \\ \frac{1}{2} & \text{при } w_{n-1}+u_{n-1} = 0, \\ 1 & \text{при } w_{n-1}+u_{n-1} < 0. \end{cases} \tag{9.40}$$

Контур C идет от $-i\infty$ до $+i\infty$, обходя начало координат по малой полуокружности справа от мнимой оси.

Докажем справедливость этого представления. Можно показать, что для отрицательных значений w_n этот интеграл равен нулю. Это делается следующим образом. Возьмем интеграл вдоль мнимой оси, в котором путь интегрирования идет от $-ia$ к ia. Если этот контур замкнуть полуокружностью в левой полуплоскости, то интеграл будет равен нулю, так как подынтегральная функция — аналитическая внутри этого контура и на его границе. Таким образом, интеграл вдоль мнимой оси и малой полуокружности равен интегралу вдоль большой полуокружности в левой полуплоскости с обратным знаком.

Теперь нужно показать, что при $a \to \infty$ интеграл вдоль большой полуокружности стремится к нулю. Это доказывается тем, что интеграл действительно сходится в секторе с вершиной в начале координат, не захватывающем мнимую ось. Это можно доказать, перейдя к представлению z в полярных координатах, взяв абсолютные значения и показав, что интеграл сходится, если радиус a стремится к бесконечности. Следовательно, при $a \to \infty$ интеграл вдоль контура C стремится к нулю при $w_{n-1}+u_{n-1} < 0$. Так как левая часть нашего уравнения в этом случае равна единице [см. формулу (9.27а)], то мы должны иметь $\varepsilon_n = 1$.

При $w_{n-1} + u_{n-1} > 0$ применяется аналогичное рассуждение, с тем лишь отличием, что большая полуокружность берется в правой полуплоскости, и, следовательно, при больших значениях a этот интеграл имеет простой полюс в точке $z = s$. При помощи теории вычетов находим, что значение этого интеграла равно $-2\pi i e^{-sw_n}$, следовательно, $\varepsilon_n = 0$. Заметим, что замкнутый контур имеет направление по часовой стрелке, отсюда и знак минус. Можно показать, что в этом случае интеграл вдоль большой полуокружности в правой полуплоскости при $a \to \infty$ стремится к нулю; таким образом, значение приведенного выше интеграла вдоль контура C остается таким же, что и значение интеграла, определенное вдоль мнимой оси.

Осталось определить, что будет при $w_{n-1} + u_n = 0$. В этом случае интегрирование обычным способом выполнить невозможно. Однако можно рассмотреть замкнутый контур, как и в том случае, когда $w_{n-1} + u_n > 0$. Получаем вычет $-2\pi i$. В этом случае интеграл вдоль полуокружности не обращается в нуль вследствие отсутствия экспоненциального множителя. Интеграл, вычисленный вдоль окружности, будет равен $-2\pi i$, следовательно, его значение вдоль полуокружности равно $-\pi i$. Таким образом, $\varepsilon_n = \frac{1}{2}$. Читателю предлагается самостоятельно проверить эту часть доказательства.

Если в обеих частях приведенного выше выражения взять средние значения (мы можем поменять местами знаки криволинейного интеграла и средних значений) и затем заметить, что обычно вероятность того, что $w_{n-1} + u_{n-1} = 0$, равна нулю, то на основании определения ε_n имеем $E[\varepsilon_n] = E\,[P(w_n = 0)] = p_0$. Это обстоятельство не должно беспокоить читателя при выводе выражения для p_0.

Имеем важное выражение, которое обычно справедливо:

$$\gamma_n(s) = -\frac{1}{2\pi i}\oint\limits_C E\,[\exp(-zw_{n-1})\exp(-zu_{n-1})]\,\frac{dz}{z-s} + p_0. \qquad (9.41)$$

Если допускается, что t_{n-1}, s_{n-1} и w_{n-1} — независимые случайные величины (заметим, что они могут иметь и взаимно зависимые распределения, но и в этом случае это представление также остается справедливым), то их средние значения будут распределены в соответствии с экспоненциальными функциями этого интеграла, т. е.

$$E\left[e^{-zw_{n-1}}e^{-zu_{n-1}}\right] = E\left[e^{-zw_{n-1}}\right]E\left[e^{-zs_{n-1}}\right]E\left[e^{zt_{n-1}}\right]. \qquad (9.42)$$

Более того, если эти распределения одинаковы для всех требований, то первое выражение в правой части (9.42) равно $\gamma(z)$, а среднее выражение есть преобразование распределения времени обслуживания $\beta(z)$. Если промежутки времени между моментами поступления требований имеют экспоненциальное распределение с параметром λ, то его преобразование равно $\frac{\lambda}{z-\lambda}$, а преобразование $\gamma(s)$, которое теперь не зависит от n, имеет вид

$$\gamma(s)=\frac{\lambda}{2\pi i}\oint\limits_{C}\frac{-\gamma(z)\beta(z)}{(z-\lambda)(z-s)}dz+p_0. \tag{9.43}$$

Так как обе функции в числителе подынтегрального выражения регулярны в правой полуплоскости и можно показать, что они сходятся к нулю на полуокружности в правой полуплоскости с радиусом, стремящимся к бесконечности, то контур C может быть замкнут такой полуокружностью, включающей в себя два простых полюса в точках λ и s.

Вычет находится обычным способом. Имеем

$$\gamma(s)=-\frac{\lambda\gamma(\lambda)\beta(\lambda)}{\lambda-s}+\frac{\lambda\gamma(s)\beta(s)}{\lambda-s}+p_0. \tag{9.44}$$

После упрощения получим

$$\gamma(s)=\frac{\lambda\gamma(\lambda)\beta(\lambda)+p_0(s-\lambda)}{s-\lambda+\lambda\beta(s)}. \tag{9.45}$$

При $s\to 0$ оба выражения стремятся к единице. Если умножить обе части этого равенства на знаменатель, то при $s\to 0$ левая часть будет стремиться к нулю. Таким образом,

$$p_0=\gamma(\lambda)\beta(\lambda). \tag{9.46}$$

После подстановки этого значения p_0 правая часть (9.45) при $s=0$ становится неопределенной. Применив правило Лопиталя, получим

$$1=\lim_{s\to 0}\frac{p_0}{\lambda[1-\beta'(s)]}. \tag{9.47}$$

Но $\lim\limits_{s\to 0}\beta'(s)\equiv\frac{1}{\mu}$, где $\frac{1}{\mu}$ — среднее время ожидания. Поэтому

$$\gamma(s)=\frac{s\left(1-\frac{\lambda}{\mu}\right)}{s-\lambda+\lambda\beta(s)}, \tag{9.48}$$

а это выражение нам уже знакомо.

3. Иной способ решения

Обозначим снова через $W(t)$ время, в течение которого требование, поступившее в момент времени t, должно ожидать начала обслуживания. Допустим, что при $t\geqslant 0$ промежутки времени между требованиями и промежутки времени обслуживания описываются одной и той же функцией $K(t)$ [это нагрузка, поступающая в канал обслуживания за промежуток времени (u, t), и, следовательно, $W(t)$ может рассматриваться как работа, которая в момент времени t еще оставалась невыполненной]. Для этого случая Бенеш [53] получил

$$W(t) = K(t) - t + \int_0^t U[-W(u)]\,du, \quad t \geqslant 0, \tag{9.49}$$

где U — функция единичного скачка, т. е.

$$U(x) = \begin{cases} 1, & x \geqslant 0, \\ 0, & x < 0. \end{cases}$$

Заметим, что $W(0) = K(0)$.

При таком подходе к системам массового обслуживания не делается допущений о независимости распределения входящего потока и независимости распределения времени обслуживания и не рассматриваются частные случаи. Нет также необходимости обращаться к марковским процессам.

Заметим, что решение предыдущего уравнения выражается через функционал верхней грани, распределение которого найти очень трудно.

Можно показать, что с вероятностью, равной 1, имеем

$$\begin{gathered} \exp[-sW(t)] = \exp\{-s[K(t) - t]\} \times \\ \times \left(1 - s\int_0^t \exp\{s[K(u) - u]\} P(u, 0)\,du\right), \end{gathered} \tag{9.50}$$

где

$$P(u, 0) = \begin{cases} 1, & \text{если } W(u) = 0, \\ 0, & \text{если } W(u) \neq 0. \end{cases}$$

Кроме того, имеем

$$P[W(t) \leqslant w] = \begin{cases} 0 \text{ при } w < 0, \\ P[K(t) - t \leqslant w] - \dfrac{\partial}{\partial w}\displaystyle\int_0^t P[K(t) - K(u) - \\ -t + u \leqslant w \text{ и } W(u) = 0]\,du \text{ при } w \geqslant 0. \end{cases} \tag{9.51}$$

Читатель, очевидно, привык к такому положению, что всегда нужно устанавливать распределение входящего потока и распределение времени обслуживания. Однако при новом подходе требуется лишь определение функции нагрузки $K(t)$.

9.4. Многоканальная система, поток с ограниченным последействием, произвольно распределенные взаимно независимые длительности обслуживания

Здесь будет приведен только конечный результат. Читатель отсылается по этому вопросу к работам Кифера и Вольфовица. Авторы указывают, что имеет смысл рассмотреть эту задачу в некоторых частных случаях распределения входящего потока и распределения времени обслуживания, но отмечают, что решение, вероятно, окажется трудным. Уравнение Линдли является частным случаем полученного ими выражения. Для стационарного состояния имеем

$$P(w) = \int P_\infty \{\psi(w, b, c)\, dB(b)\, dA(c), \tag{9.52}$$

где P_∞ — показатель, соответствующий $P(w)$;

A — распределение длительности промежутков времени между моментами поступления требований;

B — распределение времени обслуживания.

Показано, что $P(w)$ является функцией распределения, когда загрузка системы ρ меньше единицы. Однако при $\rho \geqslant 1$ это выражение вообще не имеет смысла. Здесь ρ — отношение среднего времени обслуживания к числу каналов, умноженному на среднюю длительность промежутка времени между моментами поступления требований.

Задачи

1. Поток статей в редакцию журнала задан пуассоновским распределением с параметром λ. Время обслуживания имеет усеченное экспоненциальное распределение, т. е.

$$b(t) = \begin{cases} Ce^{-\mu t}, & t \leqslant T, \\ 0, & t > T, \end{cases}$$

где C — нормирующая постоянная, зависящая от T.

С помощью уравнения Линдли найдите выражение для среднего времени ожидания.

2. Используйте интегральное уравнение Линдли для нахождения распределения времени ожидания в одноканальной системе для следующих случаев:

а) пуассоновский входящий поток с параметром λ и экспоненциальное время обслуживания с параметром μ;

б) распределение входящего потока $\left[\frac{\lambda^n}{(n-1)!}\right] t^{n-1} e^{-\lambda t}$ и экспоненциальное время обслуживания с параметром μ. Найдите распределение времени ожидания:

$$w(t) = \mu \left(\frac{\lambda}{\lambda+\mu}\right)^n \left[\sum_{k=1}^{n} \frac{(\lambda+\mu)^{n-k}}{(n-k)!} \int_0^\infty w(t+y) \times \right.$$

$$\left. \times e^{-\lambda y} y^{n-k} dy + \int_0^\infty w(t-y) e^{-\mu y} dy \right].$$

Допуская, что решение имеет вид $w(t)=1-\alpha e^{-\beta t}$, покажите, что

$$\alpha=\frac{\mu-\beta}{\mu},\quad \left(\frac{\lambda}{\lambda+\beta}\right)^n=\frac{\mu-\beta}{\mu};$$

в) входящий поток регулярный с промежутками между моментами поступления требований, равными одной временно́й единице; распределение времени обслуживания $\frac{\mu^{n+1}t^n e^{-\mu t}}{n!}$;

$$w(t)=1+\sum_{k=1}^{n} c_k e^{x_k t}.$$

Определите c_k и x_k.

3. Используя статью Смита, получите доказательство его главной теоремы.

4. Используя статью Линдли, покажите, что искомая предельная функция является функцией распределения.

5. Пусть в одноканальной системе с пуассоновским входящим потоком с параметром λ и произвольным временем обслуживания с плотностью распределения $b(x)$ функция $w_n(x, t)$ есть плотность вероятности того, что в момент времени t в очереди находится n требований, не считая обслуживаемого требования, и что обслуживание этого требования продолжается уже в течение времени x. Пусть значение $w_n(x, 0)$ задано. Вероятность того, что обслуживание этого требования закончится в промежутке времени $(x, x+\Delta)$, равна $b(x)\Delta$, где

$$b(x)=\eta(x)\exp\left[-\int_0^x \eta(y)\,dy\right],$$

а $\eta(x)\Delta$ — вероятность того, что если обслуживание требования продолжалось в момент времени x, то оно закончится в промежутке времени $(x, x+\Delta)$;

$$w_n(x+\Delta, t+\Delta)=w_n(x, t)(1-\lambda\Delta)[1-\eta(x)\Delta]+w_{n-1}(x, t)\lambda\Delta.$$

Разложив левую часть этого выражения в ряд относительно (x, t) до Δ в первой степени и проведя упрощение, получим

$$\frac{\partial w_n}{\partial t}+\frac{\partial w_n}{\partial x}+[\lambda+\eta(x)]w_n=\lambda w_{n-1},\quad n=1, 2, \ldots,$$

$$\frac{\partial w_0}{\partial t}+\frac{\partial w_0}{\partial x}+[\lambda+\eta(x)]w_0=0.$$

Если $E(t)$ — вероятность того, что система свободна, можно аналогично найти уравнение

$$\frac{dE(t)}{dt}+\lambda E(t)=\int_0^\infty \eta(x)\,w_0(x, t)\,dt$$

при заданном начальном условии $E(0)$.

Покажите справедливость граничных условий

$$w_n(0, t)=\int_0^\infty w_{n+1}(x, t)\,\eta(x)\,dx,\quad n=1, 2, \ldots,$$

$$w_0(0, t)=\int_0^\infty w_1(x, t)\,\eta(x)\,dx+\lambda E(t).$$

Кроме того, имеем

$$w_n(x, 0) = \delta_{nN}\delta(x - x_0), \quad n = 0, 1, 2, \ldots,$$

где первая величина в правой части этого выражения — символ Кронекера, а вторая величина — дельта-функция.

Условие сохранения вероятности приводит к равенству

$$\frac{d}{dt}\left[E(t) + \sum_{n=0}^{\infty}\int_0^{\infty} w_n(x, t)\,dx\right] = 0.$$

Можно ввести производящую функцию $G(z, x, t)$ для $w_n(x, t)$ и обычным способом преобразовать систему уравнений. Введем обозначение

$$G(z, x, t) = H(z, x, t)\,\mathrm{e}^{-N(x)},$$

где $N(x) = \int_0^x \eta(y)\,dy$. Покажите, что уравнение

$$\frac{dH}{dt} + \frac{dH}{dx} + \lambda(1 - z)H = 0$$

имеет решение

$$H(z, x, t) = H_0(z, t - x)\,\mathrm{e}^{-\lambda(1-z)x},$$

откуда можно получить $G(z, x, t)$.

Используя выражение для $w_n(x, 0)$, можно получить

$$H_0(z - x) = \mathrm{e}^{N(x)}\mathrm{e}^{\lambda(1-z)x}z^N(x - x_0).$$

Покажите, что граничные условия объединены в выражении

$$zG(z, 0, t) = \int_0^{\infty}\eta(x)\,G(z, x, t)\,dx + \lambda zE(t) -$$

$$- \int_0^{\infty}\eta(x)\,G(0, x, t)\,dx.$$

Можно также показать, что справедливо следующее условие:

$$\frac{dE}{dt} + \lambda E + zH_0 = \int_0^{\infty} b(x)\,H_0(z, t - x)\,\mathrm{e}^{-\lambda(1-z)x}dx + \lambda zE.$$

Если в начальный момент времени $t=0$ система свободна, при этом $E(0)=1$ и $G(z, x, 0) = 0$, то в правой части этого уравнения можно записать интеграл в пределах от 0 до t. Преобразование Лапласа по переменной x с параметром s после упрощения имеет вид

$$H_0^*(z, s) = \frac{[s + \lambda(1 - z)]\,E^*(s) - 1}{b^*[s + \lambda(1 - z)] - z}.$$

Определяя соответствующий нуль знаменателя для $0 \leqslant z \leqslant 1$, который одновременно является и нулем числителя, можно вычислить обратное преобразование функции $E^*(s)$.

При $\frac{\lambda}{\eta} < 1$, где η — среднее время обслуживания, знаменатель имеет единственный нуль z_s. Это обеспечивает существование решения для равновесного состояния, преобразование которого имеет вид

$$E^*(s)=\frac{1}{s+\lambda(1-z_s)}.$$

Применяя тауберову теорему, из которой следует, что

$$\lim_{t\to\infty}E(t)=\lim_{s\to 0}sE^*(s)=1-\frac{\lambda}{\eta},$$

докажите равенство

$$\lim_{t\to\infty}H(z,\ t)=\frac{\lambda(1-s)\left(1-\frac{\lambda}{\eta}\right)}{b^*[\lambda(1-z)]-z}.$$

Вычислите $E^*(s)$ при $b(x)=\eta e^{-\eta x}$.

Покажите также, что при $E(0)=1$ и $w_n(x,\ 0)=0$ функция $E(t)$ удовлетворяет условию $\frac{dE}{dt}=-\lambda E+\left(\text{свертка функций } \lambda E \text{ и } \frac{dG(t)}{dt}\right)$, где $G(t)$ — распределение длительности периода занятости.

Глава 10

ВХОДЯЩИЙ ПОТОК С ОГРАНИЧЕННЫМ ПОСЛЕДЕЙСТВИЕМ; ВРЕМЯ ОБСЛУЖИВАНИЯ, РАСПРЕДЕЛЕННОЕ ПО ЭКСПОНЕНЦИАЛЬНОМУ ЗАКОНУ ИЛИ ПО ЗАКОНУ ЭРЛАНГА

10.1. Введение

В этой главе рассматривается число требований, находящихся в системе, и распределение длительности периода занятости. В систему поступает поток с ограниченным последействием, а время обслуживания имеет экспоненциальное или же эрланговское распределение. Здесь используется материал нескольких статей Конолли, посвященных данному вопросу. Математический аппарат этой главы достаточно сложен и потребует терпеливого изучения. Рассматриваются моменты поступления требований и вводится достаточное число случайных величин, с тем чтобы в результате получить марковский процесс.

Для случая экспоненциального распределения времени обслуживания вводится величина $P_n(u, t)$, которая используется для вычисления вероятности $P_n(t)$ того, что в момент времени t в системе находится n требований. Кроме того, используя преобразование Лапласа, а затем применяя к обратному преобразованию методы численного интегрирования, найдем $P_n(t)$. В частном случае результат соответствует решению задачи для системы с пуассоновским входящим потоком. Для системы с экспоненциальным временем обслуживания находится также распределение длительности периода занятости. Полученный результат справедлив также в случае регулярного входящего потока.

Кратко описывается система с эрланговским распределением времени обслуживания. Для нее находится распределение вероятностей совместного появления двух событий — данного числа обслуженных требований и определенной длительности периода занятости. Метод доказательства в основном аналогичен тому, который приводится для случая экспоненциального времени обслуживания.

10.2. Число требований, находящихся в очереди (система с экспоненциальным временем обслуживания)

Рассмотрим одноканальную систему, в которой промежутки времени между моментами поступления требований взаимно независимы и имеют произвольное распределение с плотностью вероятности, равной $a(t)$, а время обслуживания имеет экспоненциальное распределение с параметром μ. Обслуживание требований происходит в порядке поступления. Обозначим через $P_n(t)$ вероятность того, что в момент времени t в системе находится n требований (включая и обслуживаемое). Обозначим через $P_n(u, t)$ плотность вероятности того, что в момент времени t в системе находится n требований при условии, что последнее требование поступило в промежутке времени $(t-u-du,\ t-u)$, а через $P_n(0, t)$ — плотность вероятности того, что в промежутке времени $(t-dt, t)$ поступит требование, в результате чего система перейдет из состояния E_{n-1} в состояние E_n. Имеем $P_0(0, t)=0$. Введем также функцию $A(t)=\int_0^t a(x)dx$.

Вероятность того, что в промежутке времени u, в течение которого не поступает ни одного требования, будет закончено обслуживание n требований при условии, что в начале промежутка времени u число требований, находящихся в системе, было больше n, равна $b_n(u)=e^{-\mu u}\left[\frac{(\mu u)^n}{n!}\right]$. Если в начале промежутка времени u в системе находится N требований, то вероятность того, что все они будут обслужены за этот промежуток, равна

$$\overline{b}_N(u)=1-\sum_{j=0}^{N-1} b_j(u). \tag{10.1}$$

Допускаем, что первое требование поступает в систему в момент времени $t=0$, когда в системе нет других требований. Исследование для произвольного начального состояния производится аналогично.

Вероятность того, что в промежутке времени $(t-dt, t)$ поступит одно требование, равна $P(0, t)=\sum_{n=1}^{\infty} P_n(0, t)$. Преобразование Лапласа этого выражения имеет вид

$$P^*(0, s)=\sum_{n=1}^{\infty} P_n^*(0, s). \tag{10.2}$$

Заметим, что $a(t-0)$ — плотность вероятности того, что, начиная с момента времени $t=0$, следующее требование поступит в промежутке времени $(t-dt, t)$. Получив выражение для $P(0, t)$, покажем, что этот ряд сходится, по крайней мере, для $\mathrm{Re}(s)>0$. Затем получим выражение для $P_n(0, t)$.

Поступление требования в промежутке времени $(t-dt, t)$ может оказаться первым после поступления начального требования в момент времени $t=0$ с вероятностью $a(t-0)dt$, или последнее поступление, предшествующее данному, может произойти в промежутке времени $(t-u-du,\ t-u)$ с вероятностью $P(0,\ t-u)a(u-0)\,du\,dt$. Таким образом, имеем

$$P(0,\ t)=a(t)+\int_0^t P(0,\ t-u)\,a(u)\,du. \tag{10.3}$$

Преобразование Лапласа

$$P^*(0,\ s)=\frac{a^*(s)}{1-a^*(s)} \tag{10.4}$$

сходится, по крайней мере, для $\operatorname{Re}(s)>0$. Допускаем, что запись в виде $a(t-0)=a(t)$ не вызовет затруднений.

Кроме того, имеем

$$P_1(0,\ t)=a(t)\,b_1(t)=\int_0^t a(u)\sum_{m=1}^{\infty} P_m(0,\ t-u)\,b_m(u)\,du,$$

$$P_2(0,\ t)=a(t)\,b_0(t)+\int_0^t a(u)\sum_{m=0}^{\infty} P_{m+1}(0,\ t-u)\,b_m(u)\,du,$$

$$P_n(0,\ t)=\int_0^t a(u)\sum_{m=0}^{\infty} P_{n+m-1}(0,\ t-u)\,b_m(u)\,du,\quad n\geqslant 3. \tag{10.5}$$

Первое выражение в правой части первого уравнения есть плотность вероятности того, что требование, поступившее в промежутке времени $(t-dt,\ t)$, будет первым после момента времени $t=0$. Второе выражение есть плотность вероятности того, что требование, поступившее до этого момента, скажем, в промежутке времени $(t-u-du,\ t-u)$, приводит к состоянию E_m, при котором все требования будут обслужены до конца промежутка времени u. Вероятность этого события должна быть просуммирована по всем m и всем u, $0\leqslant u\leqslant t$. То же относится и к остальным уравнениям.

Возьмем преобразование Лапласа последнего уравнения. Получим однородное уравнение в конечных разностях с коэффициентами, не зависящими от n:

$$P_n^*(0,\ s)=\sum_{m=0}^{\infty}\lambda_m(s)\,P_{n+m-1}^*(0,\ s),\quad n\geqslant 3. \tag{10.6}$$

Здесь

$$\lambda_m(s)=\int_0^{\infty} e^{-su}a(u)\,b_m(u)\,du=\int_0^{\infty} e^{-u(s+\mu)}(\mu u)^m\,a(u)\,\frac{du}{m!}, \tag{10.7}$$

так как имеем свертку. Кроме того, $\lambda_m(s)$ является коэффициентом при x^m в разложении преобразования

$$\int_0^{\infty} e^{-u[s+(1-x)\mu]}\,a(u)\,du=a^*[s+(1-x)\mu] \tag{10.8}$$

по степеням x. В этом можно убедиться, разложив правую часть этого выражения в ряд и написав интеграл, преобразованием которого является коэффициент при x^m.

Подставляя $x^n = P_n^*(0, s)$ в уравнение в конечных разностях (10.6) и упрощая [чтобы получить общее решение, которое принимает вид

$$P_n^*(0, s) = \sum_{j=1}^{\infty} k_j \xi_j^n,$$

если все корни ξ_i уравнения (10.9) простые], имеем

$$a^*[s + (1 - x)\mu] = x. \tag{10.9}$$

Коэффициенты k_j не зависят от n и определяются из условия, что равенство должно выполняться для всех значений n. Даже в том случае, если все корни ξ_i простые, решение не существует, если абсолютное значение каждого из них не меньше единицы, так как значение $P^*(0, s)$ конечно.

Решая (10.9) относительно x, находим единственный нуль ξ с абсолютным значением, меньшим единицы, который мы используем в выражении

$$P_n^*(0, s) = k\xi^n. \tag{10.10}$$

Это решение справедливо при $n=2$.

Преобразование Лапласа первого интегро-дифференциального уравнения имеет вид

$$P_1^*(0, s) = k\xi - 1. \tag{10.11}$$

После подстановки в $P_1^*(0, s)$ имеем

$$\frac{a^*(s)}{1 - a^*(s)} = -1 + \frac{k\xi}{1 - \xi} \tag{10.12}$$

или

$$k = \frac{1 - \xi}{\xi[1 - a^*(s)]}. \tag{10.13}$$

Интегро-дифференциальные уравнения для $P_n(t)$ совпадают с приведенными выше уравнениями (10.5), отличаясь лишь тем, что в левой части этих уравнений $P_1(0, t)$ заменено на $P_0(t)$; $P_2(0, t)$ — на $P_1(t)$ и $P_n(0, t)$ — на $P_n(t)$ (это справедливо при $n \geqslant 2$), а в правой части $a(\cdot)$ заменено на $1 - A(\cdot) \equiv A_c(\cdot)$. Заметим, что теперь подынтегральными выражениями являются условные вероятности $P_0(u, t)$, $P_1(u, t)$ и $P_n(u, t)$. Таким образом,

$$P_n(u, t) = A_c(u) \sum_{m=0}^{\infty} P_{n+m}(0, t - u)\, b_m(u).$$

Взяв преобразование Лапласа, получим

$$P_n^*(s) = \frac{(1-\xi)\,\xi^{n-1}}{1-a^*(s)} A_c^*[s+(1-\xi)\mu], \quad n \geqslant 2. \tag{10.14}$$

Заметим, что

$$A_c^*(s) = \frac{1-a^*(s)}{s}. \tag{10.15}$$

Таким образом,

$$P_n^*(s) = \frac{(1-\xi^2)\,\xi^{n-1}}{\mu\,[1-a^*(s)]\left(\frac{s}{\mu}+1-\xi\right)}, \tag{10.16}$$

что справедливо при $n \geqslant 1$, и

$$P_0^*(s) = \frac{1}{s} - \frac{1-\xi}{\mu\,[1-a^*(s)]\left(\frac{s}{\mu}+1-\xi\right)}. \tag{10.17}$$

Полученные формулы удовлетворяют соотношению $\sum_{n=0}^{\infty} P_n^*(s) = \frac{1}{s}$. Они справедливы и для пуассоновского потока. Нуль ξ — тот же самый, что и $\mu\alpha_2$ — нуль, встречавшийся при анализе этого случая в гл. 4. Используя соотношения

$$\lim_{t\to\infty} P_n(t) = \lim_{s\to 0} sP_n^*(s), \quad a \equiv \int_0^{\infty} ua(u)\,du, \tag{10.18}$$

можно вывести некоторые свойства стационарного состояния. В окрестности точки $s=0$ имеем

$$a^*(s) = 1 - as + O(s^2), \tag{10.19}$$

$$\xi(s) = \xi_0 + \xi_1 s + O(s^2), \quad \xi_0 \equiv \xi(0), \quad \xi_1 \equiv \xi'(0). \tag{10.20}$$

При $a > \frac{1}{\mu}$ $\left(\text{заметим, что для пуассоновского потока } \lambda = \frac{1}{a}\right)$ имеем $\xi_0 < 1$, а при $a \leqslant \frac{1}{\mu}$ имеем $\xi_0 = 1$. Для первого случая получим следующий основной результат:

$$\lim_{s\to 0} sP_n^*(s) = \begin{cases} \frac{1}{\mu a}(1-\xi_0)\,\xi_0^{n-1}, & n \geqslant 1, \\ 1 - \frac{1}{\mu a}, & n = 0. \end{cases} \tag{10.21}$$

Следовательно, в первом случае $\left(a \geqslant \frac{1}{\mu}\right)$ выполняется условие статистического равновесия. При $a < \frac{1}{\mu}$ оба предела равны нулю. Наконец, при $a=\frac{1}{\mu}$ функция $1-\xi$ ведет себя как дробная степень s, т. е.

$$\xi(s)=1-cs^{\frac{1}{k}}, \tag{10.22}$$

где $k>1$, а c — постоянная; оба предела равны нулю.

10.3. Длительность периода занятости в системе с экспоненциальным временем обслуживания

Первая часть этого параграфа посвящена основной теории периода занятости, а во второй части определяется распределение длительности периода занятости.

Пусть $p_m(t)\,dt$ — вероятность того, что период занятости закончится в промежутке времени $(t,\ t+dt)$ и что за это время будет обслужено m требований. Тогда вероятность того, что за период занятости будет обслужено m требований, равна

$$p_m=\int_0^\infty p_m(t)\,dt, \tag{10.23}$$

а

$$p(t)\,dt=\sum_{m=1}^{\infty} p_m(t)\,dt \tag{10.24}$$

есть вероятность того, что период занятости закончится в промежутке времени $(t,\ t+dt)$, независимо от того, сколько требований будет обслужено. Если существует статистическое равновесие, выполняется соотношение

$$\int_0^\infty p(t)\,dt=1: \tag{10.25}$$

в противном случае этот интеграл меньше единицы.

Если $c_n(t)$ — плотность вероятности того, что промежуток времени между требованиями равен t, и за это время закончится обслуживание ровно n требований, а к концу этого промежутка на обслуживание поступит следующее требование, находящееся в очереди [при условии, что к началу этого промежутка времени имеется не меньше $n+1$ требований], то

$$c_n(t)=a(t)\,b_n(t). \tag{10.26}$$

По определению, $b_n(t)$ — вероятность того, что в промежутке времени $(0,\ t)$ заключено n последовательных длительностей обслуживания; при этом в момент времени t обслуживается $(n+1)$-е требование.

Допустим, что в промежутке времени $(t, t+dt)$ после начала периода занятости поступает новое требование, и общее число требований, ожидающих в очереди, возрастает до $n \geqslant 1$. Пусть на обслуживании находится m-е требование из числа тех, которые будут обслужены в этом периоде. Пусть далее плотность вероятности этого события равна $f_{mn}(t)$. При $n=1$ имеем

$$f_{m1}(t) = \int_0^t \sum_{k=1}^{m-1} f_{m-k,\, k}(s)\, c_k(t-s)\, ds, \quad m \geqslant 2, \tag{10.27а}$$

$$f_{11}(t) = c_0(t). \tag{10.27б}$$

Первое уравнение показывает, что если такое требование поступает в промежутке времени $(s, s+ds)$, то на обслуживании может находиться любое требование от первого до $(m-1)$-го. Если обслуживается $(m-k)$-е требование, то после поступления требования должна образоваться очередь, состоящая из k требований. Последнее требование из этой очереди в момент времени t должно находиться на обслуживании. Требование, которое поступит в момент времени t, будет единственным требованием, ожидающим начала обслуживания. Аналогичное уравнение можно написать и для $n \geqslant 2$:

$$f_{mn}(t) = \int_0^t \sum_{k=0}^{m=1} f_{m-k,\, n-1+k}(s)\, c_k(t-s)\, ds. \tag{10.28}$$

Заметим, что это выражение является сверткой. Возьмем преобразование Лапласа с параметром z. Затем введем производящую функцию $f_n^*(y, z)$ (которая сама зависит от n) для преобразования Лапласа функции $f_{mn}(t)$. В результате получим уравнение в конечных разностях, которое после подстановки в производящую функцию $f_n^*(y, z)$ значений x^n имеет вид $x = c^*(xy, z)$, где

$$c^*(y, z) = \sum_{k=0}^{\infty} y^k c_k^*(z) = \int_0^{\infty} a(t) \exp\{-t[z + (1-y)\mu]\}\, dt =$$
$$= a^*[z + (1-y)\mu]. \tag{10.29}$$

Можно показать, что уравнение относительно x имеет один корень ξ, при этом $|\xi| < 1$. При $n \geqslant 1$, $|y| \leqslant 1$ для $\mathrm{Re}(z) > 0$ при условии, что ξ — простой нуль, имеем

$$f_n^*(y, z) = y\xi^n(y, z). \tag{10.30}$$

Определим теперь искомое распределение $p_m(t)$. Допустим, что преобразование Лапласа производящей функции имеет вид

$$P^*(y, z) = \sum_{m=1}^{\infty} y^m p_m^*(z). \tag{10.31}$$

Пусть

$$A_c(s) = \int_0^\infty a(t)\,dt, \tag{10.32}$$

$$d_n(t) = A_c(t)\,b_n(t), \tag{10.33}$$

$$D^*(y, z) = \sum_{n=1}^\infty y^n d_n^*(z) = \int_0^\infty A_c(s) \exp\{-[z + (1-y)\mu]\,s\}\,ds = \\ = \frac{1 - c^*(y, z)}{\left(\frac{z}{\mu} + 1 - y\right)\mu}, \tag{10.34}$$

где

$$A_c^*(z) = \frac{1 - a^*(z)}{z}. \tag{10.35}$$

Вычислим теперь распределение длительности периода занятости. Вероятность того, что период занятости окончится в промежутке времени $t+dt$ после окончания обслуживания первого требования, равна вероятности совместного наступления двух событий: 1) в промежутке времени $(0, t)$ не поступит ни одного требования и 2) обслуживание первого требования закончится в промежутке времени $(t, t+dt)$. Таким образом,

$$p_1(t) = A_c(t)\,\mu e^{-\mu t}. \tag{10.36}$$

Если при поступлении требования до момента времени t на обслуживании находится $(m-k)$-е требование, то поступление данного требования приведет к образованию очереди, содержащей k требований $(1 \leqslant k \leqslant m-1)$. Все требования, включая и то, которое находится на обслуживании, должны быть обслужены в оставшемся промежутке времени, а последнее требование будет обслужено в промежутке времени $(t, t+dt)$, при этом больше ни одно требование не поступает. Получим

$$p_m(t) = \mu \int_0^t \sum_{k=1}^{m-1} f_{m-k,\,k}(s)\,d_k(t-s)\,ds. \tag{10.37}$$

Как и ранее, можно взять преобразование Лапласа и ввести производящую функцию для преобразованных уравнений. Имеем

$$P^*(y, z) = y d_0^*(z) + \mu \sum_{k=1}^\infty y^k d_k^*(z) f_k^*(y, z) = \\ = \mu y D^*(\xi y, z) = \frac{y(1-\xi)}{\frac{z}{\mu} + 1 - \xi y}. \tag{10.38}$$

Заметим, что $P^*(y, 0)$ — производящая функция распределения p_m, а $P^*(1, 0)$ — вероятность того, что длительность периода занятости конечна.

Среднее число требований, обслуженных за период занятости, равно

$$N = \left.\frac{\partial P^*(y, 0)}{\partial y}\right|_{y=1}.$$

Если $y=1$, то

$$P^*(1, z) = \frac{1-\xi}{\frac{z}{\mu}+1-\xi}, \tag{10.39}$$

где ξ — наименьший корень уравнения

$$x = c^*(x, z) = a^*[z+(1-x)\mu]. \tag{10.40}$$

Если ввести обозначение

$$\xi(z) = \sum_{n=0}^{\infty} k_n z^n, \tag{10.41}$$

где k_n получено из $\xi(n)$ путем дифференцирования, то

$$P^*(1, z) = \frac{1-k_0-k_1 z - O(z)}{\frac{z}{\mu}+1-k_0-k_1 z - O(z)}. \tag{10.42}$$

Если $\rho = \frac{\lambda}{\mu} < 1$, где λ — интенсивность поступления требований, то $k_0 < 1$ и $P(1, z) \to 1$ при $z \to 0$. Если $\rho > 1$, то $k_0 = 1$ и

$$P^*(1, z) \to \frac{k_1}{\left(k_1 - \frac{1}{\mu}\right)} = \frac{1}{\rho} \quad \text{при } z \to 0.$$

Если $\rho = 1$, то $\xi = 1$ — двойной корень при $z=0$ и $P^*(1, z) \to 1$.

Используя эти положения, можно аналогично показать, что

$$N = \begin{cases} \dfrac{1}{1-h_0}, & \rho < 1, \\ \dfrac{(h_2-h_1)\rho}{h_1^2}, & \rho > 1, \end{cases} \tag{10.43}$$

где h_0, h_1 и h_2 — коэффициенты в разложении наименьшего корня ξ уравнения $x = a^*[(1-xy)\mu]$ в ряд $(1-y)^n$.

Для регулярного входящего потока с промежутками времени между требованиями, равными a, имеем

$$a^*(z) = e^{-az}, \tag{10.44}$$

$$\xi = \exp\left(-\frac{1-\xi y}{\rho}\right) = \frac{\rho}{y} \sum_{n=1}^{\infty} \frac{\left(\dfrac{ny}{e^{\frac{1}{\rho}}}\right)^n}{n(n!)}. \tag{10.45}$$

Для всех значений ρ Конолли получил [149]

$$p_1 = 1 - e^{-\frac{1}{\rho}},$$

$$p_2 = e^{-\frac{1}{\rho}} - e^{-\frac{2}{\rho}}\left(1 + \frac{1}{\rho}\right), \qquad (10.46)$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$h_1 = -\frac{1}{\rho - 1}, \quad h_2 = \frac{3\rho - 2}{2(\rho - 1)^3} \text{ и т. д.}$$

Кроме того, он получил таблицу значений p_m для первых пяти m и таблицу значений N для различных значений ρ — как для регулярного, так и для пуассоновского входящего потока.

10.4. Период занятости в системе с эрланговским распределением времени обслуживания

Этот случай также рассмотрен в статье Конолли, в которой главное внимание уделено вычислению вероятности и плотности вероятности совместного появления двух событий: 1) в течение периода занятости будет обслужено ровно m групп по k требований в каждой (k — параметр распределения Эрланга) и 2) весь процесс займет время, равное t. Рассматривается также одномерное распределение числа групп m (полученное путем интегрирования плотности совместного распределения по t) и одномерное распределение t (полученное путем суммирования по m). Эти положения применимы к известным частным случаям, уже рассмотренным в книге. Вспомогательные функции, которые Конолли использует в этом случае, содержат дополнительную информацию относительно элемента группы и относительно группы, обслуживание которой производится в рассматриваемом промежутке времени.

Пусть $p_m(t)$ — вероятность совместного появления таких двух событий. Возьмем преобразование Лапласа с параметром z по переменной t, а затем рассмотрим производящую функцию для преобразования Лапласа $p_m^*(z)$, используя в качестве общего члена переменную y, возведенную в степень mk (заметим, что m — число групп; умножив его на k, найдем число одиночных требований). Данная производящая функция имеет вид

$$\frac{y\sum_j\left(1 - \xi_j^k\right)}{(bz + 1 - y\xi_j)\,H'(\xi_j)}. \qquad (10.47)$$

Это выражение приводится только для того, чтобы вызвать у читателя интерес к результату исследования. Лучше всего прочитайте сами эту статью и найдите основные характеристики.

Часть IV

РАЗЛИЧНЫЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ. ПРИЛОЖЕНИЯ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ. ТЕОРИЯ ВОССТАНОВЛЕНИЯ

В этой части содержится пять глав, которые охватывают большой круг вопросов. В гл. 11 изучаются системы массового обслуживания с различными дисциплинами очереди, в отличие от того, что в предыдущих частях рассматривались системы только с одной дисциплиной очереди («первым пришел — первым обслужен»). В гл. 12 делается попытка охватить разветвленные системы массового обслуживания, но рассматриваются только элементарные формы соединения. В гл. 13 изучаются различные ситуации массового обслуживания, описанные в гл. 1. В гл. 14 содержится ряд важных приложений теории массового обслуживания к различным областям практической деятельности. Многие процессы массового обслуживания могут рассматриваться как процессы восстановления. За последние годы теории восстановления уделялось большое внимание и были получены важные результаты. Вследствие того, что прикладное значение теории восстановления возросло, некоторые из ее результатов включены в гл. 15.

Глава 11

РАЗЛИЧНЫЕ ВИДЫ ДИСЦИПЛИНЫ ОЧЕРЕДИ

11.1. Введение

До сих пор рассматривались системы, в которых обслуживание требований производилось по принципу «первым пришел — первым обслужен». Очевидно, что этот вид дисциплины очереди не является единственно возможным. Например, выбор требований на обслуживание может производиться на основании приоритета. Так, если приоритет требования высок, то независимо от его положения в очереди такое требование поступает на обслуживание раньше требований с более низкими приоритетами. Примерами этого могут служить телеграф, где срочные отправления проходят первыми, или отделение неотложной помощи в больнице, где высокий приоритет дается случаям, требующим немедленного вмешательства.

Кроме того, имеются системы, в которых порядок поступления требований не играет роли, а выбор требования на обслуживание производится случайным образом. Примером такой системы может служить телефонный коммутатор простейшего типа, в котором оператор случайным образом выбирает абонента из многих, посылающих вызовы, и дает ему требуемое соединение.

Четвертым типом дисциплины очереди является такой, при котором поступающие требования распределяются по фиксированным каналам. Существует еще одна дисциплина очереди — «последним прибыл — первым обслужен», когда поступающие требования (например, предметы массового производства) складываются друг на друга и предмет, изготовленный последним, оказывается находящимся сверху и далее используется первым.

В каждом из этих случаев главной задачей является вычисление распределения времени ожидания и его среднего значения. Последнюю величину можно легко определить, даже не находя распределения времени ожидания.

11.2. Приоритеты

Наличие приоритетов у ожидающих требований имеет большое значение. Требования с различными приоритетами образуют потоки с одинаковыми или различными распределениями. В пределах данного приоритета требования поступают на обслуживание в один или большее число каналов по принципу «первым пришел — первым обслужен». При поступлении в систему требования с более высоким приоритетом обслуживаемое требование, имеющее низший приоритет, может возвращаться в очередь (приоритет, прерывающий обслуживание) или обслуживание его может завершаться (приоритет без прерывания обслуживания).

При решении задачи для системы с приоритетами возникают важные вопросы, связанные со стоимостью обслуживания, которые следовало бы включать в модель, хотя до сих пор это делалось редко.

В таких задачах учет стоимости может оказаться полезным при сравнении потерь в системе, где обслуживание требования с низшим приоритетом не прерывается, с общими или частичными потерями в системе, где обслуживание требования с низшим приоритетом прерывается и это требование возвращается в очередь. Возвращение требования на обслуживание при наличии такого приоритета может означать возобновление процесса обслуживания сначала. Может оказаться необходимым минимизировать общую стоимость времени ожидания, стоимость времени простоя обслуживающего устройства и стоимость прерванного обслуживания.

1. Приоритет без прерывания обслуживания

Одноканальная система [132]. Пусть требования с r-м приоритетом (чем меньше номер, тем выше приоритет) поступают в одноканальную систему в соответствии с пуассоновскими процессами с параметрами λ_k $(k=1, \ldots, r)$, и, следовательно, общий поток является пуассоновским с параметром $\lambda=\sum_{k=1}^{r}\lambda_k$. Допустим, что в *пределах данного приоритета* требования обслуживаются по принципу «первым пришел — первым обслужен». Общая функция распределения времени обслуживания имеет вид $F(t)=\left(\frac{1}{\lambda}\right)\times$ $\times\sum_{k=1}^{r}\lambda_k F_k(t)$, где $F_k(t)$ — функция распределения времени обслуживания требований с k-м приоритетом при среднем времени обслуживания, равном $\frac{1}{\mu_k}$.

Самый высокий приоритет имеет индекс единица, самый низкий — индекс r. Требования с высшим приоритетом обслуживаются раньше требований с низшим приоритетом независимо от времени их поступления. Однако обслуживание требования, имеющего любой приоритет, завершается прежде, чем допускается следующее требование, т. е. прерывания обслуживания не происходит.

Пусть

$$\rho_k=\frac{\lambda_k}{\mu_k}, \quad 1\leqslant k\leqslant r, \tag{11.1}$$

и

$$\sigma_k=\sum_{i=1}^{k}\rho_i,\quad \sigma_0=0,\quad 1\leqslant k\leqslant r. \tag{11.2}$$

Допустим, что в момент времени t_0 в систему поступает требование с p-м приоритетом и в момент времени t_1 оно направляется на обслуживание. В этом случае его время ожидания равно $T=t_1-t_0$. Пусть в момент времени t_0 на обслуживании нет ни одного требования или находится одно требование, и пусть в очереди впереди поступившего требования имеется $n_k (k=1,\ldots, p)$ требований с приоритетами не ниже k-го. Пусть T_0 — время, необходимое для окончания обслуживания требования, уже находящегося на обслуживании, а T_k — время, необходимое для обслуживания n_k требований $(k=1,\ldots,p)$. За время ожидания T в систему поступит n_k $(k=1,\ldots,p-1)$ требований с приоритетами не ниже k-го, которые будут обслужены раньше этого требования. Пусть T_k' — время обслуживания n_k' требований (т. е. всех требований с k-м приоритетом, которые поступают за время T). Заметим, что

$$T\geqslant\sum_{k=1}^{p-1}T_k'+\sum_{k=1}^{p}T_k+T_0.$$

В действительности должно иметь место равенство, так как в противном случае канал будет свободен, когда требование ожидает в очереди. Возьмем средние значения и поставим знак равенства. Получим

$$E[T]=\sum_{k=1}^{p-1}E[T_k']+\sum_{k=1}^{p}E[T_k]+E[T_0]. \tag{11.3}$$

По определению $E[T]\equiv W_q$ — время ожидания требования с p-м приоритетом.

Здесь $E[T_k]$ — среднее значение времени обслуживания требования с k-м приоритетом, умноженное на n_k, т. е.

$$E[T_k]=\frac{1}{\mu_k}\lambda_k W_k, \tag{11.4}$$

где $\frac{1}{\mu_k}$ — математическое ожидание времени обслуживания требования с k-м приоритетом;

$E[n_k]=\lambda_k W_k=\left(\frac{1}{\lambda_k}\right)^{-1}W_k$ — среднее число требований, умноженное на среднее время ожидания требований с k-м приоритетом.

Аналогично находим

$$E[T_k']=\frac{1}{\mu_k}\lambda_k W_p. \tag{11.5}$$

Здесь употреблено обозначение W_p, так как среднее значение n'_k равно среднему числу поступивших требований, умноженному на среднее время, в течение которого поступили эти требования, т. е. на время T, в течение которого рассматриваемое требование с p-м приоритетом должно ожидать. Затем получим

$$W_p=\frac{\sum_{k=1}^{p}\rho_k W_k+E[T_0]}{1-\sigma_{p-1}}. \tag{11.6}$$

С помощью метода математической индукции находим

$$W_p=\frac{E[T_0]}{(1-\sigma_{p-1})(1-\sigma_p)}. \tag{11.7}$$

Для выполнения равенства необходимо, чтобы $\sigma_p<1$.

Из формулы Полячека—Хинчина имеем

$$W=\frac{L-\frac{\lambda}{\mu}}{\lambda}=\frac{\lambda\int_0^\infty t^2dF(t)}{\left(\frac{2}{\mu}\right)(\mu-\lambda)}. \tag{11.8}$$

При $p=1$ имеем классический случай одноканальной системы без приоритета с дисциплиной очереди «первым пришел — первым обслужен». Находим выражение

$$W=W_1=\frac{E[T_0]}{1-\rho}, \tag{11.9}$$

которое должно совпадать с полученным ранее выражением (11.8). Следовательно, постоянная $E[T_0]$ определена. Она имеет вид

$$E[T_0]=\frac{\lambda}{2}\int_0^\infty t^2dF(t). \tag{11.10}$$

Таким образом, мы получили выражения для времени ожидания требований с различными приоритетами.

Можно также получить $E[T_0]$ путем вероятностных рассуждений и, положив $p=1$, найти среднее время ожидания в одноканальной системе. Но данный метод является более строгим.

Многоканальная система. Допустим, что тип входящих потоков тот же, что и в предыдущем случае и что время обслуживания в каждом из c каналов имеет одинаковое распределение с параметром μ. Исследование ведется, как и ранее, с той лишь особенностью, что когда все каналы заняты, обслуженные требования покидают систему случайным образом с интенсивностью μ. Имеем

$$W_p=\frac{E[T_0]}{\left[1-\left(\frac{1}{c\mu}\right)\sum_{k=1}^{p-1}\lambda_k\right]\left[1-\left(\frac{1}{c\mu}\right)\sum_{k=1}^{p}\lambda_k\right]}. \tag{11.11}$$

И в этом случае $E[T_0]$ можно найти с помощью формулы для среднего времени ожидания в многоканальной системе, находящейся в стационарном состоянии, при обслуживании требований в порядке поступления. Получим

$$E\,[T_0] = \frac{\dfrac{(c\rho)^c}{c\mu}}{c!\,(1-\rho)\sum\limits_{j=0}^{c-1}\left[\dfrac{(c\rho)^j}{j!}\right] + \sum\limits_{j=c}^{\infty}\left(\dfrac{c^c\rho^j}{c!}\right)}, \tag{11.12}$$

где $\rho = \dfrac{\lambda}{c\mu}$.

Необходимо, чтобы выполнялось условие $\left(\dfrac{1}{c\mu}\right)\sum\limits_{k=1}^{p}\lambda_k < 1$. Тогда выражение для W_p справедливо также при $\rho > 1$.

Если система с приоритетом загружена неполностью, то разница в среднем времени ожидания для требований с различными приоритетами мала, но она становится заметной, когда нагрузка на систему возрастает. Например, увеличение числа требований с наивысшим приоритетом, поступающих в одноканальную систему, приводит к большему времени ожидания как для требований с этим приоритетом, так и для остальных требований. Поэтому важно контролировать назначение высоких приоритетов и по возможности уменьшать время обслуживания.

Обобщение результатов для одноканальной системы. Кестен и Ранненберг [440] развили дальше эти положения. Мы приведем некоторые из полученных ими результатов. Пусть

$$G_k(t) = \begin{cases} 0 & \text{при } t < 0, \\ 1 - e^{-\lambda_k t} & \text{при } t \geqslant 0 \text{ и } \lambda_k > 0 \end{cases} \tag{11.13}$$

— функция распределения длительности промежутков времени между моментами поступления требований с k-м приоритетом при общем числе приоритетов r. Пусть $F_k(t)$ — функция распределения времени обслуживания требований с этим приоритетом в одноканальной системе массового обслуживания. Допустим, что для каждого приоритета существует взаимная независимость промежутков времени между моментами поступления требований, взаимная независимость моментов времени поступления требований и взаимная независимость длительностей обслуживания. Пусть

$$\lambda = \sum_{i=1}^{r}\lambda_i, \quad \mu_k^{(l)} = \int_{-0}^{\infty} t^l dF_k(t), \tag{11.14}$$

$$\varphi_k(\alpha) = \int_{-0}^{\infty} e^{-\alpha t} Fd_k(t), \quad \psi_k(\alpha) = \int_{-0}^{\infty} e^{-\alpha t} dH_k(t), \tag{11.15}$$

где $H_k(t)$ — функция распределения времени ожидания требований с k-м приоритетом.

Заметим, что если функция $\psi_k(\alpha)$ определена, то с помощью обратных преобразований можно определить $H_k(t)$.

При отсутствии насыщения

$$\sum_{i=1}^{r} \lambda_i \mu_i^{(1)} < 1,$$

$$\psi_k \alpha =$$

$$= \frac{\left[1 - \sum_1^r \lambda_i \mu_i^{(1)}\right]\left(-\sum_1^{k-1} \lambda_i + z_k^* - \alpha\right) - \sum_{k-1}^{r} \lambda_i \left[1 - \varphi_i\left(\sum_1^{k-1} \lambda_j - z_k^* + \alpha\right)\right]}{\lambda_k - \alpha - \lambda_k \varphi_k \left(\sum_1^{k-1} \lambda_j - z_k^* + \alpha\right)}, \quad (11.16)$$

где $z_1^* = 0$, а $z_k^* = z_k^*(\alpha)$ удовлетворяет уравнению

$$z_k^* = \sum_1^{k-1} \lambda_i \varphi_i \left(\sum_1^{k-1} \lambda_j - z_k^* + \alpha\right) = 0, \quad k \geqslant 2. \quad (11.17)$$

Так, например,

$$\psi_1(\alpha) = -\frac{\left[1 - \sum_1^r \lambda_i \mu_i^{(1)}\right] \alpha + \sum_0^{\tau} \lambda_i \left[1 - \varphi_i(\alpha)\right]}{\lambda_1 - \alpha - \lambda_1 \varphi_1(\alpha)}. \quad (11.18)$$

Выражение для среднего времени ожидания имеет вид

$$W_k \equiv E(w_k) = \frac{\sum_1^r \lambda_i \mu_i^{(2)}}{2\left[1 - \sum_1^{k-1} \lambda_i \mu_i^{(1)}\right]\left[1 - \sum_1^{k} \lambda_i \mu_i^{(1)}\right]}. \quad (11.19)$$

Второй момент равен

$$E(w_k^2) = \frac{\sum_1^r \lambda_i \mu_i^{(3)}}{3\left[1 - \sum_1^{k-1} \lambda_i \mu_i^{(1)}\right]^2 \left[1 - \sum_1^{k} \lambda_i \mu_i^{(1)}\right]} +$$

$$+ \frac{\sum_1^r \lambda_i \mu_i^{(2)} \sum_1^{k-1} \lambda_j \mu_j^{(2)}}{2\left[1 - \sum_1^{k-1} \lambda_i \mu_i^{(1)}\right]^2 \left[1 - \sum_1^{k} \lambda_i \mu_i^{(1)}\right]^2} +$$

$$+ \frac{\sum_1^r \lambda_j \mu_i^{(2)} \sum_1^{k-1} \lambda_j \mu_j^{(2)}}{2\left[1 - \sum_1^{k-1} \lambda_i \mu_i^{(1)}\right]\left[1 - \sum_1^{k} \lambda_i \mu_i^{(1)}\right]}, \quad k = 1, \ldots, r. \quad (11.20)$$

В случае насыщения $\sum_{1}^{r}\lambda_i\mu_i^{(1)} \geqslant 1$, и можно найти такое целое число $s\,(0 \leqslant s \leqslant r)$, что $\sum_{i=1}^{s}\lambda_i\mu_i^{(1)} < 1$, $\sum_{1}^{s+1}\lambda_i\mu_i^{(1)} \geqslant 1$;

$$W_k \equiv E(w_k) = \frac{\sum_{i=1}^{s}\lambda_i\mu_i^{(2)} + \left(\frac{\mu_{s+1}^{(2)}}{\mu_{s+1}^{(2)}}\right)\left[1-\sum_{1}^{s}\lambda_i\mu_i^{(1)}\right]}{2\left[1-\sum_{1}^{k-1}\lambda_i\mu_i^{(1)}\right]\left[1-\sum_{1}^{k}\lambda_i\mu_i^{(1)}\right]}, \qquad (11.21)$$

$$k = 1,\ 2, \ldots,\ s,$$

$$W_k \equiv E(w_k) = \infty, \quad k = s+1, \ldots,\ r.$$

Непрерывное множество приоритетов. Фиппс [644] распространил результаты, полученные для одноканальной системы массового обслуживания с приоритетами, на случай выхода станков из строя. Число имеющихся станков принимается равным бесконечности. Приоритет назначается в зависимости от времени, необходимого для обслуживания станка, — чем меньше время обслуживания, тем выше приоритет. Так как длительность времени обслуживания может соответствовать любому действительному числу, то и приоритет также может принимать любое значение, т. е. имеется бесконечное число приоритетов.

Допустим, что поступление новых заявок на ремонт станков подчиняется закону Пуассона и что среднее число заявок, поступающих в единицу времени, равно λ. Допустим также, что интенсивность поступления заявок с приоритетом t (где t — время, необходимое для обслуживания каждого из этих станков) задана величиной λ_t. Пусть $F(t)$ — функция распределения среднего времени обслуживания станков, имеющих любой приоритет. Тогда

$$\lambda_t dt = \lambda dF(t). \qquad (11.22)$$

По аналогии со случаем дискретных приоритетов среднее время ожидания W_t для станка, время ремонта которого равно t, определяется выражением

$$W_t = \frac{W_0}{\left[1-\lambda\int_0^t s\,dF(s)\right]^2}, \quad t \leqslant \tau, \qquad (11.23)$$

$$W_t = \infty, \quad t > \tau,$$

где

$$W_0 = \frac{\lambda}{2}\int_0^{\tau} t^2 dF(t), \qquad (11.24)$$

а τ — критическая длительность обслуживания, аналогичная s в (11.21).

Величина τ определяется из равенства

$$\lambda \int_0^{\tau} t dF(t) = 1. \tag{11.25}$$

Если длительность обслуживания не меньше τ, то наступает насыщение, и, следовательно, очередь растет до бесконечности.

Среднее число станков, ожидающих ремонта, равно

$$L_q = \lambda W_0 \int_0^{\tau} \frac{dF(t)}{\left[1 - \lambda \int_0^t s dF(s)\right]^2}. \tag{11.26}$$

Если, например, $F(t) = 1 - e^{-\mu t}$, что соответствует экспоненциальному времени ремонта, то при $\lambda < \mu$ время τ равно бесконечности и

$$W_0 = \frac{\lambda}{\mu^2} \text{ (случай отсутствия насыщения).} \tag{11.27}$$

Среднее время ожидания равно

$$W_t = \frac{W_0}{\left\{1 - \frac{\lambda}{\mu}\left[1 - e^{-\mu t}(1 + \mu t)\right]\right\}^2}, \tag{11.28}$$

и среднее число станков, ожидающих ремонта, равно

$$L_q = \frac{\lambda^2}{\mu} \int_0^{\infty} \frac{e^{-\mu t} dt}{\left\{1 - \frac{\lambda}{\mu}\left[1 - e^{-\mu t}(1 + \mu t)\right]\right\}^2}. \tag{11.29}$$

При насыщении L_q равно бесконечности.

2. Приоритеты, прерывающие обслуживание

В случае дисциплины очереди с приоритетом, прерывающим обслуживание, при поступлении требования с высшим приоритетом находящееся на обслуживании требование с низшим приоритетом возвращается в очередь, в которой ожидают требования с низшим приоритетом. Имеется два вида приоритетов с прерыванием обслуживания:

1) прерывание с повторением: при возвращении требования в обслуживающее устройство время, затраченное на обслуживание, теряется (здесь рассматривается только этот случай);

2) прерывание с продолжением прерванного обслуживания. Обслуживание требования возобновляется в той точке, где оно было прервано.

Джейсуол исследовал задачу для двух приоритетов, прерывающих обслуживание, в системе с пуассоновским входящим потоком и экспоненциальным распределением времени обслуживания при повторении прерванного обслуживания. Он применил метод, изложенный в гл. 9 (задача 5). Джейсуол представил распределения времени обслуживания так, как показано в гл. 9, и использовал их математические ожидания совместно с интенсивностью поступления требований в системе дифференциально-разностных уравнений для нахождения вероятности того, что в очереди находится данное число требований каждого приоритета при условии, что обслуживание требования, первого в очереди, составленной из требований с низшими приоритетами, было прервано после того, как обслуживание уже продолжалось в течение времени y, а длительность обслуживания требования равна x. Необходимо также рассмотреть выражения для этих вероятностей в том случае, когда обслуживание требования с низшим приоритетом не прерывается. Используются также методы, основанные на определении производящих функций и преобразований Лапласа, как в случае повторения прерванного обслуживания, рассматриваемого ниже.

Рассмотрим приоритеты, прерывающие обслуживание с повторением. В пределах каждого приоритета поступающие требования образуют пуассоновский поток и обслуживаются в порядке поступления. Для требований любого приоритета время обслуживания в одном канале имеет экспоненциальное распределение.

Когда поступает требование с высшим приоритетом, то требование с низшим приоритетом возвращается в начало очереди, состоящей из требований с таким же приоритетом, и ожидает обслуживания.

Ниже будут приведены дифференциально-разностные уравнения для системы с тремя приоритетами, прерывающими обслуживание. Чтобы получить решение для системы с двумя приоритетами, достаточно положить λ_3 и μ_3 равными нулю, опустить третий индекс и отбросить лишние уравнения.

Допустим, что имеется:

n требований с высшим приоритетом с интенсивностью поступления λ_1 и интенсивностью обслуживания μ_1;

m требований со средним приоритетом с интенсивностью поступления λ_2 и интенсивностью обслуживания μ_2;

k требований с низшим приоритетом с интенсивностью поступления λ_3 и интенсивностью обслуживания μ_3.

Пусть $P_{nmk}(t)$ — вероятность того, что в момент времени t в системе находится n требований с первым приоритетом, m требований со вторым приоритетом и k требований с третьим приоритетом. Имеем:

$$P'_{nmk}(t) = -(\lambda_1 + \lambda_2 + \lambda_3 + \mu_1) P_{nmk}(t) +$$
$$+ \lambda_1 P_{n-1,\, mk}(t) \lambda_2 P_{n,\, m-1,\, k}(t) + \lambda_3 P_{nm,\, k-1}(t) + \mu_1 P_{n+1,\, mk}(t),$$
$$P'_{000}(t) = -(\lambda_1 + \lambda_2 + \lambda_3) P_{000}(t) + \mu_1 P_{100}(t) +$$
$$+ \mu_2 P_{010}(t) + \mu_3 P_{001}(t),$$
$$P'_{n00}(t) = -(\lambda_1 + \lambda_2 + \lambda_3 + \mu_1) P_{n00}(t) + \lambda_1 P_{n-1,\, 00}(t) +$$
$$+ \mu_1 P_{n+1,\, 00}(t), \tag{11.30}$$

$$P'_{0m0}(t) = -(\lambda_1+\lambda_2+\lambda_3+\mu_2)P_{0m0}(t)+\mu_1 P_{1m0}(t)+$$
$$+\lambda_2 P_{0,\,m-1,\,0}(t)+\mu_2 P_{0,\,m+1,\,0}(t),$$
$$P'_{00k}(t) = -(\lambda_1+\lambda_2+\lambda_3+\mu_3)P_{00k}(t)+\mu_1 P_{10k}(t)+$$
$$+\mu_2 P_{01k}(t)+\lambda_3 P_{00,\,k-1}(t)+\mu_3 P_{00,\,k+1}(t),$$
$$P'_{nm0}(t) = -(\lambda_1+\lambda_2+\lambda_3+\mu_1)P_{nm0}(t)+\lambda_1 P_{n-1,\,m0}(t)+$$
$$+\lambda_2 P_{n,\,m-1,\,0}(t)+\mu_1 P_{n+1,\,m0}(t),$$
$$P'_{n0k}(t) = -(\lambda_1+\lambda_2+\lambda_3+\mu_1)P_{n0k}(t)+\lambda_1 P_{n-1,\,0k}(t)+$$
$$+\lambda_3 P_{n0,\,k-1}(t)+\mu_1 P_{n+1,\,0k}(t),$$
$$P'_{0mk}(t) = -(\lambda_1+\lambda_2+\lambda_3+\mu_2)P_{0mk}(t)+\mu_1 P_{1mk}(t)+$$
$$+\lambda_2 P_{0,\,m-1,\,k}(t)+\mu_2 P_{0,\,m+1,\,k}(t)+\lambda_3 P_{0m,\,k-1}(t).$$

В общем случае имеется $2p$ уравнений, описывающих систему с приоритетами, прерывающими обслуживание.

Заметим, что если требования с высшим приоритетом ожидают, то требование с низшим приоритетом не может поступить на обслуживание.

Упражнение. Напишите уравнения для системы с двумя приоритетами: n и m. Они нам потребуются позднее.

Хиткоут [340] рассмотрел такую задачу для системы с двумя приоритетами, в которой в момент времени $t=0$ находится начальное число $m_0=h$ требований со вторым приоритетом и $n_0=0$ требований с первым приоритетом. После преобразования по Лапласу $g_{nm}(s)$ четырех уравнений, описывающих систему с двумя приоритетами и содержащих вероятности $P_{nm}(t)$, вводится производящая функция, определенная для $|z|<1$ и $|x|<1$,

$$F(z,\ x,\ s)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} g_{nm}(s)\, z^n x^m. \tag{11.31}$$

Используется также частичная производящая функция

$$H_n(x,\ s)=\sum_{m=0}^{\infty} g_{nm}(s)\, x^m. \tag{11.32}$$

Чтобы найти выражения

$$\mu_1 H_1-\left(s+\mu_2+\lambda_1+\lambda_2-\lambda_2 x-\frac{\mu_2}{x}\right)H_0=$$
$$=g_{00}\mu_2\left(\frac{1}{x}-1\right)-x^h \tag{11.33}$$

и

$$\mu_1 H_{n+1}-(s+\mu_1+\lambda_1+\lambda_2-\lambda_2 x)H_n+\lambda_1 H_{n-1}=0,$$

производится умножение на x^m и суммирование по m. Затем, умножая на z^n и суммируя по n, находим

$$F(z, x, s)=$$

$$=\frac{g_{00}\mu_2\left[\left(\frac{1}{x}\right)-1\right]-x^h+\left\{\mu_1\left[\left(\frac{1}{z}\right)-1\right]-\mu_2\left[\left(\frac{1}{x}\right)-1\right]\right\}H_0(x, s)}{\lambda_1 z-(s+\mu_1+\lambda_1+\lambda_2-\lambda_2 x)+\frac{\mu_1}{z}}. \qquad (11.34)$$

Здесь $H_0(x, s)$ — функция, полученная при решении системы уравнений для H_n. Имеем

$$H_0(x, s)=\frac{g_{00}\mu_2\left[\left(\frac{1}{x}\right)-1\right]-x^h}{\mu_1 v_1-\left(s+\mu_2+\lambda_1+\lambda_2-\lambda_2 x-\frac{\mu_2}{x}\right)}. \qquad (11.35)$$

Если $v_1(x, s)$ — величина, обратная нулю знаменателя (11.34), лежащего вне единичного круга $|z|=1$, то можно написать

$$F(z, x, s)=$$

$$=\frac{[\mu_2(1-x)g_{00}(s)-x^{h+1}][1-v_1(x, s)]}{\{\mu_2(1-x)[1-\rho_2 x-v_1(x, s)]-sx\}[1-zv_1(x, s)]}, \qquad (11.36)$$

где $\rho_i=\frac{\lambda_i}{\mu}$, $i=1, 2$. Заметим, что $v_2(x, s)$ — величина, обратная другому нулю знаменателя. Кроме того, $s+\mu_1+\lambda_1+\lambda_2-\lambda_2 x=\mu_1(v_1+v_2)$, $\rho_1=v_1 v_2$. Теперь можно показать, что, коль скоро $\rho_1+\rho_2<1$, то

$$\lim_{t\to\infty}P_{00}(t)=\lim_{s\to 0}sg_{00}(s)=1-\rho_1-\rho_2.$$

Для производящей функции вероятностей стационарных состояний имеем

$$\lim_{s\to 0}sF(z, x, s)=\frac{(1-\rho_1-\rho_2)[1-v_1(x, 0)]}{[1-\rho_2 x-v_1(x, 0)][1-zv_1(x, 0)]}. \qquad (11.37)$$

Из $F(z, x, s)$ определяется $g_{00}(s)$, для чего находится нуль ζ, лежащий внутри единичного круга $|x|=1$. В результате получим $g_{00}(s)=\frac{\zeta^{h+1}}{\mu_2(1-\zeta)}$. Так как производящая функция определена, то путем дифференцирования можно найти вероятности.

Если $\mu_1=\mu_2=\mu$, то, оперируя по-прежнему внутри кругов $|z|<1$ и $|x|<1$, найдем нуль $\zeta=\alpha_2$, встречавшийся ранее в случае одноканальной системы, с той лишь особенностью, что здесь λ заменено на $\lambda_1+\lambda_2$:

$$F(z, x, s)=\frac{\zeta^{h+1}(1-x)(1-\zeta)^{-1}-x^{h+1}}{\mu[1-zv_1(x, s)][1-xv_2(x, s)]}. \qquad (11.38)$$

Произведя обратное преобразование функции $\frac{\zeta^{h+1}}{\mu(1-\zeta)}$, получим

$$P_{00}(t)=\frac{e^{-t(\mu+\lambda_1+\lambda_2)}}{\mu t}\sum_{j=0}^{\infty}\left(\sqrt{\frac{\mu}{\lambda_1+\lambda_2}}\right)^{j+h+1}\times$$
$$\times(j+h+1)I_{j+h+1}\left[2t\sqrt{\mu(\lambda_1+\lambda_2)}\right]. \qquad (11.39)$$

Преобразование Лапласа математического ожидания $E(m\mid h,\ t)$ числа требований с низшим приоритетом при условии, что начальное число требований равно h, имеет вид

$$\left.\frac{\partial P(z,\ x,\ s)}{\partial x}\right|_{\substack{z=1\\ x=1}}.$$

Произведя обратное преобразование, получим

$$E(m\mid h,\ t)=h+(\lambda_2-\mu)t+\mu\int_0^t P_{00}(u)\,du+$$
$$+\sqrt{\lambda_1\mu}\int_0^t(t-u)\frac{e^{-u(\mu+\lambda_1)}}{u}I_1(2u\sqrt{\lambda_1\mu})\,du. \qquad (11.40)$$

При $\lambda_1=0$ эти формулы приводятся к формулам, полученным для одноканальной системы с параметрами λ_2 и μ_2.

Распределение длительности периода занятости для системы без приоритетов находится из исходной системы уравнений. Кроме того, в данном случае в начальной точке находится поглощающий барьер, так что процесс прекращается, когда система в первый раз переходит в это состояние. Вместо $P_{nm}(t)$ используем $Q_{nm}(t)$. Область случайного блуждания, связанного с этой задачей, ограничена отражающими барьерами вдоль $n=0$, $m=1$ и в точке $(0,\ 0)$. Взяв, как и ранее, преобразование Лапласа, можно определить искомое распределение. В этом случае $Q^*_{00}(s)=\frac{\zeta^h}{s}$, где ζ — нуль функции $F(z,\ x,\ s)$ при $\mu_1\neq\mu_2$.

Вычислив обратное преобразование функции ζ^h, найдем искомое распределение $\frac{dQ_{00}}{dt}$. Соответствующие ему математическое ожидание и дисперсия равны соответственно

$$E(t\mid h)=\frac{h}{\mu_2(1-\rho_1+\rho_2)} \qquad (11.41)$$

и

$$V(t\mid h)=\frac{h\left[1+\rho_2-\rho_1\left(1-\frac{2\mu_2}{\mu_1}\right)\right]}{\mu_2^2(1-\rho_1-\rho_2)^3}. \qquad (11.42)$$

Если $\mu_1 = \mu_2 = \mu$, то

$$\frac{dQ_{60}}{dt} = h\left(\sqrt{\frac{\mu}{\lambda_1+\lambda_2}}\right)^h \frac{e^{-t(\mu+\lambda_1+\lambda_2)}}{t} I_h\left[2t\sqrt{\mu(\lambda_1+\lambda_2)}\right]. \quad (11.43)$$

Рассмотрим r приоритетов, прерывающих обслуживание, в системе с пуассоновским входящим потоком с параметром λ_k и экспоненциальным временем обслуживания с параметром μ_k для требований с k-м приоритетом. Среднее число требований, ожидающих начала обслуживания в системе, находящейся в стационарном состоянии, т. е. при $\sum\limits_{k=1}^{r} \frac{\lambda_k}{\mu} < 1$, равно [171]

$$\frac{\rho_k}{\left(1-\sum\limits_{n=1}^{k-1}\rho_n\right)\left(1-\sum\limits_{n=1}^{k}\rho_n\right)}.$$

Дисперсия равна

$$\frac{2\left(\sum\limits_{n=1}^{k-1}\rho_n\right)\rho_k^3}{\left(1-\sum\limits_{n=1}^{k-1}\rho_n\right)\left(1-\sum\limits_{n=1}^{k}\rho_n\right)} + \frac{\rho_k^2+\rho_k\left(1-\sum\limits_{n=1}^{k-1}\rho_n\right)\left(1-\sum\limits_{n=1}^{k}\rho_n\right)}{\left(1-\sum\limits_{n=1}^{k-1}\rho_n\right)^2\left(1-\sum\limits_{n=1}^{k}\rho_n\right)},$$

где $\rho_k = \frac{\lambda_k}{\mu}$.

Хиткоут рассмотрел также и общий случай. Уайт и Кристай [876] и Стефан [784] также исследовали эту задачу главным образом для стационарного случая. Стефан приводит некоторые вычисления для этой задачи, рассматривая среднее время ожидания и другие моменты для требований с низшим приоритетом. Он рассмотрел также время пребывания в системе и время ожидания, включая время ожидания вследствие прерывания обслуживания.

3. Динамические приоритеты

Дж. Джексон исследовал динамические приоритеты в одноканальной системе с дискретным временем. Исходя из заданного распределения вероятностей, поступающему требованию назначается определенный номер разряда срочности. Вновь поступающее требование получает преимущество перед ожидающим требованием в том и только в том случае, если разность между номером разряда срочности вновь поступающего требования и номером разряда срочности предыдущего требования не меньше времени, в течение которого ожидает предыдущее требование. Этот случай является примером того, как администрация проявляет заботу о тех клиентах, которые ожидали в течение длительного времени. Эта система приводится к обычной системе с «приоритетом состояний», если рассмотренное выше распределение вероятностей становится непрерывным, и к системе с обслуживанием по принципу «первым пришел — первым обслужен», если оно сводится к единственному значению.

Введем некоторые обозначения: p — вероятность поступления требования в течение рассматриваемого интервала; q — вероятность окончания обслуживания требования, если канал занят; π_n — вероятность того, что поступающее требование получит номер разряда срочности, не превышающий n; $\rho=\frac{p}{q}$ — загрузка системы; $\rho_n=\rho\pi_n$ — загрузка системы требованиями с номером разряда срочности, не превышающим n,

$$Z=\frac{1-q}{1-p}, \quad A=\rho Z.$$

Пусть в рассматриваемом периоде s_k — вероятность того, что в стационарном состоянии непосредственно после того, как требование покинуло систему (нижняя точка), и непосредственно перед тем, как поступило новое требование (если это происходит), в системе находится n требований. Обычным способом находим уравнения

$$\begin{gathered} p(1-q)s_0=q(1-p)s_1, \\ [p(1-q)+q(1-p)]s_k=p(1-q)s_{k-1}+q(1-p)s_{k+1}, \\ k=1,\ 2,\ \ldots \end{gathered} \tag{11.44}$$

Решение их имеет вид

$$s_k=(1-A)A^k. \tag{11.45}$$

Если $x_i(t)$ — вероятность того, что в стационарном состоянии в системе остается i требований после того, как обслуживание требований, находящихся в заданной нижней точке, продолжалось в течение t единиц времени, то

$$\begin{gathered} x_0(t)=x_0(t-1)+qx_1(t-1), \\ x_i(t)=(1-q)x_i(t-1)+qx_{i+1}(t-1), \quad i=1,\ 2,\ \ldots \end{gathered} \tag{11.46}$$

Решение этих уравнений имеет вид

$$\begin{gathered} x_i(t)=(1-A)A^iZ^t, \quad t=0,\ 1,\ \ldots,\quad i=1,\ 2,\ \ldots, \\ x_0(t)=1-\sum_{i=1}^{\infty}x_i(t)=1-\rho Z^{t+1}. \end{gathered} \tag{11.47}$$

Пусть W — общее среднее время ожидания требований всех разрядов срочности в стационарном состоянии. Оно равно математическому ожиданию общего времени обслуживания требований, для которых время поступления в систему находится в данном интервале.

Для системы с обслуживанием по принципу «первым пришел — первым обслужен» выражение для вероятности того, что время ожидания не больше t, имеет вид

$$P(\leqslant t)=1-\rho Z^{t+1}. \tag{11.48}$$

Для случая статического приоритета имеем

$$\frac{W_j}{W}=\frac{1-\rho}{(1-\rho_{j-1})(1-\rho_j)}, \tag{11.49}$$

где W_j — среднее время ожидания требований с номером разряда срочности j в равновесном состоянии.

Вывод этого выражения для дискретного времени аналогичен выводу, который приводит Кобхем.

Дж. Джексон указывает следующие ограничения для среднего времени ожидания W_j' требования с номером разряда срочности j при равновесном состоянии в случае динамического приоритета:

$$\begin{gathered}\frac{W_j'}{W}\geqslant\frac{1-\rho}{1-\rho_j},\\ \frac{W_j'}{W}\leqslant\frac{1}{1-\rho_{j-1}},\\ \frac{W_j'}{W}\leqslant 1+(1-Z)\sum_{i<j}\rho_i\end{gathered} \tag{11.50}$$

и

$$\frac{1-\rho}{1-\rho_j}\leqslant\frac{W_j}{W_j'}\leqslant\frac{1}{1-\rho_{j-1}}.$$

Первое неравенство доказывается путем замены этой системы такой же системой, в которой все номера разрядов срочности, меньшие j, увеличиваются до j, а затем полученная система заменяется системой со статическим приоритетом, имеющей те же значения p и q и те же распределения номеров разрядов срочности.

Дж. Джексон предложил методику вычисления последовательности вероятностей времени ожидания в системе с динамическим приоритетом.

11.3. Случайный выбор для обслуживания

1. Многоканальная система с экспоненциальным временем обслуживания

Первым исследовал задачу случайного выбора на обслуживание Меллор. Однако только Воло дал точную формулировку этой задачи. Основываясь на работах Воло, Полячек получил асимптотические формулы. Пальм, Полячек и Риордан пытались получить решение задачи Воло. В большинстве случаев полученные решения являются достаточно приближенными. Нашей задачей является нахождение распределения времени ожидания. Выполним это, введя вспомогательные понятия.

Допустим, что в многоканальную систему поступает пуассоновский поток с параметром λ, а каждый из c каналов имеет одинаковое экспоненциальное распределение времени обслуживания с параметром μ. Допустим также, что требование остается в системе до тех пор, пока обслуживание его не закончится. Выбор ожидающих требований на обслуживание в один из каналов производится случайным образом. Следовательно, вероятность того, что ожидающее требование поступит на обслуживание, зависит от числа требований, ожидающих в это время, и не зависит от того, сколько времени это требование уже ожидало. Оказывается, что чем дольше требование ожидает, тем меньше вероятность того, что оно будет обслужено в последующем промежутке времени t. Парадокс заключается в том, что требование может покинуть систему и вновь поступить в нее; при этом вероятность того, что оно будет обслужено, увеличится. Объяснение этого парадокса состоит в том, что чем дольше требование ожидает, тем больше вероятность того, что в систему, в которой находится большое число требований, поступит новое требование.

Рассматривая уход из системы и повторное поступление требований, Пальм показал, что вычисленное заранее распределение времени ожидания при случайном выборе на обслуживание при перегрузках не применимо к вычислению дополнительного времени ожидания t. В случае обслуживания требований в порядке поступления (время ожидания не зависит от числа требований, поступивших после данного требования) при сделанных ранее допущениях время ожидания имеет экспоненциальное распределение; оставшееся время ожидания не увеличивается с течением времени, а остается таким же. Так, обычно при прочих равных условиях для клиентов более выгодно, если обслуживание производится в порядке поступления; это положение проверялось в отношении стоимости обслуживания и с другой точки зрения, например с целью относительного уменьшения времени ожидания.

Обозначим через $f_n(t)$ вероятность того, что данное из n требований, ожидающих в системе в произвольный момент времени, будет ожидать еще в течение времени t. Можно допустить, что за начало отсчета берется произвольный момент времени. Рассматриваемое требование ожидает в течение времени t, и, хотя к моменту его поступления в системе находилось $n-1$ других требований, возможно, что за это время некоторые требования покинут очередь и поступят новые требования. Следовательно, возможно, что данное требование будет продолжать ожидать при большем или при меньшем числе требований, находящихся в очереди, и вероятность того, что в момент времени t оно еще не поступит на обслуживание, зависит от числа требований, находящихся в очереди. На основании марковского свойства этого процесса, т. е. полугруппового свойства $P(t+s)=P(t)P(s)$, вероятность $f_n(t)$ можно определить, рассмотрев, что происходит в промежутке времени $(0, \Delta t)$ и в примыкающем к нему промежутке $(t, t-\Delta t)$.

1. За время Δt с вероятностью $\lambda\Delta t$ может поступить одно требование, а затем данное требование вместе с другими n требова-

ниями начинает ожидать в течение времени $t-\Delta t$ (всего имеется $n+1$ ожидающих требований). Вероятность этого события равна $f_{n+1}(t-\Delta t)$, а $\lambda\Delta t f_{n+1}(t-\Delta t)$ — вероятность того, что происходят оба эти события.

2. За время Δt будет обслужено одно из $n-1$ остальных ожидающих требований. Вероятность того, что за время Δt будет обслужено одно из этих требований, равна $\mu c\Delta t$, а вероятность того, что это будет любое требование, кроме данного, равна $\frac{n-1}{n}$. Следовательно, вероятность того, что будет обслужено любое требование, кроме данного, составляет $\frac{n-1}{n}\mu c\Delta t$. Вероятность того, что данное требование вместе с другими $n-2$ требованиями начинает ожидать в течение времени $t-\Delta t$, равна $f_{n-1}(t-\Delta t)$, и, следовательно, вероятность обоих событий составляет $\frac{n-1}{n}c\mu\Delta t f_{n-1}(t-\Delta t)$.

3. За время Δt ни одно требование не поступает и не заканчивается ни одно обслуживание. Вероятность этого события равна $(1-\lambda\Delta t)(1-\mu c\Delta t)$. Вероятность того, что данное требование вместе с другими $n-1$ требованиями начинает ожидать в течение времени $t-\Delta t$, равна $f_n(t-\Delta t)$. Снова рассматривается произведение этих двух событий.

Заметим, что в первом случае необходимо учитывать вероятность того, что ни одно требование не будет обслужено, равную $(1-\mu c\Delta t)$, а во втором случае — вероятность того, что не поступит ни одно новое требование, равную $(1-\lambda\Delta t)$. Можно было бы также рассмотреть случаи поступления более одного требования, однако для пуассоновского входящего потока и экспоненциального распределения времени обслуживания этими величинами можно пренебречь.

По закону полной вероятности для описанного случая имеем

$$f_n(t)=(1-\lambda\Delta t)(1-\mu c\Delta t)f_n(t-\Delta t)+\lambda f_{n+1}(t-\Delta t)\Delta t+$$
$$+\frac{n-1}{n}\mu c f_{n-1}(t-\Delta t)\Delta t.$$

Это уравнение справедливо при $n>1$. Если $n=1$ и канал свободен, то данное требование поступает на обслуживание, и в этом случае имеем

$$f_1(t)=(1-\lambda\Delta t)(1-\mu c\Delta t)f_1(t-\Delta t)+\lambda f_2(t-\Delta t)\Delta t.$$

Проведя обычные упрощения, разделив на Δt и положив $\Delta t\to 0$, эти уравнения можно записать в следующем виде:

$$f'_n(t)=-(\lambda+\mu c)f_n(t)+\frac{n-1}{n}\mu c f_{n-1}(t)+\lambda f_{n+1}(t),\ n>1,$$
$$f'_1(t)=-(\lambda+\mu c)f_1(t)+\lambda f_2(t). \tag{11.51}$$

Рассмотрим решение этой системы при $f_n(0)=1$ $(n=1, 2, \ldots)$. Заметим, что условие $0 \leqslant f_n(t) \leqslant 1$ должно выполняться для всех n, так как $f_n(t)$ — некоторая вероятность.

Если в индексе условных вероятностей $f_n(t)$ не учитывается рассматриваемое требование, то эти уравнения приводятся к системе, которую рассмотрели Воло, Полячек, Риордан и Вилкинсон.

Для этого случая примем обозначение $F_{n-1}(t)=f_n(t)$ и подставим его в систему уравнений (11.51). Заменив n на $n+1$, получим

$$F'_n(t) = -(\lambda+\mu c)F_n(t) + \frac{n}{n+1}\mu c F_{n-1}(t) + \lambda F_{n+1}(t), \quad n \geqslant 1,$$

$$F'_0(t) = -(\lambda+\mu c)F_0(t) + \lambda F_1(t). \qquad (11.52)$$

Пусть $\rho = \frac{\lambda}{\mu}$. Разделим обе части уравнений (11.51) на λ и примем $\lambda=1$, т. е. рассмотрим случай, когда параметр λ равен одной временно́й единице. Затем введем производящую функцию

$$\varphi(z, t) = \sum_{n=1}^{\infty} f_n(t) z^n. \qquad (11.53)$$

Эта функция удовлетворяет граничным условиям

$$\varphi(0, t)=0, \quad \varphi(z, 0) = \sum_{n=1}^{\infty} z^n = \frac{z}{1-z}. \qquad (11.54)$$

Умножая n-е уравнение (11.51) на nz^n и суммируя, получим гиперболическое дифференциальное уравнение

$$\frac{\partial^2\varphi(z, t)}{\partial z\,\partial t} + \frac{(1-z)(z-\rho)}{z}\frac{\partial\varphi(z, t)}{\partial z} + \frac{1}{z^2}\varphi(z, t) = 0. \qquad (11.55)$$

Взяв преобразование Лапласа по t и решив полученное линейное дифференциальное уравнение в частных производных первого порядка, получим преобразование Лапласа производящей функции. Предоставим это выполнить читателю.

Пальм [640], решив систему уравнений (11.51), получил разложение в ряд по степеням t

$$\varphi(z, t) = \sum_{n=0}^{\infty} A_n(z)\frac{t^n}{n!}, \qquad (11.56)$$

где $A_n(z) = \varphi^{(n)}(z, 0)$, в соответствии с разложением в ряд Маклорена, есть n-я производная от φ при $t=0$.

Таким образом,

$$A_n(z) = \sum_{m=0}^{\infty} f_m^{(n)}(0) z^m, \qquad (11.57)$$

а коэффициент $f_m^{(n)}(0)$ можно определить, дифференцируя систему уравнений (11.51) $n-1$ раз и положив $\lambda=1$. Так как

$f_m(0)=1$, то при $n=1$ имеем $f_m^{(1)}(0)=-\left[\frac{1}{m-\rho}\right]$, $\rho=\frac{\lambda}{\mu c}$. Это выражение можно использовать для определения $f_m^{(2)}(0)$ и т. д. Таким образом, если функция $\varphi(z, t)$ определена, то $f_m(t)$ найдем как коэффициент разложения функции φ в ряд по возрастающим степеням z. Применяя разложение в ряд к дифференциальному уравнению в частных производных для производящей функции и приравнивая t нулю, получаем выражение

$$A_n(\rho)=-\lim_{z\to 0}\frac{A_{n-1}(z)}{z}-\frac{1}{\rho}\int_0^{\rho}A_{n-1}(z)\,dz, \tag{11.58}$$

в котором необходимо вычислить все значения A_n. Моменты производящей функции имеют вид

$$\mu_n(z)=-\int_0^{\infty}t^n\frac{\partial\varphi(z, t)}{\partial t}\,dt=n\int_0^{\infty}t^{n-1}\varphi(z, t)dt. \tag{11.59}$$

Интегрируя по t дифференциальное уравнение для φ и используя второе граничное условие, получим

$$\mu_1(z)=\frac{\rho}{2-\rho}z\frac{2-z}{(1-z)^2}. \tag{11.60}$$

Вероятность того, что время ожидания больше t, равна

$$P(>t)=(1-\rho)\rho\sum_{n=1}^{\infty}\rho^n f_n(t)=\frac{1-\rho}{\rho}\varphi(\rho, t). \tag{11.61}$$

Интегрируя по t, получим среднее время ожидания

$$W_q=\frac{1-\rho}{\rho}\mu_1(\rho). \tag{11.62}$$

Второй момент распределения времени ожидания равен

$$\frac{1-\rho}{\rho}\mu_1(\rho)=\frac{\rho}{(1-\rho)^2}\cdot\frac{4}{2-\rho}. \tag{11.63}$$

Разложив функции $\mu_1(z)$ и $\mu_2(z)$ в ряд, получим первый и второй моменты для отдельных значений $f_n(t)$. Они равны коэффициентам при z^n в этих разложениях.

Полячек [680] применил другой способ решения, использовав систему (11.52) при $\rho=\frac{\lambda}{c}$ и $\mu=1$.

Вначале заметим, что

$$P(<t)=1-\sum_{j=0}^{\infty} p_{c+j}F_j(t)=1-p_c\sum_{j=0}^{\infty}\rho^j F_j(t)=$$

$$=1-\left[\sum_{n=0}^{c-1}\frac{\lambda^n}{n!}+\frac{\lambda^c}{c!(1-\rho)}\right]^{-1}\frac{\lambda^c}{c!}\sum_{j=0}^{\infty}\rho^j F_j(t), \qquad (11.64)$$

где p_{c+j} — вероятность того, что в стационарном состоянии в системе находится $c+j$ требований.

Это выражение совпадает с формулой, полученной для системы с обслуживанием требований в порядке поступления.

Если умножить (11.52) на $\left(\frac{1}{\sqrt{\lambda c}}\right)\rho^{\frac{n}{c}}e^{(c+\lambda)t}$ и ввести обозначения $s=t\sqrt{\lambda c}$ и $g_n(s)=\rho^{\frac{n}{2}}e^{(c+\lambda)t}F_n(t)$, то уравнения (11.52) примут вид

$$\frac{dg_n(s)}{ds}=\frac{n}{n+1}g_{n-1}(s)+g_{n+1}(s), \quad n=0, 1, \ldots, \qquad (11.65)$$

и

$$g_n(0)=\rho^{\frac{n}{2}}, \quad n=0, 1, \ldots. \qquad (11.66)$$

Так как $0<F_n\leqslant 1$, то интеграл

$$\varphi_n(z)=\int_0^{\infty} e^{-2zs}g_n(s)\,ds \qquad (11.67)$$

сходится для $\operatorname{Re}(z)>\frac{(c+\lambda)}{2\sqrt{\lambda c}}=\frac{1+\rho}{2\sqrt{\rho}}$.

Применив это преобразование к последней системе уравнений, получим

$$2z\varphi_n(z)-\rho^{\frac{n}{2}}=\frac{n}{n+1}\varphi_{n-1}(z)+\varphi_{n+1}(z), \quad n=0, 1, \ldots \qquad (11.68)$$

Умножая n-е уравнение на $(n+1)x^n$ и суммируя, имеем

$$2z\sum_{n=0}^{\infty}(n+1)x^n\varphi_n(z)-\frac{1}{(1-x\sqrt{\rho})^2}=$$

$$=\sum_{n=1}^{\infty}nx^n\varphi_{n-1}(z)+\sum_{n=0}^{\infty}(n+1)x^n\varphi_{n+1}(z). \qquad (11.69)$$

Введем обозначение

$$\Phi(x,\ z)=\sum_{n=0}^{\infty} x^n \varphi_n(z). \qquad (11.70)$$

После упрощения при $|x|<\frac{1}{\sqrt{\rho}}$ имеем

$$(x^2-2xz+1)\frac{d\Phi}{dx}+(x-2z)\Phi=-\frac{1}{(1-x\sqrt{\rho})^2}. \qquad (11.71)$$

Нам нужны решения этого уравнения, непрерывные по x при $|x|<\frac{1}{\sqrt{\rho}}$ и достаточно больших вещественных z. Если коэффициент при $\frac{d\Phi}{dx}$ разложить на множители $(x-\alpha)(x-\beta)$, где α имеет положительный знак перед радикалом, и затем разложить β в ряд при $|z|>1$, то получим единственное решение для $\beta\approx\frac{1}{2z}$ при $|x|<\frac{1}{\sqrt{\rho}}$. Разложим $\Phi(x,\ z)$ и правую часть равенства в ряд Тейлора по $(x-\beta)$. Выбирая для постоянной, которая появляется при решении линейных уравнений первого порядка, такое значение, что при $x=\beta$ функция $\Phi(x,\ z)$ остается конечной, имеем

$$\Phi(x,\ z)=(\alpha-x)^m(\beta-x)^{-(m-1)}\times$$

$$\times\int_x^{\beta}(\alpha-y)^{-(m-1)}(\beta-y)^m\frac{dy}{(1-y\sqrt{\rho})^2}, \qquad (11.72)$$

где $m=\frac{\beta}{(\alpha-\beta)}=\left(\frac{z}{2\sqrt{z^2-1}}\right)-\frac{1}{2}$, а величины $(\alpha-y)^m$ и $(\beta-y)^m$ при $y=x$ совпадают с $(\alpha-x)^m$ и $(\beta-x)^m$.

Заметим, что

$$\Phi(x,\ z)=\int_0^{\infty}e^{-2zs}\sum_{n=0}^{\infty}x^n g_n(s)\,ds. \qquad (11.73)$$

Применив обратное преобразование, получим

$$\sum_{n=0}^{\infty}(x\sqrt{\rho})^n F_n(t)=\exp[-c(1+\rho)t]\times$$

$$\times\frac{1}{\pi i}\int_{-i\infty+a}^{i\infty+a}e^{2ct\sqrt{\rho}z}\,\Phi(x,\ z)\,dz, \qquad (11.74)$$

где

$$a>\frac{(1+\rho)}{2\sqrt{\rho}},\quad t>0$$

Переходя в комплексную плоскость, получим

$$\sum_{n=0}^{\infty} \rho^n F_n(t) = 2\int_0^{\pi} \exp\left[-\left(1+\rho-2\sqrt{\rho}\cos w\right)ct\right] \times$$
$$\times\left(1-\sqrt{\rho}e^{iw}\right)^{i\,\mathrm{ctg}\,w-2}\left(1-\sqrt{\rho}\,e^{-iw}\right)^{-i\,\mathrm{ctg}\,w-2} e^{w\,\mathrm{ctg}\,w} \times$$
$$\times\left(e^{\pi\,\mathrm{ctg}\,w}+1\right)^{-1}\sin w\,dw. \qquad (11.75)$$

При $c\to\infty$ имеем

$$P(<t) = 1-\left(1+2be^{b^2}\int_{\infty}^{b} e^{-t^2}dt\right)^{-1} 2\sqrt{2b\theta}\,\times$$
$$\times K_1(2\sqrt{2b\theta})\left[1+O\left(\frac{1}{\sqrt{c}}\right)\right], \qquad (11.76)$$

где $K_1(z) = \int_0^{\infty} e^{-z\,\mathrm{ch}\,t}\,\mathrm{ch}\,t\,dt$ — представление модифицированной функции Бесселя третьего рода первого порядка в виде интеграла Шлефли при $b=(1-\rho)\sqrt{\frac{c}{2}}$ и $\theta = t\sqrt{\frac{c}{2}}$.

Заметим, что среднее время ожидания будет таким же, как и в системе с обслуживанием в порядке поступления.

Риордан [725] исследовал и вывел выражение для $P(>t)$, применяя разложение системы (11.52) в ряд по экспонентам. Он получил для всех значений t при $\rho=\frac{\lambda}{c}$, $\mu=1$ выражение

$$P(>t) = \frac{1}{2}\left(1-\sqrt{\frac{\rho}{2}}\right)e^{-t(1-\rho)\left(1-\sqrt{\frac{\rho}{2}}\right)}+$$
$$+\frac{1}{2}\left(1+\sqrt{\frac{\rho}{2}}\right)e^{-t(1-\rho)\left(1+\sqrt{\frac{\rho}{2}}\right)}, \qquad (11.77)$$

которое справедливо при $\rho<0{,}7$.

Он исходил из представления

$$P(>t) = A_1 e^{-\frac{(1-\rho)t}{x_1}} + A_2 e^{-\frac{(1-\rho)t}{x_2}} + \ldots. \qquad (11.78)$$

Разлагая его в ряд вида $A_1x_1^k + A_2x_2^k + \ldots$, этот автор получил моменты M_k и, приравняв коэффициенты, нашел значения A_i.

Важные вычисления выполнил Вилкинсон [882]. Леруа [505] исследовал эту задачу, применяя матричные методы.

2. Одноканальная система с постоянным временем обслуживания

Бёрк [100] исследовал задачу случайного выбора для обслуживания в случае одноканальной системы с пуассоновским входящим потоком с параметром λ и постоянным временем обслуживания при равновесном состоянии. Время обслуживания принимается равным одной временно́й единице. На основании доказательства А. Я. Хинчина вероятности стационарного состояния будут одними и теми же независимо от того, рассматриваются ли они для всего процесса или только для множества дискретных моментов ухода из системы обслуженных требований. Это видно из того, что полученная Кендаллом производящая функция вероятностей, рассматриваемых в моменты ухода требований, совпадает с производящей функцией вероятностей, рассматриваемых в произвольный момент, которую получил Кроммелин. Необходимо определить распределение времени ожидания $P(\leqslant t)$.

В момент ухода обслуженного требования, который происходит после поступления ожидающего требования, данное требование будет находиться в системе вместе с $n-1$ другими [с вероятностью $P(n|\lambda)$], если в момент ухода, непосредственно предшествующий поступлению данного требования, в системе находилось k требований ($k=0, \ldots, n$) и до окончания обслуживания следующего требования поступило еще $n-k$ требований (включая и рассматриваемое). Имеем

$$P(n|\lambda)=(p_0+p_1)\frac{\lambda^{n-1}e^{-\lambda}}{(n-1)!}+\ldots+p_n e^{-\lambda}. \qquad (11.79)$$

Это выражение является уравнением Кроммелина для p_{n-1}, и, следовательно,

$$P(n|\lambda)=p_{n-1}. \qquad (11.80)$$

Обозначим через $q(t|n)$ вероятность того, что время ожидания не превышает t при условии, что ожидающее требование является одним из n требований, ожидающих начала обслуживания в момент, когда происходит уход первого обслуженного требования после поступления данного требования в систему. Тогда

$$P(\leqslant t)=\sum_{n=1}^{\infty} P(n|\lambda)\, q(t|n). \qquad (11.81)$$

Но промежуток времени между моментом поступления данного требования и моментом ухода первого обслуженного требования однозначно определяется распределением пуассоновского входящего потока в интервале обслуживания и не зависит от длительности ожидания в будущем вследствие того, что выбор на обслуживание производится случайным образом.

Если время ожидания T выражается как сумма части его T', которая является целым числом временных единиц, и остальной дробной части T'' и если промежуток времени t аналогично разделен на t' и t'', то, применяя предыдущее замечание о независимости, найдем

$$q(t \mid n) = P(T \leqslant t \mid n) = P(T' < t' \mid n) = \\ = P(T' = t' \mid n) = P(T'' < t''), \quad (11.82)$$

или, так как T'' имеет равномерное распределение,

$$q(t' \mid n) = \sum_{i=0}^{t-1} P(T' = i \mid n) + t'' P(T' = t' \mid n). \quad (11.83)$$

Если ввести обозначение $P(T' = i \mid n) \equiv Q_i(n)$, то

$$Q_0(n) = \frac{1}{n} \quad (11.84)$$

и

$$Q_i(n) = [1 - Q_0(n)] \sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!} Q_{i-1}(n + j - 1), \quad i = 1, \ldots, t'.$$

Решив это уравнение, найдем $q(t \mid n)$, а после того как будет получено p_n, найдем и $P(\leqslant t)$. Бёрк приводит графические вычисления распределения времени ожидания для 130 значений времени обслуживания и сравнивает результаты, полученные для системы с постоянным временем обслуживания при обслуживании требований в порядке поступления, с результатами, полученными для системы с экспоненциальным временем обслуживания при случайном выборе на обслуживание для большого уровня нагрузки.

11.4. Распределение требований по каналам

Вслед за Полячеком О. А. Вольберг [968] исследовал многоканальную систему, имеющую c параллельных каналов, с пуассоновским входящим потоком с параметром λ и произвольным распределением времени обслуживания $B(t)$, одинаковым для всех каналов и всех требований. Требования распределяются по каналам в порядке поступления: первое требование направляется в первый канал, второе — во второй и т. д., c-е требование — в последний, а затем $(c+1)$-е требование направляется в первый канал, таким же образом все каналы заполняются во второй раз и т. д. О. А. Вольберг вычислил распределение времени ожидания в стационарном состоянии для одного канала. Как уже указывалось в гл. 3, этому случаю соответствует понятие о фильтрованном пуассоновском потоке, из которого в одноканальную систему допускается каждое c-е требование, т. е. входящий поток является эрланговским и исследуется система $E_c/G/1$.

Пусть время ожидания $(n+1)$-го требования, поступившего в канал в момент времени t, есть случайная величина $w_n(t)$, а время обслуживания — случайная величина u_n. Если обозначить через $\tau_{n,c}$ промежуток времени между данным требованием и следующим [т. е. $(n+c+1)$-м] требованием, поступающим в ту же самую очередь, то найдем

$$w_{n+c}(t)=w_n(t-\tau_{n,c})+u_n-\tau_{n,c}. \qquad (11.85)$$

Пусть

$$P_n(w,\ t)=P\,[w_n(t)\leqslant w] \qquad (11.86)$$

и

$$F_{n,c}(\tau,\ t)=P(\tau_{n,c}\leqslant\tau). \qquad (11.87)$$

Затем О. А. Вольберг получил следующее выражение:

$$\frac{\partial}{\partial\tau}F_{n,c}(\tau,\ t)=\frac{(t-\tau)^n}{n!}\,\frac{c-1}{(c-1)!}\,\frac{(n+c)!}{t^{n+c}},\quad t>0. \qquad (11.88)$$

Это выражение вместе с первым соотношением он использовал для вывода рекуррентной формулы, по которой вычисляется $P_n(w,\ t)$ (если известна вероятность того, что начальное время занятости каждого канала не превосходит заданного). Тогда для любого требования, ожидающего перед каналом, распределение времени ожидания имеет вид

$$P(w,\ t)=\sum_{n=0}^{\infty}\frac{(\lambda t)^n e^{-\lambda t}}{n!}P_n(w,\ t). \qquad (11.89)$$

Произведя некоторые выкладки и принимая, что $t\to\infty$, О. А. Вольберг получил преобразование Лапласа—Стилтьеса $\psi(s)$ функции распределения времени ожидания в стационарном состоянии (т. е. когда загрузка $\rho=\frac{\lambda}{c\mu}<1$, где μ — интенсивность обслуживания):

$$\psi(s)=\frac{sQ(s)}{\beta(s)-(1-\lambda s)^c}. \qquad (11.90)$$

Здесь $\beta(s)$ — преобразование функции распределения времени обслуживания,

$$Q(s)=\frac{c\lambda(1-\rho)(s_1-s)(s_2-s)\ldots(s_{c-1}-s)}{s_1s_2\ldots s_{c-1}}, \qquad (11.91)$$

а s_i $(i=1,\ \ldots,\ c-1)$ — нули функции $[\beta(s)-(1-\lambda s)^c]$, которые лежат в правой полуплоскости. О. А. Вольберг показал, что при

$c=1$ эта функция совпадает с результатом для одноканальной системы, полученным А. Я. Хинчиным. При $\rho>1$ имеем

$$\lim P\left[\sqrt{\lambda t}w+(\rho-1)t,\ t\right]=$$

$$=\frac{1}{2\pi}\int\limits_{-\infty}^{\frac{w}{\sqrt{\frac{(\mu_2-\mu_1)}{c}+\frac{\mu_1^2}{c^2}}}} e^{-\frac{x^2}{2}}dx, \qquad (11.92)$$

где $\mu_1=\frac{1}{\mu}$, а μ_2 — второй момент распределения времени обслуживания.

11.5. Дисциплина очереди „пришел последним — обслужен первым"

Рассмотрим случай, когда поступающие требования, например, предметы массового производства, располагаются сверху других, ранее изготовленных предметов. Если такой предмет должен использоваться, то он берется сверху (поступает на обслуживание). В то время когда этот предмет используется, поступают другие предметы и складываются в порядке поступления. Это случай дисциплины очереди «пришел последним — обслужен первым». Требуется вычислить время ожидания для стационарного состояния.

Допустим, что требование, поступающее в одноканальную систему по закону Пуассона с параметром λ, является n-м в очереди и время обслуживания имеет произвольное распределение $B(t)$ с математическим ожиданием $\frac{1}{\mu}$. Для вычисления времени ожидания этого требования используется метод, аналогичный тому, который применяется при вычислении периода занятости в гл. 8. Поступившее требование ожидает, пока закончится обслуживание требования, находящегося в обслуживающем устройстве. Если y_0 — длительность оставшегося времени обслуживания, то за это время может поступить m требований, и все они будут обслужены, как и те требования, которые поступят за время их обслуживания, прежде чем рассматриваемое требование поступит на обслуживание.

Таким образом, по существу, распределение времени ожидания для системы с дисциплиной очереди «пришел последним — обслужен первым» является распределением длительности периода занятости; число требований, находящихся в очереди, равное вначале n, снова становится равным n. Однако рассматриваемые переходы не зависят от n, и можно считать, что в данном случае длительность периода занятости для системы с обслуживанием в порядке поступления, т. е. время, необходимое для первого возвращения к состоянию, когда очереди нет, если в начальный момент очередь отсутствовала, в обоих случаях одна и та же. Таким образом, вычисления производятся точно так же, как и для распределения длительности периода занятости системы с обслуживанием требований в порядке поступления при единственном изменении, сущность которого выяснится ниже.

Распределение длительности y для стационарного состояния в данном случае не является распределением времени обслуживания, так как длительность y также определяется временем, когда рассматриваемое требование поступает в систему. В общем виде преобразование Лапласа—Стилтьеса функции распределения времени ожидания записывается как

$$\gamma(s) = \psi\{s + [1 - \Gamma(s)]\lambda\}, \tag{11.93}$$

где $\psi(s)$ и $\Gamma(s)$ соответственно преобразования Лапласа—Стилтьеса функции распределения y и распределения длительности периода занятости (см. гл. 8).

Рассмотрим дифференциально-разностные уравнения для вероятностей $p_n(y, t)$, определяющих число n требований, находящихся в системе в момент времени t, и время y, необходимое для завершения обслуживания требования, а затем введем производящую функцию этих вероятностей и возьмем ее преобразования Лапласа—Стилтьеса по y и t. Переходя к стационарному состоянию, Уишарт получил

$$\psi(s) = \frac{1 - \rho + \lambda[1 - \beta(s)]}{s}. \tag{11.94}$$

Здесь $\beta(s)$ — преобразование Лапласа—Стилтьеса функции $B(t)$. Окончательно находим

$$\gamma(s) = \frac{1 - \rho + \lambda[1 - \Gamma(s)]}{\left[\dfrac{s}{\lambda + 1 - \Gamma(s)}\right]}. \tag{11.95}$$

Упражнение. Покажите, что в данном случае среднее время ожидания то же самое, что и для марковской системы с дисциплиной очереди «первым пришел — первым обслужен», а дисперсия для рассматриваемого случая велика. Напомним, что для вычисления моментов производится дифференцирование преобразования Лапласа—Стилтьеса при $s=0$.

Пусть в одноканальную систему поступает поток с ограниченным последействием и время обслуживания имеет экспоненциальное распределение; тогда остаточное время обслуживания будет также распределено по экспоненциальному закону; следовательно, время ожидания будет таким же, как период занятости при времени обслуживания, распределенном, в свою очередь, как и период занятости. В гл. 10 из статьи Коноллн было выведено преобразование Лапласа—Стилтьеса функции распределения времени ожидания

$$\frac{1-\lambda+\lambda(1-\xi)}{\frac{s}{\mu}+1-\xi}, \tag{11.96}$$

которое также имеет бо́льшую дисперсию, чем в случае дисциплины очереди «первым пришел — первым обслужен».

Задачи

1. Рассмотрите с помощью числового примера систему с приоритетом, прерывающим обслуживание.

2. Допустим, что в c-канальной системе с пуассоновским входящим потоком с параметром λ и одинаковым экспоненциальным временем обслуживания с параметром μ выбор требований на обслуживание производится любым способом. Пусть $f_n(t)$ — вероятность того, что в промежутке времени t после того, как число ожидающих требований стало равным n, перегрузка не уменьшилась и, следовательно, число требований, ожидающих в очереди, не стало равным n—1. Таким образом, $f_0(t)$, искомое распределение длительности перегрузки, есть вероятность того, что перегрузка сохраняется по крайней мере в течение времени t. Покажите, что уравнения имеют вид

$$\frac{1}{\lambda}f'_n(t)=-\frac{1+\rho}{\rho}f_n(t)+f_{n+1}(t)+\frac{1}{\rho}f_{n-1}(t),\quad n>0,$$

$$\frac{1}{\lambda}f'_0(t)=-\frac{1+\rho}{\rho}f_0(t)+f_1(t),\quad n=0,$$

$$\frac{1}{\lambda}f_n(0)=1,\quad \rho=\frac{\lambda}{c\mu}.$$

Покажите также, что если

$$F(x,\ t)=\sum_{k=0}^{\infty}f_n(t)\,x^n$$

[этот ряд сходится при $|x|<1$, так как $0\leqslant f_n(t)\leqslant 1$], то

$$\frac{1}{\lambda}\frac{\partial F}{\partial t}+\frac{(1-x)(x-\rho)}{\rho x}F+\frac{f_0(t)}{x}=0$$

при $F(0,\ t)=f_0(t)$, $F(x,\ 0)=\dfrac{1}{1-x}$.

Применяя предложенный Пальмом метод разложения в ряд при случайном выборе на обслуживание, покажите, как можно получить решение приведенной выше системы уравнений. Покажите также, что $(1-\rho)F(\rho,\ t)$ — вероятность того, что за время t, истекшее после случайно выбранного момента времени, перегрузка не исчезла. Преобразовав по Лапласу дифференциальное уравнение в частных производных и рассматривая нули знаменателя преобразования, покажите, что $f_n(t)$ задается той же формулой, что и распределение длительности периода занятости в одноканальной системе, в которой обслуживание требований производится в порядке поступления, если заменить μ на $c\mu$ [795].

3. Покажите справедливость выражения, полученного для $P(>t)$ в многоканальной системе со случайным выбором на обслуживание.

4. Интересным, хотя и не связанным непосредственно с данной главой, является уравнение второго порядка

$$\frac{\partial P(z, t)}{\partial t} = -\frac{\partial b(z) P(z, t)}{\partial z} + \frac{1}{2} \frac{\partial^2 a(z) P(z, t)}{\partial z^2},$$

которое превращается в уравнение Фоккера—Планка, если $a(z)$ и $b(z)$ — постоянные. Уравнение Фоккера—Планка, которое является общим уравнением теплопроводности или диффузии, часто встречается в теории случайных процессов. Это уравнение справедливо только в том случае, если случайный процесс является марковским и имеет нормальное распределение. Здесь $P(z, t)$ — вероятность того, что в момент времени t случайная величина имеет значение z.

Другой известный вариант этого уравнения получим, положив $b(z) = -\beta z$ и $a(z) = 2D$, где β и D — постоянные. Если $D = 0$, то имеем линейное уравнение, которое нам часто встречалось в тексте. Решите уравнение теплопроводности, положив $b(z) = 0$ и $a(z) = a$. Решение его имеет вид

$$P(z, t) = \frac{1}{(2\pi a t)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2at}\right).$$

Найдите также решение этого уравнения при произвольных постоянных $a > 0$ и b. В этом случае решение будет таким же, что и приведенное выше, с тем лишь отличием, что x^2 заменяется на $(x - bt)^2$.

Глава 12

МНОГОФАЗОВОЕ ОБСЛУЖИВАНИЕ (ПОСЛЕДОВАТЕЛЬНО РАСПОЛОЖЕННЫЕ КАНАЛЫ)

12.1. Введение

В большинстве работ, посвященных системам массового обслуживания с несколькими обслуживающими устройствами, расположенными последовательно, исследование ограничивается рассмотрением систем с пуассоновским входящим потоком и экспоненциальным временем обслуживания при обслуживании требований в порядке поступления. Требование, находящееся в системе, в процессе обслуживания может пройти все фазы (последовательные каналы) вплоть до последней или может поступить на обслуживание в определенную фазу. Заметим, что рассмотренная выше система с многофазовым обслуживанием соответствует системе с последовательными каналами при отсутствии ожидания перед всеми каналами, кроме первого. Новое требование допускается на обслуживание после того, как предыдущее требование пройдет все фазы обслуживания. В этой главе мы допускаем, что ожидание разрешается перед различными фазами. Исследование почти целиком посвящено выводу уравнений для стационарного состояния.

Важным вопросом, который необходимо исследовать при изучении систем с последовательными каналами, является распределение выходящего потока одного канала, который затем является входящим потоком для следующего канала. Эту задачу исследовали Бёрк [99], Рейч [708]) и Коэн [137]. Заметим, что распределение выходящего потока системы, находящейся в стационарном состоянии, не зависит от выбранной дисциплины очереди. Для стационарного состояния системы $M/M/c$ Бёрк показал, что промежутки времени между требованиями, покидающими систему, взаимно независимы. В этой системе обслуженные требования, покидающие систему, образуют пуассоновский поток с тем же параметром λ, который характеризует распределение входящего потока. Это обстоятельство ожидалось ранее интуитивно, на основании свойств стационарного состояния. Бёрк показал также, что в стационарном режиме длительность случайных промежутков времени между моментами, когда требования покидают систему, не влияет на состояние системы в конце этого промежутка.

Рейч, применяя другие методы, показал, что если промежутки времени между требованиями, поступающими в одноканальную систему, и промежутки времени их обслуживания распределены по нормированному закону хи-квадрат с четырьмя степенями свободы (некоторое изменение допущений Бёрка), то распределение промежутков времени между требованиями, покидающими систему, не подчиняется закону хи-квадрат с четырьмя степенями свободы. Следовательно, в общем случае не следует ожидать, что выходящий поток будет совпадать с входящим потоком даже в стационарном состоянии. Кроме того, Рейч показал, что в одноканальной системе с пуассоновскими входящим и выходящим потоками время обслуживания либо с вероятностью 1 равно нулю, либо имеет экспоненциальное распределение. Эта теорема является обратной теореме Бёрка.

Финч [245] показал, что выходящий поток будет в точности пуассоновским только в том случае, когда допускается очередь бесконечной длины. Он показал, что $N=\infty$ и экспоненциальное время обслуживания являются необходимыми и достаточными условиями того, чтобы промежутки времени между требованиями в выходящем потоке были взаимно независимы и число требований, находящихся в очереди после ухода из системы обслуженных требований, не зависело от промежутка времени, истекшего от момента ухода из системы предыдущего требования.

О'Брайен [615] рассмотрел два последовательных канала в стационарном состоянии с пуассоновским входящим потоком и экспоненциальным временем обслуживания и нашел среднее число требований, находящихся в очереди, и среднее время ожидания.

Р. Джексон [369, 370] систематически исследовал марковские модели с последовательными каналами (в стационарном случае) при различном экспоненциальном распределении времени обслуживания в каждой фазе и получил вероятность того, что в различных очередях находится определенное число требований; он также определил среднее число требований и распределение времени ожидания в каждой фазе.

Дж. Джексон [366] рассмотрел систему, в которой в каждой фазе разрешен пуассоновский поток требований, поступающих как из системы, так и извне. Вероятность поступления требований в различные фазы различна. Время обслуживания имеет экспоненциальное распределение, а каждая фаза состоит из нескольких параллельных каналов. Для системы с пуассоновским входящим потоком и экспоненциальным временем обслуживания, в которой требования проходят через все каналы, начиная с первого, он получил выражение для вероятности того, что в некоторой фазе находится определенное число требований.

Хант [361] исследовал загрузку (отношение средней интенсивности входящего потока к средней интенсивности обслуживания) системы с пуассоновским входящим потоком и экспоненциальным временем обслуживания, когда требования не покидают систему до тех пор, пока не пройдут через все каналы, начиная с первого, для следующих случаев: 1) перед каждым каналом разрешена неограниченная очередь; 2) разрешена ограниченная положительным объемом очередь перед каждым каналом, кроме первого, который может иметь неограниченную очередь; 3) очередь отсутствует перед каждым каналом, кроме первого, где она может быть неограниченной (с задержкой обслуженного требования в канале до того момента времени, когда оно сможет поступить на обслуживание в следующий канал), и, наконец, 4) очередь отсутствует перед каждым каналом, кроме первого, где она может быть неограниченной; нет свободных обслуживающих устройств (вся очередь передвигается, как одно требование). Для последнего случая получена также вероятность того, что обслуживание будет закончено в момент времени t после того, как очередь начала передвигаться. Когда перед последовательными каналами очередь не разрешается, то эффективная интенсивность обслуживания канала равна данной интенсивности обслуживания, умноженной на ту долю времени, в течение которого канал не был закрыт.

Допуская, что в каждом из последовательных каналов время обслуживания, превышающее заданную длительность, имеет экспоненциальное распределение, Нелсон [604] вычислил для каждого канала вероятность того, что время ожидания больше заданного. Дебон и Кац [170], аппроксимируя сумму показательных функций функцией хи-квадрат, упростили выкладки Нелсона. Сакс [751] исследовал эргодичность систем с последовательными каналами.

12.2. Выходящий поток системы

При рассмотрении этого вопроса важно определить, является ли пуассоновским выходящий поток системы с пуассоновским входящим потоком. После этого можно рассмотреть несколько последовательных каналов, зная, что их входящие и выходящие потоки являются пуассоновскими. Дадим доказательство этого важного положения, которое приводит Бёрк.

Теорема 12.1. В системе с c параллельными каналами, пуассоновским входящим потоком с параметром λ и одинаковым для каждого канала экспоненциальным распределением времени обслуживания с параметром μ при стационарном состоянии выходящий поток имеет пуассоновское распределение.

Доказательство. Доказательство проведем в несколько этапов. Первая часть доказательства состоит в вычислении числа требований, находящихся в системе в момент времени, когда обслуженное требование покидает систему. Затем это выражение используется для определения начальных условий для вероятностей совместного появления двух событий: 1) через промежуток времени t после ухода обслуженного требования в системе находится данное число требований и 2) этот промежуток времени t меньше промежутка времени между требованиями, покидающими систему. Суммируя по числу требований, находящихся в системе, находим распределение промежутков времени между требованиями, покидающими систему. Затем определяем, что это распределение является экспоненциальным, или, что то же самое, уход требований из системы есть пуассоновский процесс. Число требований, находящихся в системе в момент времени, когда происходит уход обслуженного требования, и промежутки времени между моментами ухода требований взаимно независимы и, таким образом, вероятность совместного появления этих событий равна произведению отдельных вероятностей. Заметим, что для этого случая вероятность p_n того, что в системе при стационарном состоянии находится n требований, была вычислена в предыдущей главе. Когда требование покидает систему, то система переходит из состояния E_{n+1} в состояние E_n; обратный переход происходит при поступлении требования, когда система находится в состоянии E_n. Так как число переходов $E_n \to E_{n+1}$ не может отличаться от числа переходов $E_{n+1} \to E_n$ более чем на единицу, то доля требований, покидающих систему, когда она находится в состоянии E_n, приближается к тому же пределу (к вероятности для стационарного состояния), что и доля требований, поступающих в систему, когда она находится в состоянии E_n.

Обозначим через τ длительность произвольного промежутка времени между требованиями, покидающими систему, а через $k(t)$ — состояние системы в момент времени t после ухода предыдущего требования. Введем обозначения

$$F_n(t) \equiv P\,[k(t) = n \text{ и } t < \tau], \tag{12.1}$$

$$F(t) \equiv \sum_{n=0}^{\infty} F_n(t), \quad t < \tau. \tag{12.2}$$

Так, $F(t)$ — одномерное распределение, так как мы просуммировали по всем значениям другой переменной. Заметим, что $1-F(t)$ есть распределение длительности промежутка времени между требованиями, покидающими систему.

Имеем начальные условия

$$F_n(0) \equiv p_n. \tag{12.3}$$

Теперь запишем

$$\begin{aligned} &F_0(t+\Delta t) = (1-\lambda\Delta t)\,F_0(t),\\ &F_n(t+\Delta t) = (1-\lambda\Delta t)\,(1-n\mu\Delta t)\,F_n(t) + \lambda F_{n-1}(t)\,\Delta t, \quad n < c,\\ &F_n(t+\Delta t) = (1-\lambda\Delta t)\,(1-c\mu\Delta t)\,F_n(t) + \lambda F_{n-1}(t)\,\Delta t, \quad n \geqslant c. \end{aligned} \tag{12.4}$$

Произведя обычные упрощения, получим систему дифференциально-разностных уравнений, решение которой имеет вид

$$F_n(t) = p_n e^{-\lambda t}, \quad t < \tau. \tag{12.5}$$

Упражнение. Проверьте, что (12.5) действительно является упрощенной формой записи системы (12.4).

Суммируя по n от нуля до бесконечности, найдем, что промежутки времени между требованиями, покидающими систему, имеют то же самое распределение, что и промежутки времени между поступающими требованиями.

Отсюда вытекает также независимость τ и $k(\tau)$, так как

$$P[k(\tau+0)=n \text{ и } t<\tau<t+dt]=$$
$$=\begin{cases} F_{n+1}(t)(n+1)\mu dt \text{ при } n+1\leqslant c, & (12.6)\\ F_{n+1}(t)c\mu dt \text{ при } n+1>c. & (12.7)\end{cases}$$

После подстановки значения $F_n(t)$ это выражение распадается на функцию распределения $k(\tau)$ и функцию распределения τ, что доказывает взаимную независимость этих двух величин. Справедливо также, что τ не зависит от длительности всех (последующих) промежутков времени между требованиями, покидающими систему после момента времени τ. Это доказывается путем использования последнего результата и на основании марковского свойства системы. Распределение выходящего потока в любой момент времени зависит только от состояния системы в момент времени t и не зависит от предыдущего состояния системы.

Рейч приводит следующую теорему, обратную данной.

Теорема 12.2. Если входящий и выходящий потоки одноканальной системы имеют пуассоновское распределение, то распределение времени обслуживания является экспоненциальным или время обслуживания с вероятностью 1 равно нулю.

Можно также показать, что если входящий поток имеет постоянную интенсивность поступления требований λ и если выходящий поток является пуассоновским, однородным при всех значениях параметра обслуживания μ, то и входящий поток является пуассоновским. Это следует из допущения об однородности потока и при $\mu\to\infty$. В этом случае время обслуживания равно нулю, и распределение выходящего потока совпадает с распределением входящего потока.

В многоканальной системе с постоянными промежутками времени между требованиями, покидающими систему, и постоянным временем обслуживания выходящий поток (если имеется очередь) также является постоянным с промежутками времени между моментами ухода требований, равными времени обслуживания. Экспоненциальному распределению промежутков времени между поступающими требованиями и экспоненциальному распределению времени обслуживания соответствует распределение по закону хи-квадрат с двумя степенями свободы, а постоянным промежуткам времени между моментами поступления требований и постоянному времени обслуживания соответствует распределение по закону хи-квадрат с бесконечным числом степеней свободы. Для промежуточного случая Рейч приводит следующее положение.

Теорема 12.3. Распределение промежутков времени между требованиями, покидающими одноканальную систему, в которой промежутки времени между поступающими требованиями и длительности обслуживания являются суммами двух случайных величин, имеющих экспоненциальное распределение с параметрами λ и μ, не является суммой двух случайных величин, имеющих экспоненциальное распределение.

Доказательство производится методом reductio ad absurdum. Если бы распределение промежутков времени τ между требованиями, покидающими систему, действительно являлось суммой двух случайных величин, имеющих экспоненциальное распределение, то одномерное распределение было бы то же самое, что и для входящего потока. Отсюда следует, что вероятность того, что после ухода требования в системе не остается других требований, равна $1-\frac{\lambda}{\mu}$. Обращаясь к работе О. А. Вольберга, рассмотренной в предыдущей главе, видим, что это невозможно.

Финч [245] показал, что для стационарного состояния одноканальной системы с пуассоновским входящим потоком, когда место для ожидания рассчитано на неограниченное число требований, следующие положения справедливы в том и только в том случае, если время обслуживания имеет экспоненциальное распределение:

1) число требований, находящихся в очереди после того, как требование покидает систему, не зависит (в пределе) от того, когда покинуло систему предыдущее требование;

2) два последовательных промежутка времени между требованиями, покидающими систему, взаимно независимы (в пределе).

12.3. Два последовательных канала

В случае неограниченного входящего пуассоновского потока с параметром λ, поступающего в первый из двух последовательных каналов, имеющих экспоненциальное время обслуживания с параметрами μ_1 и μ_2 соответственно (требование должно пройти через оба обслуживающих канала), имеем следующие уравнения для переходного состояния (вывод их производится обычным способом) [369]:

$$P'=(0,\ 0,\ t)=-\lambda P(0,\ 0,\ t)+\mu_2 P(0,\ 1,\ t),\quad n_1=n_2=0,\quad (12.8а)$$

$$P'(0,\ n_2,\ t)=-(\lambda+\mu_2)P(0,\ n_2,\ t)+\mu_1 P(1,\ n_2-1,\ t)+ \\ +\mu_2 P(0,\ n_2+1,\ t),\qquad n_1=0,\ n_2>0,\quad (12.8б)$$

$$P'(n_1,\ 0,\ t)=-(\lambda+\mu_1)P(n_1,\ 0,\ t)+\mu_2 P(n_1,\ 1,\ t)+ \\ +\lambda P(n_1-1,\ 0,\ t),\qquad n_1>0,\quad n_2=0.\quad (12.8в)$$

$$P'(n_1,n_2,t)=-(\lambda+\mu_1+\mu_2)P(n_1,n_2,t)+ \\ +\mu_1 P(n_1+1,n_2-1,t)+\mu_2 P(n_1,\ n_2+1,\ t)+ \\ +\lambda P(n_1-1,\ n_2,\ t),\quad n_1>0,\quad n_2>0.\quad (12.8г)$$

Здесь $P(n_1, n_2, t)$ — вероятность того, что в момент времени t в первой фазе имеется n_1 требований (включая и то, которое находится на обслуживании), а во второй фазе находится n_2 требований. Решение уравнений для стационарного состояния имеет вид

$$p(n_1, n_2) = \rho_1^{n_1}\rho_2^{n_2}p(0, 0),$$

$$\rho_1 = \frac{\lambda}{\mu_1}, \quad \rho_2 = \frac{\lambda}{\mu_2}. \tag{12.9}$$

Так как

$$\sum_{\substack{n_1=0\\n_2=0}}^{\infty} p(n_1, n_2) = 1, \tag{12.10}$$

то

$$p(0, 0) = (1-\rho_1)(1-\rho_2). \tag{12.11}$$

Математическое ожидание числа требований, находящихся в системе, можно вычислить, умножая $p(n_1, n_2)$ на n_1+n_2 и суммируя по n_1, а затем по n_2. Получим

$$L = \frac{\rho_1}{1-\rho_1} + \frac{\rho_2}{1-\rho_2}. \tag{12.12}$$

Упражнение. Проверьте, что (12.9) действительно является решением уравнений для стационарного состояния.

Упражнение. Получите (12.11) и (12.12).

Вероятность того, что в первой фазе находится n_1 требований, найдем, произведя суммирование по n_2. Получим $\rho_1^{n_1}(1-\rho_1)$. Аналогично, $\rho_2^{n_2}(1-\rho_2)$ — вероятность того, что во второй фазе находится n_2 требований. (Проверьте это.)

Если имеется N одинаковых источников входящего потока, и, следовательно, для каждого требования существует вероятность того, что за время Δt оно поступит в систему (равная $\lambda\Delta t$), а все остальное останется без изменений, то уравнения примут вид:

1. То же, что и (12.8а).
2. То же, что и (12.8б) при $n_1=0$; $n_2=1, \ldots, N-1$.
3. $P'(0, N, t) = -\mu_2 P(0, N, t) + \mu_1 P(1, N-1, t)$;
$n_1 = 0$; $n_2 = N$.
4. То же, что и (12.8в) при $n_1 = 1, \ldots, N-1$; $n_2 = 0$.
5. $P'(N, 0, t) = -\mu_1 P(N, 0, t) + \lambda P(N-1, 0, t)$;
$n_1 = N$; $n_2 = 0$.
6. То же, что и (12.8г) при $n_1 > 0$, $n_2 > 0$, $n_1 + n_2 < N$.
7. $P'(n_1, n_2, t) = -(\mu_1 + \mu_2) P(n_1, n_2, t) +$
$+ \mu_1 P(n_1+1, n_2-1, t) + \lambda P(n_1-1, n_2, t)$;

$$n_1 > 0,\ n_2 > 0,\ n_1 + n_2 = N.$$

И в этом случае решение для стационарного состояния имеет вид

$$p(n_1, n_2) = \rho_1^{n_1}\rho_2^{n_2}p(0, 0). \tag{12.13}$$

Просуммировав по n_1 и n_2 от нуля до N $(n_1+n_2 \leqslant N)$ и приравняв результат единице, получим

$$p(0, 0) = \frac{(\rho_1-\rho_2)(1-\rho_1)(1-\rho_2)}{(\rho_1-\rho_2)-\left(\rho_1^{N+2}-\rho_2^{N+2}\right)+\rho_1\rho_2\left(\rho_1^{N+1}-\rho_2^{N+1}\right)}. \qquad (12.14)$$

Среднее число требований, находящихся в системе при $n_1 + n_2 \leqslant N$, равно

$$L = \sum_{n_1=0}^{N}\sum_{n_2=0}^{N}(n_1+n_2)\,p(n_1, n_2) =$$

$$= \frac{1}{\rho_1-\rho_2}\left\{\frac{\rho_1^2\left[1-(N+1)\rho_1^N+N\rho_1^{N+1}\right]}{(1-\rho_1)^2} - \right.$$

$$\left. - \frac{\rho_2^2\left[1-(N+1)\rho_2^N+N\rho_2^{N+1}\right]}{(1-\rho_2)^2}\right\}p(0, 0). \qquad (12.15)$$

Среднее число обслуженных требований равно

$$\sum_{n_1=1}^{N} p(n_1, 0) + \sum_{n_2=1}^{N} p(0, n_2) + 2\sum_{n_1=1}^{N}\sum_{n_2=1}^{N} p(n_1, n_2) =$$

$$= \frac{(\rho_1+\rho_2)\left[(\rho_1-\rho_2)-\left(\rho_1^{N+1}-\rho_2^{N+1}\right)+\rho_1\rho_2\left(\rho_1^N-\rho_2^N\right)\right]p(0, 0)}{(1-\rho_1)(1-\rho_2)(\rho_1-\rho_2)}. \qquad (12.16)$$

Среднее число ожидающих требований равно

$$\sum_{n_1=2}^{N}(n_1-1)p(n_1, 0) + \sum_{n_2=2}^{N}(n_2-1)p(0, n_2) + \sum_{n_1=2}^{N}(n_1-1)p(n_1, 1) +$$

$$+ \sum_{n_2=2}^{N-1}(n_2-1)p(1, n_2) + \sum_{\substack{n_1=2\\ n_1+n_2 \leqslant N}}^{N-2}\sum_{n_2=2}^{N-2}(n_1+n_2-2)p(n_1, n_2) =$$

$$= \left\{\frac{\rho_2^2(1+\rho_1)\left[1-N\rho_1^{N-1}+(N-1)\rho_1^N\right]}{(1-\rho_1)^2} + \right.$$

$$+ \frac{\rho_1^2(1+\rho_2)\left[1-N\rho_2^{N-1}+(N-1)\rho_2^N\right]}{\left(1-\rho_2\right)^2} +$$

$$+ \frac{\rho_1\rho_2}{\rho_1-\rho_2}\left[\frac{1-(N-1)\rho_1^{N-2}+(N-2)\rho_1^{N-1}}{(1-\rho_1)^2} - \right.$$

$$\left.\left. - \frac{1-(N-1)\rho_2^{N-2}-(N-2)\rho_2^{N-1}}{(1-\rho_2)^2}\right]\right\}p(0, 0). \qquad (12.17)$$

Теперь эти положения можно применить к трехфазной системе при неограниченном источнике входящего потока.

Упражнение. Дайте вывод формулы (12.13) и проверьте все остальные выражения.

12.4. Многоканальная система с последовательными и параллельными каналами

Изложенные выше результаты можно распространить на k последовательных фаз с очередью перед каждой фазой и неограниченным пуассоновским потоком, поступающим в каждую фазу. i-я фаза, где i — любое число, состоит из r_i параллельных каналов с экспоненциальным временем обслуживания с параметром μ_i [таким образом, вероятность того, что за время Δt будет закончено обслуживание одного из n_i требований, находящихся в i-й фазе, равна $n_i\mu_i\Delta t + o(\Delta t)$ при $n_i < r_i$ и равна $r_i\mu_i\Delta t + o(\Delta t)$ при $n_i \geqslant r_i$]. При $\rho_i = \frac{\lambda}{r_i\mu_i} < 1$ $(i = 1, \ldots, k)$ уравнения для стационарного состояния в обозначениях Р. Джексона [370] имеют вид

$$\left[\lambda + \sum_{j=1}^{k} \delta(n_j)\, a(n_j)\, \mu_j\right] p(n_1, \ldots, n_k) =$$
$$= \sum_{j=1}^{k} \delta(n_j + 1)\, a(n_j + 1)\, \mu_j p(n_1, \ldots, n_j + 1,\ n_{j+1} - 1,$$
$$n_{j+2}, \ldots, n_k) + \lambda p(n_1 - 1,\ n_2, \ldots, n_k). \qquad (12.18)$$

Если любой из аргументов является отрицательным, то вероятность равна нулю и последняя сумма не содержит членов с -1 и содержит один член с $+1$. Кроме того,

$$a(n_j) = \begin{cases} n_j, & n_j < r_j, \\ r_j, & n_j \geqslant r_j, \end{cases} \qquad (12.19)$$

$$\delta(n_j) = \begin{cases} 1, & n_j \neq 0, \\ 0, & n_j = 0, \end{cases} \quad j = 1,\ 2, \ldots, k. \qquad (12.20)$$

Решение имеет вид

$$p(n_1, \ldots, n_k) = p(0, \ldots, 0) \prod_{j=1}^{k} b(n_j), \qquad (12.21)$$

$$b(n_j) = \begin{cases} \dfrac{1}{n_j!}(r_j\rho_j)^{n_j}, & n_j < r_j, \\[2mm] \dfrac{1}{r_j!}(r_j\rho_j)^{r_j}(\rho_j)^{n_j - r_j}, & n_j \geqslant r_j. \end{cases} \qquad (12.22)$$

Так как сумма по всем значениям n_j должна равняться единице, то получим

$$\sum_{n_1=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} \left[\prod_{j=1}^{k} b(n_j)\right] = \prod_{j=1}^{k} \sum_{j=0}^{\infty} b(n_j) \equiv \prod_{j=1}^{k} A_j, \quad j = 1, \ldots, k, \qquad (12.23)$$

и

$$p(0, \ldots, 0) = \prod_{j=1}^{k} A_j^{-1}. \qquad (12.24)$$

Р. Джексон приводит также доказательство единственности этого решения. Распределение вероятностей для любой фазы находится путем суммирования по числу требований, находящихся во всех других фазах. Отсюда вероятность того, что в j-й фазе имеется n_j требований, равна

$$p(n_j) = \frac{b(n_j)}{A_j}.$$

(Чтобы найти вероятность того, что в j-й фазе находится n требований, положим $n_j = n$.) Этот результат можно получить также, решая задачу для системы с c параллельными каналами при $c = r_j$.

Положив $r_j = 1$ $(j = 1, \ldots, k)$, для одного из последовательных каналов получим

$$p(n_1, \ldots, n_k) = p(0, \ldots, 0) \prod_{j=1}^{k} \rho_j^{n_j}, \qquad (12.25)$$

$$p(0, \ldots, 0) = \prod_{j=1}^{k} (1 - \rho_j). \qquad (12.26)$$

Вследствие того что фазы взаимно независимы, вероятность того, что в j-й фазе находится n требований, равна

$$\rho_j^n (1 - \rho_j). \qquad (12.27)$$

Среднее число требований, находящихся в этой фазе, составляет

$$\sum_{n=0}^{\infty} n\rho_j^n (1 - \rho_j) = \frac{\rho_j}{1 - \rho_j}. \qquad (12.28)$$

Среднее число требований, находящихся на обслуживании в j-й фазе, равно

$$\sum_{n=1}^{\infty} \rho_j^n (1 - \rho_j) = \rho_j. \qquad (12.29)$$

Среднее число требований, ожидающих начала обслуживания в j-й фазе, равно

$$\rho_j^2 (1 - \rho_j)^{-1}. \qquad (12.30)$$

Математическое ожидание числа требований, находящихся в системе, равно сумме математических ожиданий числа требований, находящихся в фазах.

Распределение времени ожидания для клиента, поступающего из $(j-1)$-й фазы в j-ю фазу, как показано для одноканальной системы, имеет вид

$$f_j(\xi)\,d\xi=\sum_{n=1}^{\infty}(1-\rho_j)\rho_j^n\mu_j^n\xi^{(n-1)}\frac{e^{-\mu_j\xi}}{(n-1)!}\,d\xi=$$

$$=\lambda(1-\rho_j)\,e^{-(\mu_j-\lambda)\xi}d\xi. \tag{12.31}$$

Вероятность отсутствия ожидания в j-й фазе равна $1-\rho_j$, а вероятность ожидания, когда система занята, —

$$(\mu_j-\lambda)\,e^{-(\mu_j-\lambda)\xi}d\xi. \tag{12.32}$$

При $\mu_j=\mu$ положим $\sum n_j=n$. Пусть $p(n)$ — вероятность того, что в системе ожидает n требований. Тогда

$$p(n)=\binom{n+k-1}{k-1}\rho^n(1-\rho)^k, \tag{12.33}$$

где $\rho=\dfrac{\lambda}{\mu}$.

Упражнение. Убедитесь в справедливости полученных уравнений для стационарного состояния и проверьте все остальные формулы этого параграфа.

12.5. Нерегулярный процесс последовательного обслуживания

Если в j-ю фазу (с r_j параллельными каналами) допускается также внешний пуассоновский поток с параметром λ_j, а после окончания обслуживания в этой фазе с вероятностью q_{jm} требования поступают в m-ю фазу и затем обслуживание производится в порядке поступления, то интенсивность поступления требований в j-ю фазу равна

$$\overline{\lambda}_j=\lambda_j+\sum_m q_{mj}\overline{\lambda}_m. \tag{12.34}$$

Параметр $\overline{\lambda}_j$ играет в j-й фазе ту же самую роль, что и параметр λ в предыдущем случае. Имеем

$$p(n_1,\ldots,n_k)=p^1_{n_1}p^2_{n_2}\ldots p^k_{n_k}, \tag{12.35}$$

где

$$\overline{\lambda}_j<\mu_j,\quad j=1,\ldots,k,$$

а вероятность того, что в j-й фазе находится n_j требований, равна

$$p_{n_j}^j = \begin{cases} \dfrac{p_0^j \left(\dfrac{\bar{\lambda}_j}{\mu_j}\right)^{n_j}}{n_j!}, & n_j = 0, 1, \ldots, r_j, \\ \dfrac{p_0^j \left(\dfrac{\bar{\lambda}_j}{\mu_j}\right)^{n_j}}{r_j! \, (r_j)^{n_j - r_j}}, & n_j = r_j, \ldots \end{cases} \tag{12.36}$$

Постоянная p_0^j находится из соотношения $\sum\limits_{n_j} p_{n_j}^j = 1$. Этот результат основан на том, что распределение числа требований, находящихся в каждой фазе, не зависит от распределения числа требований, находящихся в любой другой фазе. Дж. Джексон [366] написал уравнения для неустановившегося состояния и, переходя к стационарному состоянию, получил эту формулу.

12.6. Система, в которой не разрешается очередь перед последовательными фазами

В данном случае имеется две фазы. В первую фазу поступает пуассоновский входящий поток с параметром λ. В обеих фазах время обслуживания имеет одинаковое экспоненциальное распределение с параметром μ, обслуживание требований производится по принципу «первым пришел — первым обслужен». Выходящий поток первой фазы является входящим потоком для второй фазы, и образование очереди перед второй фазой не разрешается, в то время как перед первой фазой допускается очередь бесконечной длины. Это приводит к запиранию первого канала (требования не поступают на обслуживание) даже в том случае, если обслуживание требования окончилось и имеется очередь. Первый канал открывается для обслуживания только в том случае, если второй канал освобождается, и данное требование может перейти на обслуживание из первого канала во второй. Значение загрузки $\rho = \frac{\lambda}{\mu}$ в этом случае уменьшается, так как эффективная интенсивность обслуживания должна умножаться на ту долю времени, в течение которого первый канал не закрыт.

Пусть $P(., ., t)$ — вероятность того, что в момент времени t в первой фазе находится n требований (включая и обслуживаемое).

1. Положим в $P(., ., t)$ первый аргумент равным нулю, если требование все еще обслуживается в первом канале, или единице, если обслуживание его закончено, но оно задерживается, так как

второй канал занят; следовательно, второй аргумент также полагается равным единице.

2. Если второй канал освобождается, то второй аргумент заменяется нулем.

Заметим, что нулевой индекс обозначает отсутствие ожидающих требований или отсутствие обслуживаемых требований в первой фазе; таким образом, в этом случае ясно, что если первый аргумент заменен нулем, то первый обслуживающий канал свободен.

При $n=0$ имеем

$$P_0'(0, 0, t) = -\lambda P_0(0, 0, t) + \mu P_0(0, 1, t),$$

$$P_0'(0, 1, t) = -(\lambda+\mu) P_0(0, 1, t) + \mu P_0(1, 1, t) + \mu P_1(0, 0, t),$$

$$P_0'(1, 1, t) = -(\lambda+\mu) P_0(1, 1, t) + \mu P_1(0, 1, t). \qquad (12.37)$$

При $n \geqslant 1$

$$P_n'(0, 0, t) = -(\lambda+\mu) P_n(0, 0, t) + \lambda P_{n-1}(0, 0, t) + \mu P_n(0, 1, t),$$

$$\begin{aligned} P_n'(0, 1, t) = -(\lambda+2\mu) P_n(0, 1, t) + \lambda P_{n-1}(0, 1, t) + \\ + \mu P_n(1, 1, t) + \mu P_{n+1}(0, 0, t), \end{aligned} \qquad (12.38)$$

$$P_n'(1, 1, t) = -(\lambda+\mu) P_n(1, 1, t) + \lambda P_{n-1}(1, 1, t) + \mu P_{n+1}(0, 1, t).$$

Здесь μ — интенсивность обслуживания, а λ — интенсивность поступления требований.

Приравняв левые части этих уравнений нулю, можно получить уравнения для стационарного состояния. Перейдя к стационарному состоянию, введем операцию сдвига $Ep_n = p_{n+1}$. Решив эту систему уравнений, найдем, что определитель равен нулю [361], и получим уравнение, корни которого равны

$$r_1 = 1, \quad r_2 = \frac{\rho}{\rho+1}, \quad r_{3,4} = \frac{\rho}{4}\left[(\rho+3) \pm \sqrt{\rho^2+6\rho+1}\right], \qquad (12.39)$$

где $\rho = \frac{\lambda}{\mu}$ — загрузка. Так как вероятность $p_n(.,.)$ того, что в стационарном состоянии в системе находится n требований, есть линейная комбинация n-х степеней этих корней, то для обеспечения сходимости нужно отбросить все корни, значение которых больше единицы, и все значения ρ, которые приводят к корням, большим единицы.

Критическое значение ρ, при котором можно найти эти вероятности, есть максимально допустимая загрузка. Таким образом, отбрасываем r_1, а при $\rho \geqslant \frac{2}{3}$ нужно отбросить также и r_3. Поэтому $\rho = \frac{2}{3}$ и есть максимально допустимая загрузка. Общее решение для каждого значения вероятности стационарного состояния при $\rho < \frac{2}{3}$ и $n \geqslant 4$ имеет вид

$$p_n(0, 0) = \sum_{j=2}^{4} c_{1j} r_j^n, \quad p_n(0, 1) = \sum_{j=2}^{4} c_{2j} r_j^n,$$
$$p_n(1, 1) = \sum_{j=2}^{4} c_{3j} r_j^n. \qquad (12.40)$$

При $n<4$ результат находится непосредственно из уравнений для стационарного состояния. Вероятность $p_0(0, 0)$ вычисляется с помощью выражения

$$\sum_{n=0}^{\infty} [p_n(0, 0) + p_n(0, 1) + p_n(1, 1)] = 1. \qquad (12.41)$$

Математическое ожидание числа требований, находящихся в системе, равно

$$L = \sum_{n=0}^{\infty} [np_n(0, 0) + (n+1)p_n(0, 1) + (n+2)p_n(1, 1). \qquad (12.42)$$

Простой способ вычисления максимально допустимой загрузки для двухфазной системы заключается в определении величин

$$Q(0, 0, t) = \sum_{n=0}^{\infty} P_n(0, 0, t), \qquad Q(1, 1, t) = \sum_{n=0}^{\infty} P_n(1, 1, t),$$
$$Q(0, 1, t) = \sum_{n=0}^{\infty} P_n(0, 1, t).$$

Принимая, что интенсивность обслуживания в первой фазе равна μ_1, а во второй фазе составляет μ_2, получим

$$Q'(0, 0, t) = -\mu_1 Q(0, 0, t) + \mu_2 Q(0, 1, t) + \mu_1 P_0(0, 0, t),$$
$$Q'(0, 1, t) = -(\mu_1 + \mu_2) Q(0, 1, t) + \mu_1 Q(0, 0, t) +$$
$$+ \mu_2 Q(1, 1, t) - \mu_1 P_0(0, 0, t) + \mu_2 P_0(0, 1, t), \qquad (12.43)$$
$$Q'(1, 1, t) = -\mu_2 Q(1, 1, t) + \mu_1 Q(0, 1, t) - \mu_1 P_0(0, 1, t).$$

Пусть $\mu_1 = \mu_2 = \mu$. Заметим, что, перейдя к стационарному состоянию, при увеличении коэффициента использования можно пренебречь вероятностями $P_0(0, 0, t)$ и $P_0(0, 1, t)$. Таким образом, все вероятности

$$\lim_{t \to \infty} Q(., ., t) \equiv q(., .) \qquad (12.43)$$

равны между собой.

Отсюда найдем, что доля времени, в течение которого обслуживающее устройство первой фазы свободно, составляет

$$\frac{q(0, 0) + q(0, 1)}{q(0, 0) + q(0, 1) + q(1, 1)} = \frac{2}{3}.$$

Найденное значение максимально допустимой загрузки является таким же, что и полученное выше при рассмотрении корней.

В случае наличия k фаз в первой фазе происходит более длительная задержка, чем в любой другой, и максимально допустимая загрузка для первой фазы, обозначаемая через $\rho_{\text{макс}}$, является максимально допустимой загрузкой также для всей системы.

Для последнего случая при $\mu_1 \neq \mu_2$ имеем

$$\rho_{\text{макс}} = \frac{\mu_2(\mu_1 + \mu_2)}{\mu_1^2 + \mu_1\mu_2 + \mu_2^2}. \tag{12.45}$$

Предельное значение интенсивности потока требований, которая может допускаться в стационарном состоянии, равно $\lambda_{\text{макс}} = \mu_1\rho_{\text{макс}}$. Применяя этот метод к трехфазной системе, в которой интенсивность обслуживания во всех фазах одинакова, получим

$$\rho_{\text{макс}} = \frac{22}{39} \approx 0{,}5641.$$

При этом рассматриваются вероятности восьми состояний. Для четырех фаз рассматриваются вероятности 21 состояния, и аналогично находим

$$\rho_{\text{макс}} = 0{,}5115.$$

Если во второй фазе разрешается конечная очередь, содержащая $q-1$ требований, то имеется $q+2$ состояний и

$$\rho_{\text{макс}} = \frac{\mu_2\left(\mu_1^{q+1} - \mu_2^{q+1}\right)}{\mu_1^{q+2} - \mu_2^{q+2}}. \tag{12.46}$$

При $\mu_1 = \mu_2$ имеем $\rho_{\text{макс}} = \frac{q+1}{q+2}$. Если во второй и третьей фазах разрешается очередь, в которой может находиться одно требование, то получим пятнадцать вероятностей состояний. Если интенсивности обслуживания во всех фазах одинаковы, то

$$\rho_{\text{макс}} = \frac{8529}{12\,721} \approx 0{,}6705.$$

Для случая, когда перед первой фазой разрешается бесконечная очередь, а перед остальными $k-1$ фазами очередь отсутствует и не допускается наличие свободных обслуживающих устройств (т. е. требования ожидают в обслуживающем канале до тех пор, пока следующий канал не станет свободным), имеем

$$\rho_{\text{макс}} = \frac{1}{\mu\tau_k}.$$

Здесь τ_k — математическое ожидание вероятности того, что в момент времени t после начала обслуживание во всех k фазах будет закончено, если интенсивность обслуживания во всех фазах одинакова:

$$\tau_k = k\int_0^\infty t\mu e^{-\mu t}(1-e^{-\mu t})^{k-1}dt = -k\left[\frac{dB\left(\frac{p}{\mu}, N\right)}{dp}\right]\Bigg|_{p=\mu}. \qquad (12.47)$$

Последнее выражение в правой части найдем, применяя преобразование Лапласа. Здесь B — бета-функция.

Распределение, записанное в подынтегральном выражении, является обобщением выражения, полученного для двух фаз, где доказывается, что вероятность завершения обслуживания во всех фазах равна сумме двух одинаковых выражений. Каждое выражение равно произведению двух величин, первая из которых — вероятность того, что в промежутке времени $(t, t+dt)$ закончится обслуживание в одной из фаз, а вторая — вероятность того, что во второй фазе обслуживание закончится не позднее момента времени t.

Для трехфазной системы $\rho_{\text{макс}}=\frac{6}{11}$, а для четырех фаз $\rho_{\text{макс}}=\frac{12}{25}$. Этот метод применим и в том случае, когда интенсивности обслуживания различны.

12.7. Некоторые формулы для времени ожидания

Допустим, что

$$P(x_j > r) = K_j e^{c_j r} \qquad (12.48)$$

есть вероятность того, что время ожидания в j-й фазе, состоящей из r_j параллельных каналов, не больше t. Пусть в этом выражении [см. (12.31)]

$$c_j = -r_j\mu_j(1-\rho_j), \qquad (12.49)$$

$$K_j = P(x_j > 0). \qquad (12.50)$$

Можно показать, что в этом случае функция распределения времени ожидания во всех k фазах имеет вид

$$P\left(\sum_{j=1}^{k} x_j \leqslant \tau\right) = 1 - \sum_{j=1}^{k} A_{jk} e^{c_j\tau}, \qquad (12.51)$$

где

$$A_{jk} = K_j\left(\prod_{\substack{i=1\\ j\neq j}}^{k}\frac{1-K_i c_j}{c_j - c_i}\right), \quad j=1,\ldots,k, \qquad (12.52)$$

и $c_i \neq c_j$ при $i \neq j$.

Доказательство производится по индукции. Принимается результат при $y=\sum_{j=1}^{k-1} x_j$ и доказывается его справедливость для k. Таким образом,

$$P(y+x_k\leqslant\tau)=P(y=0,\ x_k=0)+P(y=0,\ 0<x_k\leqslant\tau)+ \\ +P(0<y\leqslant\tau,\ x_k=0)+P(0<y\leqslant\tau,\ 0<x_k\leqslant\tau). \quad (12.53)$$

Конечную формулу получим, заменив каждую величину ее эквивалентом.

Нелсон [604], рассматривая число параллельных каналов в каждой фазе, вычислил каждое значение c_j и каждое значение K_j, использовал эти результаты для нахождения вероятности того, что в каждом из нескольких параллельных каналов (предполагаемых на основании работ Бёрка и Рейча независимыми) время ожидания больше нуля, а затем с помощью приведенной выше формулы вычислил A_{jk}. В общем виде приводится метод вычисления для случая $c=c_j$ при некоторых i. Каждая из двух величин изменяется на малую величину Δ, которая прибавляется к одной величине и вычитается из другой.

Дебон и Кац [170] использовали распределение по закону хи-квадрат для аппроксимации суммы нескольких экспоненциальных распределений путем приравнивания математических ожиданий и дисперсий математическим ожиданиям и дисперсиям плотности распределения, полученной из функции распределения (12.51). Этот метод находит важные приложения; произведена достаточно тщательная проверка его применимости [745].

Замечание 1. Сакс [751] рассмотрел проблему эргодичности систем массового обслуживания с последовательными каналами. Если в системе с двумя последовательными каналами время ожидания n-го требования в первой очереди равно w_n, а время ожидания во второй очереди равно w_n^* (требование должно поступить в оба канала) и совместное распределение $(w_n,\ w_n^*)$ при $n\to\infty$ стремится к собственному распределению вероятностей, то говорят, что система массового обслуживания является эргодической.

Замечание 2. Коэн [137] исследовал поступление пуассоновского потока в несколько параллельных каналов, выходящий поток которых направляется в другую группу параллельных каналов и т. д. Рассматривается несколько таких параллельных систем. В некоторой точке выходящий поток всех этих групп является входящим потоком для одноканальной системы.

Глава 13

НЕКОТОРЫЕ ИНТЕРЕСНЫЕ АСПЕКТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

13.1. Введение

В этой главе будут рассмотрены некоторые интересные вопросы, связанные со структурой систем массового обслуживания. В первой главе уже упоминалось о различных явлениях, которые могут наблюдаться при массовом обслуживании. Здесь будут проанализированы некоторые из этих явлений иногда кратко, а в некоторых случаях достаточно подробно. Попытаемся установить некоторую последовательность и начнем с тех явлений, которые влияют на входящий поток, на очередь, на обслуживающее устройство, а затем и на выходящий поток.

Например, клиент может отказаться стать в очередь или может уйти из очереди до начала обслуживания. Кроме того, до начала обслуживания клиенту может потребоваться некоторая информация. Клиенты могут опаздывать, даже если время прибытия назначено расписанием. Если имеется несколько очередей, то клиенты могут непрерывно переходить из одной очереди в другую. Число каналов может изменяться при изменении длины очереди. Каналы могут объединяться для обслуживания различных потребностей одного клиента. Они могут производить и специальное обслуживание. Кроме того, каналы могут иметь различную интенсивность обслуживания. Может иметься единственное обслуживающее устройство, которое перемещается вместе с линией обслуживания. Наконец, обслуженные клиенты могут возвращаться в очередь за дополнительным обслуживанием.

Рассмотрение здесь этих явлений отнюдь не является исчерпывающим. Оно лишь имеет цель вызвать интерес к различным вопросам. Для начала укажем на последнюю работу, посвященную случаю, когда клиенты обращаются за основным и дополнительным обслуживанием. Свенссон [790] рассмотрел несколько интересных задач, связанных с ожиданием в очереди. Конечное число клиентов обращается за обслуживанием в систему с несколькими обслуживающими устройствами. Если свободных устройств нет, то клиент ожидает, пока одно из них не освободится. Каждый клиент может обращаться за дополнительным обслуживанием или в то время, когда он обслуживается, или когда ожидает обслуживания. Как поступление требований на основное обслуживание, так и поступление требований на дополнительное обслуживание имеет пуассоновское распределение. Автор предполагает, что интенсивности поступления требований на основное обслуживание и поступления требований на два вида дополнительного обслуживания различны, но исследует главным образом частные случаи, когда по крайней мере два потока из трех имеют одинаковую интенсивность. В рассматриваемом частном случае интенсивности основного и дополнительного обслуживания равны между собой. Исследуется два вида процесса обслуживания: один, когда требования обслуживаются в порядке поступления, и другой, при котором обслуживание производится в порядке поступления требований на основное обслуживание. Для обоих случаев получены приближенные решения.

13.2. Случай, когда клиенты отказываются становиться в очередь

Сразу же после поступления в систему клиент может принять решение не становиться в очередь, либо найдя ее слишком длинной, либо основываясь на информации относительно длительности обслуживания. Клиент может уходить из очереди (т. е. он может оказаться нетерпеливым и покинуть систему необслуженным) после присоединения к очереди, если он считает, что время ожидания будет больше приемлемого для него. Последнее явление наиболее вероятно для систем, в которых дисциплиной очереди является случайный выбор на обслуживание. Оно изучается в следующем параграфе.

1. Первый вариант системы, в которой клиенты отказываются становиться в очередь

Рассмотрим одноканальную систему с пуассоновским входящим потоком с параметром λ и экспоненциальным временем обслуживания с параметром μ.

Обозначим через $P_n(t)$ вероятность того, что в момент времени t в очереди находится не более n клиентов. Следует заметить, что здесь $P_n(t)$ получает новый смысл. Пусть K — наибольшая длина очереди, при которой прибывший клиент не покидает систему. Она имеет одно и то же распределение для всех поступающих требований. Пусть $F(n)=P(K\leqslant n)$ — вероятность того, что клиент отказывается становиться в очередь, если в ней находится n клиентов. Таким образом, клиент становится в очередь, если длина очереди меньше той, при которой он покинул бы систему. Вероятность этого события равна $1-F(n-1)$. Введем обозначения $G(n)=1-F(n)$, $f(n)=F(n)-F(n-1)$ и $p_n(t)=P_n(t)-P_{n-1}(t)$. Обычно $p_0(t)=P_0(t)\neq 0$ и $f(0)=$

$=F(0)\neq 0$ — это выражения для вероятности отсутствия очереди в момент времени t и вероятности того, что клиент определенно относится к категории неприсоединяющихся к очереди. Уравнения имеют вид

$$p_n(t+\Delta t)=[1-(\lambda+\mu)\Delta t]\,p_n(t)+\mu p_{n+1}(t)\Delta t+ \\ +\lambda p_{n-1}(t)\,G(n-2)\Delta t+\lambda p_n(t)\,F(n-1)\Delta t+O(\Delta t)^2. \quad (13.1)$$

Последние два члена перед $O(\Delta t)^2$ выражают вероятность того, что в течение времени Δt прибывает клиент, который становится в очередь (первое выражение) или покидает систему (второе выражение).

Как и обычно, эту систему можно привести к дифференциально-разностному уравнению, которое удобно распространить и на случай $n=0$, положив $F(-1)=F(-2)=0$.

Просуммировав эти уравнения по n, получим

$$\frac{dP_n(t)}{dt}=\mu p_{n+1}(t)-\lambda p_n(t)\,G(n-1), \quad (13.2)$$

или, положив $\lambda G(n)\equiv\lambda_n$, это уравнение можно написать в следующем виде:

$$\frac{dP_n(t)}{dt}=\lambda_{n-1}P_{n-1}(t)-(\mu+\lambda_{n-1})\,P_n(t)+\mu P_{n+1}(t). \quad (13.3)$$

Если $F(n)=0$ для всех n, то (13.3) превращается в уравнение для частного случая процесса рождения и гибели.

Решение уравнения (13.2) для стационарного состояния имеет вид

$$p_1=\rho p_0, \quad p_{n+1}=\rho G(n-1)\,p_n,$$
$$p_n=\rho^n c_n p_0. \quad (13.4)$$

Здесь

$$\rho=\frac{\lambda}{\mu}, \quad c_0=c_1=1, \quad c_n=\prod_{i=0}^{n-2} G(i), \quad i=2, 3, \ldots, \quad (13.5)$$

а p_0 определяется из соотношения $\sum\limits_{n=0}^{\infty} p_n=1$. Можно показать [319], что сходимость этого ряда является необходимым и достаточным условием существования равновесного состояния.

Заметим, что если $p_n\neq 0$, а $p_{n+1}=0$, то и $G(n-1)=0$, следовательно, при $m>n$ имеем $p_m=0$, т. е. прибывающие клиенты не становятся в очередь, если в очереди находится n или большее число клиентов. Кроме того, если $p_n\neq 0$ и $p_{n+1}\neq 0$, то

$$f(n)=G(n-1)-G(n)=\frac{p_0}{p_1}\left(\frac{p_{n+1}}{p_n}-\frac{p_{n+2}}{p_{n+1}}\right) \quad (13.6)$$

и

$$p_{n+1}^2\geqslant p_n p_{n+2}. \quad (13.7)$$

Любое распределение неограниченной очереди, удовлетворяющее последнему неравенству, соответствует распределению вероятностей того, что прибывающий клиент отказывается становиться в очередь.

Чтобы найти L и дисперсию, введем производящую функцию для (13.2), перейдем к стационарному состоянию и свяжем эти результаты с полученным ранее решением:

$$P(z, t)=\sum_{n=0}^{\infty} p_n(t) z^n=(1-z) \sum_{n=0}^{\infty} P_n(t) z^n \tag{13.8}$$

и

$$Q(z, t)=\sum_{n=0}^{\infty} p_{n+1}(t) F(n) z^n. \tag{13.9}$$

Умножая (13.2) на z^n и суммируя, получим уравнение

$$\frac{\partial}{\partial t} \frac{P(z, t)}{1-z}=-\lambda p_0(t)+$$
$$+\left(\frac{\mu}{z}-\lambda\right)\left[P(z, t)-p_0(t)\right]+\lambda z Q(z, t). \tag{13.10}$$

При стационарном состоянии (т. е. при $t \to \infty$) это уравнение имеет вид

$$P(z)-\rho z P(z)-p_0+\rho z^2 Q(z)=0. \tag{13.11}$$

Отсюда при $z=1$

$$P(0) \equiv p_0=1-\rho+\rho Q(1). \tag{13.12}$$

Так как $0<p_0 \leqslant 1$, то это означает, что $\frac{(\rho-1)}{\rho}<Q(1) \leqslant 1$. Эффективное значение загрузки для тех клиентов, которые становятся в очередь, равно

$$\rho'=\rho-\rho Q(1), \tag{13.13}$$

так как величина ρ была вычислена для всех клиентов, — как тех, которые становятся в очередь, так и тех, которые покидают очередь.

Основываясь на решении, полученном для стационарного состояния, и определении $P(z)$, получим

$$P(z)=p_0 \sum_{n=0}^{\infty} \rho^n c_n z^n. \tag{13.14}$$

Теперь, используя (13.11), можно найти $Q(z)$ в явном виде. Таким образом, поскольку p_0 известно, то распределение числа клиентов, находящихся в очереди, и распределение вероятностей того, что прибывающий клиент отказывается становиться в очередь, определяются однозначно.

Итак,

$$\rho \frac{\partial}{\partial\rho} \log p_n =$$

$$= \rho \frac{\partial}{\partial\rho} \left(\log c_n + n \log \rho - \log \sum_{n=0}^{\infty} \rho^n c_n \right) = n - L, \quad (13.15)$$

где L — математическое ожидание числа клиентов, находящихся в очереди.

Легко показать, что дисперсия длины очереди равна

$$\sum_{n=0}^{\infty} (n-L)^2 p_n = \sum_{n=0}^{\infty} (n-L) \rho \frac{\partial p_n}{\partial\rho} = \rho \frac{\partial}{\partial\rho} L. \quad (13.16)$$

Из выражений для $P(z)$ и $Q(z)$ находим

$$\left(\rho \frac{\partial}{\partial\rho} - z \frac{\partial}{\partial z} \right) \begin{cases} \log P(z) \\ \log Q(z) \end{cases} = \begin{cases} -L, \\ 1 - L. \end{cases} \quad (13.17)$$

Дифференцируя (13.11) по z и положив $z=0$, найдем математическое ожидание, а после вторичного дифференцирования при $z=0$ найдем дисперсию:

$$L = \frac{\rho}{1-\rho} [1 - Q'(1) - 2Q(1)], \quad (13.18)$$

$$\sigma^2 = \frac{\rho}{(1-\rho)^2} \{1 - \rho [Q'^2(1) + 4Q^2(1) + 4Q(1) Q'(1)] -$$
$$- (1-\rho) [Q''(1) + 5Q'(1) + 4Q(1)]\}. \quad (13.19)$$

Хейт применил эти положения к распределениям, удовлетворяющим условию (13.7).

Кратко проиллюстрируем некоторые из ранее приведенных результатов на следующем примере.

Пример. Для биномиального распределения имеем

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n,$$

$0 < p < 1$. Тогда $\rho = \frac{np}{(1-\rho)}$ и $L = \frac{np}{(n+p)} \to n$ при $\rho \to \infty$. Заметим, что, подставив в (13.4) значение p_0 и $c_1 = 1$, найдем ρ. Распределение вероятностей того, что прибывающий клиент не становится в очередь, имеет вид

$$G(k) = \begin{cases} \frac{n-k-1}{n(k+2)}, & 0 \leqslant k \leqslant n-1, \\ 0, & n \leqslant k, \end{cases}$$

$$f(n) = \begin{cases} \frac{n+1}{n} \frac{1}{(k+1)(k+2)}, & 0 \leqslant k \leqslant n-1, \\ 0, & n \leqslant k, \end{cases}$$

и не зависит от p. Выражая p через ρ, получим

$$p_0=\left(\frac{n}{n+\rho}\right)^n, \quad P(z)=\left(\frac{n+\rho z}{n+\rho}\right)^n, \quad \sigma^2=\left(\frac{n}{n+\rho}\right)^2.$$

Упражнения. Читатель может теперь проверить эти положения для следующих распределений.

1. Отрицательное биномиальное:

$$p_k=\binom{N+k-1}{N-1}x^k(1+x)^{-N-k}, \quad k=0, 1, \ldots,$$

где $x>0$, $N>1$.

2. Пуассоновское:

$$p_k=\frac{\rho^k e^{-\rho}}{k!}, \quad k=0, 1, \ldots$$

3. Распределение

$$p_k=A(k+a)^v e^{-\lambda k},$$

где

$$a>0, \quad v>0, \quad \lambda>0,$$

а A является функцией этих величин.

4. Нормальное распределение:

$$p_k=Ae^{-\frac{(k-m)^2}{2v}},$$

где m и v являются приближенными значениями математического ожидания и дисперсии, если v и $\frac{m^2}{v}$ велики (т. е. больше девяти).

5. Детерминированные отказы, когда для всех клиентов существует одна и та же максимальная длина очереди, при которой прибывающие клиенты отказываются становиться в очередь:

$$f(n)=\begin{cases}1 & \text{при} \quad n=K,\\ 0 & \text{при} \quad n\neq K.\end{cases}$$

В этом случае

$$G(n)=\begin{cases}1, & 0\leqslant n\leqslant K-1,\\ 0, & K\leqslant n,\end{cases} \qquad c_n=\begin{cases}1, & 0\leqslant n\leqslant K+1,\\ 0, & K+2\leqslant n.\end{cases}$$

2. Другой вариант системы, в которой клиенты отказываются становиться в очередь

Допустим, что в одноканальной системе с пуассоновским входящим потоком с параметром λ и экспоненциальным временем обслуживания с параметром μ имеем $b_0=1$,

$$0\leqslant b_{i+1}\leqslant b_i\leqslant 1 \quad (i=0, 1, 2, \ldots).$$

Здесь b_i — вероятность того, что прибывающий клиент становится в очередь, если в системе находится i клиентов. Хомма [358] получил для этой задачи уравнения для вероятностей $P_n(t)$ того, что в системе в момент времени t находится n клиентов:

$$P'_n(t) = \lambda[b_{n-1}P_{n-1}(t) - b_nP_n(t)] - \mu[P_n(t) - P_{n+1}(t)], \quad n \geqslant 1,$$
$$P'_0(t) = \mu P_1(t) - \lambda P_0(t). \qquad (13.20)$$

Пусть $W(t)$ — время ожидания клиента, который становится в очередь в момент времени t. Введем функцию $P(w, t)$ и вычислим ее значение:

$$P(w, t) \equiv P[W(t) \leqslant w] = \int_0^w \sum_{n=0}^{\infty} P_n(t)\, e^{-\mu y} \frac{(\mu y)^{n-1}}{(n-1)!} \mu dy. \qquad (13.21)$$

Преобразование Лапласа—Стилтьеса функции $P(w, t)$ по w имеет вид

$$\gamma(s, t) = \int_0^{\infty} e^{-sw} dP(w, t) = \sum_{n=0}^{\infty} P_n(t) \left(\frac{\mu}{\mu+s}\right)^n. \qquad (13.22)$$

Взяв производную по времени, подставив значение $P_n(t)$ из уравнений для переходного состояния и положив

$$\frac{\partial\gamma(s, t)}{\partial t} = 0,$$

получим преобразование функции распределения времени ожидания в стационарном состоянии в виде

$$\gamma(s) = \frac{\lambda}{\mu+s} \sum_{n=0}^{\infty} b_n p_n \left(\frac{\mu}{\mu+s}\right)^n + p_0. \qquad (13.23)$$

Положив $s = 0$, имеем

$$p_0 = 1 - \rho \sum_{n=0}^{\infty} b_n p_n. \qquad (13.24)$$

Получив коэффициенты при s и s^2 в выражении для $\gamma(s)$, найдем среднее значение $E(W)$ и дисперсию $\sigma^2(W)$ времени ожидания W:

$$E(W) = \frac{1-p_0}{\mu} + \frac{\rho}{\mu} \sum_{n=1}^{\infty} n b_n p_n, \qquad (13.25)$$

$$\sigma^2(W) = \frac{\rho}{\mu^2} \sum_{n=0}^{\infty} (n+1)(n+2) b_n p_n - [E(W)]^2. \qquad (13.26)$$

Заметим, что, положив $b_n = 1$ $(n \geqslant 0)$, получим известный результат для среднего времени ожидания в стационарном состоянии:

$$E(W) = \frac{\rho}{\mu(1-\rho)},$$

так как $\lim_{t\to\infty} P_n(t) = p_n = (1-\rho)\rho^n, \; n \geqslant 0$, и

$$\sum_{n=1}^{\infty} np_n = \frac{\rho}{1-\rho}.$$

Кроме того, рассматривая условия существования стационарного состояния и используя метод вложенных цепей Маркова, Хомма рассмотрел задачу такого типа для произвольного распределения входящего потока, поступающего в многоканальную систему.

Упражнение 6. Убедитесь в справедливости уравнений, описывающих данную систему, и проверьте выражения для математического ожидания и дисперсии.

3. Система с возможностью неприсоединения к очереди при произвольном входящем потоке

Рассмотрим одноканальную систему с произвольным входящим потоком и экспоненциальным распределением времени обслуживания. Обозначим через $\tau_1, \ldots, \tau_n, \ldots$ моменты поступления требований, а через $A(x)$ — распределение промежутков времени между требованиями $\tau_n - \tau_{n-1}$ (предполагается, что они взаимно независимы и имеют одинаковые распределения). Пусть случайная величина $n(t)$ означает число требований, находящихся в очереди в момент времени t. Введем обозначение $\eta_n = \eta(\tau_n - 0)$.

Пусть $\{\gamma_n\}$ — последовательность взаимно независимых одинаково распределенных случайных величин, принимающих целочисленные значения. Максимально допустимое число требований, находящихся в очереди, равно $N+1$, включая обслуживаемое требование. Когда очередь достигает такой длины, поступающие требования не становятся в очередь. Требование, поступающее в момент времени τ_n, становится в очередь в том и только в том случае, если $\eta_n \leqslant \min(N, \gamma_n)$. Невозрастающая последовательность неотрицательных действительных чисел b_m при $b_0 = 1$ определяется как $b_m = P(\gamma_n \geqslant m)$, $m = 0, 1, \ldots$. Это выражение есть вероятность того, что поступающее требование становится в очередь, если в ней находится m требований. Необходимо найти распределение числа требований, находящихся в системе (включая обслуживаемое требование):

$$p_k = \lim_{n\to\infty} P(\eta_k = k), \quad k = 0, 1, \ldots, N+1.$$

Финч [244] доказал существование p_k и его независимость от начального состояния.

Для произвольного входящего потока имеем

$$p_k = \frac{q_{N+1-k}}{\sum_{j=0}^{N+1} q_j}, \quad k = 0, 1, \ldots, N+1, \tag{13.27}$$

где q_j определяется из выражения

$$q(z)=\frac{(1-z)\widetilde{A}[\mu(1-z)]}{1-\widetilde{A}[\mu(1-z)]}\overline{q\beta(z)}, \quad q_0=1. \qquad (13.28)$$

Здесь

$$q(z)=\sum_{i=0}^{\infty} q_i z^i; \qquad \overline{q\beta(z)}=\sum_{j=1}^{\infty} q_j\beta_j z^{j-1};$$

$$\widetilde{A}(s)=\int_0^{\infty} e^{-sx}dA(x),$$

а $\beta_i(i=0, 1, \ldots)$, $0<\beta_i\leqslant 1$ — произвольные постоянные, причем $\beta_k=1$, $k>N+1$. Значения q_i можно получить, дифференцируя (13.28), или из рекуррентного соотношения:

$$q_k=r_0\beta_{k+1}q_{k+1}+\sum_{i=0}^{k-1}(r_{i+1}-r_i)\beta_{k-i}q_{k-i}+$$

$$+\sum_{i=0}^{k} r_i q_{k-i}, \quad k=0, 1, \ldots, \qquad (13.29)$$

где

$$r_i=\int_0^{\infty} e^{-\mu x}\frac{(\mu x)^i}{i!}dA(x). \qquad (13.30)$$

Эта формула для p_k согласуется с аналогичной формулой, полученной Такачем при $b_m=1(m=0, \ldots, N)$, и формулой, которую получил Хейт для пуассоновского входящего потока.

Если входящий поток имеет эрланговское распределение E_2, т. е.

$$dA(x)=\lambda^2 x e^{-\lambda x}, \quad x\geqslant 0, \qquad (13.31)$$

то $\widetilde{A}(s)=\frac{\lambda^2}{(\lambda+s)^2}$. При $\rho=\frac{\lambda}{\mu}$ получим

$$p_k=\left(\frac{\rho^2}{2\rho+1}\right)^k \frac{b_{k-1}b_{k-2}\ldots b_0}{w_{k-1}w_{k-2}\ldots w_0}p_0, \quad k=1, \ldots, N+1, \qquad (13.32)$$

где p_0 определяется обычным способом, w_k выражается в виде непрерывной дроби

$$w_k=1-\frac{db_{k+1}}{1-}\frac{db_{k+2}}{1-}\ldots\frac{db_N}{1-}, \quad k=0, 1, \ldots, N-1,$$

$w_N=1$, и

$$d=\left[\frac{\rho}{(2\rho+1)}\right]^2.$$

Теорема 13.1. Если $A(x)$ не является решетчатой функцией распределения, т. е. если это распределение не сосредоточено на целочисленных значениях x, то можно показать, что распределение числа требований, находящихся в очереди при стационарном состоянии,

$$\lim_{t\to\infty} P[\eta(t)=k]=p_k, \quad k=0, 1, \ldots, N+1,$$

существует и не зависит от начального состояния. Это распределение имеет вид

$$p_k=\frac{\lambda}{\mu} b_{k-1}p_{k-1}, \quad k=1, 2, \ldots, N+1, \tag{13.33}$$

$$p_0=1-\frac{\lambda}{\mu}\sum_{n=0}^{N} b_n p_n.$$

Доказательство. Покажем, что существование распределения p_k зависит от следующих вероятностей перехода конечной цепи Маркова η_n $(n=1, 2, \ldots)$:

$$P(\eta_{n+1}=k\,|\,\eta_n=j)=b_j r_{j,k}+(1-b_j)r_{j-1,k}\equiv\pi_{j,k}, \tag{13.34}$$

$$j,\ k=0,\ 1,\ \ldots,\ N,$$

$$\pi_{N+1,k}=r_{N,k},$$

где

$$r_{N,k}=\begin{cases} 0, & j+1-k<0,\\ \displaystyle\int_0^\infty e^{-\mu x}\frac{(\mu x)^{j+1-k}}{(j+1-k)!}dA(x)=r_{j+1-k}, & j+1-k\geqslant 0. \end{cases} \tag{13.35}$$

Данная цепь Маркова является эргодической, следовательно, распределение p_k существует и не зависит от распределения η_1.

Значение p_k однозначно определяется из выражения

$$p_k=\sum_{j=k-1}^{N+1}\pi_{j,k}p_j,\ k=0,\ 1,\ \ldots,\ N, \tag{13.36}$$

$$\sum_{k=0}^{N+1}p_k=1.$$

Подставляя значение $\pi_{j,k}$, вводя обозначения $Q_k=p_{N+1-k}$ $(k=0,\ 1,\ \ldots,\ N+1)$ и $\beta_k=\beta_{N+1-k}$ $(k=1,\ 2,\ \ldots,\ N+1)$, получаем решение для Q_k, выраженное через Q_0, и находим $Q_k=p_0 q_k$, что является искомым результатом.

13.3. Уход из очереди до начала обслуживания

В этом случае клиенты покидают очередь после того, как они пробудут в ней некоторое время. Обычно это происходит вследствие того, что длительность ожидания является значительной.

1. Система, в которой клиенты отказываются становиться в очередь и покидают очередь до начала обслуживания

Примем те же обозначения, что и в первом варианте системы, в которой клиенты отказываются становиться в очередь, когда рассматривалась одноканальная система с пуассоновским входящим потоком и экспоненциальным временем обслуживания. Обозначим через $r(n)\Delta t$ вероятность того, что за время Δt клиент покинет очередь, если в ней находится n клиентов.

Рассматривая вероятность отсутствия переходов в другое состояние, вероятность прибытия клиента, который отказывается становиться в очередь, вероятность прибытия клиента, который становится в очередь, вероятность ухода обслуженного клиента и вероятность ухода клиента до начала обслуживания, получим систему дифференциально-разностных уравнений. Пренебрегая переходами высших порядков, имеем

$$\begin{aligned} p_n(t+\Delta t) = {} & (1-\lambda\Delta t)(1-\mu\Delta t)[1-r(n)\Delta t]\, p_n(t) + \\ & + \lambda\Delta t(1-\mu\Delta t)[1-r(n)\Delta t]\, p_n(t) F(n-1) + \\ & + \lambda\Delta t(1-\mu\Delta t)[1-r(n-1)\Delta t]\, p_{n-1}(t) G(n-2) + \\ & + (1-\lambda\Delta t)\mu\Delta t[1-r(n+1)\Delta t]\, p_{n+1}(t) + \\ & + (1-\lambda\Delta t)(1-\mu\Delta t) r(n+1)\Delta t p_{n+1}(t). \end{aligned} \tag{13.37}$$

При $n=0$ в первом члене μ опускается и $r(0)=r(1)=0$. После обычных упрощений найдем уравнение

$$\begin{aligned} \frac{d}{dt} p_n(t) = {} & -[\lambda+\mu+r(n)]\, p_n(t) + \lambda p_n(t) F(n-1) + \\ & + \lambda p_{n-1}(t) G(n-2) + \mu p_{n+1}(t) + r(n-1) p_{n+1}(t). \end{aligned} \tag{13.38}$$

Просуммировав от нуля до n и произведя упрощения, получим

$$\frac{d}{dt} P_n(t) = \mu p_{n+1}(t) - \lambda G(n-1) p_n(t) + r(n+1) p_{n+1}(t). \tag{13.39}$$

Допуская, что при $t\to\infty$ вероятности $p_n(t)$ и $P_n(t)$ стремятся к пределу, получим уравнение

$$\mu p_{n+1} + r(n+1) p_{n+1} = \lambda G(n-1) p_n, \tag{13.40}$$

которое имеет решение

$$p_n = c_n \left(\frac{\lambda}{\mu}\right)^n p_0. \tag{13.41}$$

Здесь p_0 определяется из соотношения $\sum\limits_{n=0}^{\infty} p_n = 1$. Если обозначить

$$R(n) = \frac{\lambda}{\mu + r(n)}, \tag{13.42}$$

то

$$c_n = \prod_{j=0}^{n-2} G(j) R(j+2). \tag{13.43}$$

При соответствующем ограничении функций можно получить решение для случая, когда клиенты остаются необслуженными либо вследствие того, что они покидают очередь до начала обслуживания, либо вследствие того, что они отказываются становиться в очередь.

Упражнение 7. Проверьте это положение для рассмотренного ранее случая, когда клиент отказывается становиться в очередь.

2. Нетерпеливые клиенты

Баррер [28, 29] исследовал систему, в которой клиент уходит из очереди, если время ожидания больше допустимого. Он ввел понятие о клиенте, принятом на обслуживание. Клиент может ожидать в очереди и не будучи принятым на обслуживание. Если клиент принят на обслуживание, то он будет ожидать до заранее установленного предела, в противном случае он уходит из очереди. В другом рассматриваемом случае клиент независимо от того, принят он на обслуживание или нет, покидает систему, как только достигается намеченный заранее предел времени ожидания. В первом случае рассматривается многоканальная система с обслуживанием в порядке прибытия клиентов, а во втором случае — одноканальная система со случайным выбором на обслуживание. Результат, полученный для одноканальной системы со случайным выбором на обслуживание, применим и к c-канальной системе, $c>1$ (см. работу Б. В. Гнеденко, результат которой приведен в гл. 9).

Рассмотрим одноканальную систему с нетерпеливыми клиентами и случайным выбором на обслуживание, в которую поступает пуассоновский входящий поток, а время обслуживания имеет экспоненциальное распределение. Обозначим через $P_n(t)$ вероятность того, что в момент времени t в очереди находится n клиентов, а через $C_n(\tau_0,\ t)\Delta t+o(\Delta t)$ — вероятность того, что в промежутке времени $(t,\ t+\Delta t)$ клиент уйдет из очереди, если в момент времени t в очереди находится n клиентов; предельное время ожидания этого клиента равно τ_0. Обычным способом находим систему уравнений:

$$P_0'(t) = -\lambda P_0(t) + [\mu + C_1(\tau_0,\ t)]\, P_1(t),$$

$$\begin{aligned} P_n'(t) = \lambda P_{n-1}(t) - [\lambda + \mu + C_n(\tau_0,\ t)]\, P_n(t) + \\ + [\mu + C_{n+1}(\tau_0,\ t)]\, P_{n+1}(t),\quad n \geqslant 1. \end{aligned} \qquad (13.44)$$

Баррер показал, что в отличие от обычной системы с ожиданием, когда клиенты остаются на обслуживание несмотря ни на что, точное задание числа клиентов, находящихся в очереди в момент времени t, как правило, не определяет полностью поведение очереди в будущем. Это положение справедливо, так как клиенты могут стать нетерпеливыми и покинуть очередь. Здесь необходимо найти время или распределение времени, в течение которого клиент ожидает, за исключением частных случаев, где это не требуется. Это нужно для того, чтобы определить $C_n(\tau_0,\ t)$, а затем найти начальные условия, необходимые для решения этой системы уравнений. Для определения $C_n(\tau_0, t)$ положим, что $C_n(\tau, t, \Delta\tau)\Delta\tau$ — вероятность того, что в очереди имеется клиент, который ожидал от момента $\tau-\Delta\tau$ до момента τ, если в момент t в очереди находится n клиентов. Так как при случайном выборе вероятность того, что клиент поступит на обслуживание, равна

$$\frac{\mu\Delta\tau}{n}+o(\Delta\tau),$$

то

$$C_n(\tau, t, \Delta\tau)\left(1-\frac{\mu\Delta\tau}{n}\right)=C_n(\tau+\Delta\tau, t+\Delta\tau, \Delta\tau)\Delta\tau. \quad (13.45)$$

После упрощения приходим к уравнению

$$\frac{\partial}{\partial t}C_n(\tau, t, 0)+\frac{\partial}{\partial t}C_n(\tau, t, 0)=-\frac{\mu}{n}C_n(\tau, t, 0), \quad (13.46)$$

которое имеет решение

$$C_n(\tau, t, 0)=f_n(t-\tau)e^{-\frac{\mu\tau}{n}}. \quad (13.47)$$

Здесь функция $f_n(t-\tau)$ определяется из соответствующих начальных условий. Если левая часть выражения (13.47) принимается непрерывной в точке $\Delta\tau=0$, то можно записать

$$C_n(\tau, t, \Delta\tau)\Delta\tau=f_n(t-\tau)e^{-\frac{\mu\tau}{n}}\Delta\tau+o(\Delta\tau).$$

Искомая величина $C_n(\tau_0, t)\Delta\tau$ — левая часть последнего соотношения при $\tau=\tau_0$. Отсюда

$$C_n(\tau, t)=f_n(t-\tau)e^{-\frac{\mu\tau}{n}}. \quad (13.48)$$

Решение данной системы уравнений для стационарного случая можно получить, рассмотрев решение уравнений стационарного процесса рождения и гибели. Введем обозначение $C_n=\lim\limits_{t\to\infty}C_n(\tau_0, t)=$ $=Ae^{-\frac{\mu\tau_0}{n}}$, где A_n определяется выражением

$$\int_0^{\tau_0}C_n(\tau, t)\,d\tau=n=A_n\frac{n}{\mu}\left(1-e^{-\frac{\mu\tau_0}{n}}\right). \quad (13.49)$$

Интенсивность, с которой нетерпеливые клиенты уходят из очереди, равна

$$C=\sum_{n=1}^{\infty}p_nC_n=\lambda-\mu(1-p_0). \quad (13.50)$$

Таким образом, вероятность того, что клиент будет потерян, составляет

$$\frac{1-(1-\rho)\sum\limits_{n=1}^{\infty}\rho^{n-1}\prod\limits_{k=1}^{n}\left(1-e^{-\frac{\mu\tau_0}{k}}\right)}{1+\sum\limits_{n=1}^{\infty}\rho^{n}\prod\limits_{k=1}^{n}\left(1-e^{-\frac{\mu\tau_0}{k}}\right)}. \quad (13.51)$$

Система с c параллельными каналами, пуассоновским входящим потоком и экспоненциальным временем обслуживания, одинаковым для каждого канала, описывается уравнениями

$$P'_0(t) = -\lambda P_0(t) + \mu P_1(t),$$

$$P'_n(t) = \lambda P_{n-1}(t) - (\lambda + n\mu) P_n(t) + (n+1)\mu P_{n+1}(t), \quad n < c,$$

$$\begin{aligned} P'_n(t) = \lambda P_{n-1}(t) - [\lambda + c\mu + C_n(t)] P_n(t) + \\ + [c\mu + C_{n+1}(t)] P_{n+1}(t), \quad n \geqslant c. \end{aligned} \tag{13.52}$$

Здесь $C_n(t)\Delta t + o(\Delta t)$ — вероятность того, что в промежутке времени $(t, t+\Delta t)$ клиент будет потерян, если в момент времени t в очереди находится n клиентов.

Пальм решил задачу для стационарного состояния марковской c-канальной системы с потерями. (См. также статью Коэна в гл. 14, задача 10.)

13.4. Ориентирование требований

Леруа и Воло [503] рассмотрели следующую задачу. Допустим, что в системе массового обслуживания с c параллельными каналами (телефонные линии) требуется определенное время для ориентирования вызова (перед началом обслуживания). Это время имеет экспоненциальное распределение со средней длительностью b. Время обслуживания (длительность телефонного разговора) также имеет экспоненциальное распределение. Примем математическое ожидание времени обслуживания равным одной временно́й единице. Пусть λ — параметр пуассоновского потока. Обозначим через p_i вероятность того, что $i\,(0 \leqslant i \leqslant c)$ каналов действительно заняты разговором, а через $q_j\,(0 \leqslant j \leqslant c)$ — вероятность того, что занято c каналов, причем $j-1$ каналов занято разговором, а $c-j+1$ каналов — ориентированием вызовов. Если все каналы заняты, то вызов теряется. В этом случае уравнения для стационарного состояния имеют вид

$$cp_c = \frac{q_c}{b},$$

$$(\lambda + i)p_i = (i+1)p_{i+1} + \frac{q_i}{b}, \quad 0 \leqslant i \leqslant c,$$

$$\left(c - 1 + \frac{1}{b}\right) q_c = \lambda p_{c-1}, \tag{13.53}$$

$$\left(j - 1 + \frac{1}{b}\right) q_j = \lambda p_{j-1} + jq_{j+1}, \quad 0 \leqslant j \leqslant c,$$

при

$$\sum_{i=0}^{c} p_i + \sum_{j=1}^{c} q_j = 1.$$

Суммируя эту систему уравнений и используя последнее равенство, получим

$$b\lambda \sum_{i=0}^{c-1} p_i = \sum_{j=1}^{c} q_j = 1 - \sum_{i=0}^{c} p_i. \tag{13.54}$$

Доля потерянных требований равна

$$p_c + \sum_{j=1}^{c} q_j = \frac{b\lambda + p_c}{b\lambda + 1}. \tag{13.55}$$

Уравнение для стационарного состояния позволяет выразить p_i и q_j через p_c. Значения q_j удовлетворяют соотношению

$$\frac{q_j}{b} = j\left(\frac{1}{b\lambda} + \frac{\lambda + j}{\lambda}\right) q_{j+1} - j\,\frac{j+1}{\lambda} q_{j+2}, \quad 0 < j < c. \tag{13.56}$$

Используя обозначение $(N,\ n) = N(N+1)\ \dots\ (N+n-1)$, получим

$$\frac{q_{c-j}}{(c+1,\ j+1)\, b p_c} = \left(c + \lambda + \frac{1}{b};\ 1\right)\left(\frac{b}{\lambda}\right)^j -$$
$$- \frac{(j-1;\ 1)}{(1;\ 1)}\left(c + \lambda + \frac{1}{b};\ 2\right)\left(\frac{b}{\lambda}\right)^{j-1} +$$
$$+ \frac{(j-3;\ 2)}{(1;\ 2)}\left(c + \lambda + \frac{1}{b};\ 3\right)\left(\frac{b}{\lambda}\right)^{j-2} -$$
$$- \frac{(j-5;\ 3)}{(1;\ 3)}\left(c + \lambda + \frac{1}{b};\ 4\right)\left(\frac{b}{\lambda}\right)^{j-3} + \dots \tag{13.57}$$

Отсюда находим

$$\frac{\sum\limits_{j=1}^{c} q_j}{b p_c} = (c;\ 1) + (c-1;\ 2)\left(\lambda + \frac{1}{b} + 1;\ 1\right)\left(\frac{b}{\lambda}\right) +$$
$$+ (c-2;\ 3)\left(\lambda + \frac{1}{b} + 1;\ 2\right)\left(\frac{b}{\lambda}\right)^2 + \dots +$$
$$+ (0;\ c)\left(\lambda + \frac{1}{b} + 1;\ c-1\right)\left(\frac{b}{\lambda}\right)^c. \tag{13.58}$$

Обозначив правую часть этого выражения через Q, получим вероятность того, что все каналы заняты:

$$p_c = \frac{\lambda}{\lambda + (1 + b\lambda)\, Q}. \tag{13.59}$$

Доля потерянных требований равна

$$\frac{b\lambda}{b\lambda + 1} + \frac{\lambda}{b\lambda + 1}\,\frac{1}{\lambda + (1 + b\lambda)\, Q} = \frac{\lambda(1 + bQ)}{\lambda + (1 + b\lambda)\, Q}.$$

При фиксированных значениях b и λ при $c\to\infty$ это выражение стремится к $\frac{b\lambda}{b\lambda+1}$. При $\lambda\to\infty$ доля теряемых требований стремится к единице при любом c. При $b\to 0$ имеем известную формулу Эрланга для системы с потерями

$$\frac{\frac{\lambda^c}{c!}}{1+\lambda+\frac{\lambda^2}{2!}+\ldots+\frac{\lambda^c}{c!}}. \tag{13.60}$$

Если $\lambda\to\infty$ и $c\to\infty$ и при этом их отношение $\frac{\lambda}{c}$ остается постоянным, то это выражение стремится к нулю.

Более изящную формулировку этой задачи в другом изложении см. в гл. 14 при рассмотрении системы с потерями, в которой производится управление входящим потоком. Эти положения распространяются также и на системы с ожиданием.

13.5. Запаздывание требований относительно запланированных моментов поступления

Уинстен [887] рассмотрел одноканальную систему с регулярным входящим потоком и экспоненциальным временем обслуживания. Изучая различные траектории, по которым очередь может пройти между двумя точками, которые являются моментами поступления требований, он получил распределение числа требований, находящихся в очереди при стационарном состоянии. Уинстен получил вероятность $\bar{\pi}_n$ того, что в момент времени, предшествующий поступлению требования, в системе находится n требований:

$$\bar{\pi}_n=\pi^n\bar{\pi}_0, \tag{13.61}$$

где π находится из соотношения

$$\pi e^{\lambda}=e^{\lambda\pi} \tag{13.62}$$

и, как обычно, $\bar{\pi}_0=1-\pi$.

Для случая, когда требования запаздывают, хотя и предусматривается поступление требований через постоянные промежутки времени, $P(\leqslant t)$ — вероятность того, что длительность задержки не превышает t. Требуется выполнение условия $0\leqslant t<1$, т. е. хотя требование и запаздывает, однако оно должно поступить до запланированного момента поступления следующего за ним требования. Можно получить выражения для распределения числа требований, находящихся в очереди: 1) в момент, когда запланировано поступление требования, 2) в момент, когда оно фактически поступает и 3) в произвольный момент времени. Распределение для первого случая имеет вид

$$\pi_n=\pi_1\pi^{n-1},\quad n\geqslant 1. \tag{13.63}$$

Так, из соотношения

$$\sum_{n=0}^{\infty} \pi_n = 1$$

находим

$$\pi_1 = (1-\pi)(1-\pi_0), \qquad (13.64)$$

$$\pi_0 = 1 - \pi \frac{\int\limits_0^1 e^{\lambda t} dP(\leqslant t)}{\int\limits_0^1 e^{\lambda \pi t} dP(\leqslant t)}. \qquad (13.65)$$

Здесь π определяется, как и ранее.

Чтобы найти число требований, находящихся в очереди в моменты фактического поступления требований, необходимо выразить $\pi_n(t)$ через π_n и усреднить по t. Получим

$$\pi_n(t) = \sum_{i=0}^{\infty} \pi_{n+i} e^{-\lambda t} \frac{(\lambda t)^i}{i!}, \quad n \geqslant 1. \qquad (13.66)$$

Таким образом,

$$\pi_n(t) = \pi_1 \pi^{n-1} e^{-\lambda t (1-\pi)},$$

$$\pi_0(t) = 1 - \frac{\pi_1}{1-\pi} e^{-\lambda t (1-\pi)}. \qquad (13.67)$$

Распределение числа требований, находящихся в очереди непосредственно перед запланированными моментами поступления, имеет вид

$$\bar{\pi}_n = \bar{\pi}_1 \pi^{n-1}, \quad n \geqslant 1, \qquad (13.68а)$$

$$\bar{\pi}_1 = \pi_1 \int\limits_0^1 e^{-\lambda t (1-\pi)} dP(\leqslant t) =$$

$$= \pi(1-\pi) \frac{\int\limits_0^1 e^{\lambda t} dP(\leqslant t) \int\limits_0^1 e^{-\lambda t + \lambda \pi t} dP(\leqslant t)}{\int\limits_0^1 e^{\lambda \pi t} dP(\leqslant t)}, \qquad (13.68б)$$

$$\bar{\pi}_0 = \int\limits_0^1 \pi_0(t)\, dP(\leqslant t). \qquad (13.68в)$$

Заметим, что если $P(\leqslant t)$ имеет единственный скачок, т. е. все требования запаздывают на одно и то же время, то задача сводится к случаю, когда требования поступают в запланированные моменты времени. В этом случае можно считать, что поступление требований запланировано на более позднее время, включающее и задержку.

И, наконец, распределение числа требований, находящихся в очереди, для общего случая поступления требований с задержкой равно взвешенному среднему значению:

$$p_n = \left(1 - \frac{\bar{\pi}_1}{1-\pi}\right) P(n=0) + \frac{\bar{\pi}_1}{1-\pi} P(n \geqslant 1), \qquad (13.69)$$

где n — число требований, находящихся в очереди. Но $P(n \geqslant 1) = \pi^{n-1}(1-\pi)$, следовательно $P(n=0)$ можно определить из соотношения $\sum\limits_{n=0}^{\infty} p_n = 1$. Аналогично можно вычислить время ожидания. Автор указывает на ряд обобщений, не приводя, однако, решений в явном виде.

13.6. Переход клиентов из одной очереди в другую

Представляет интерес рассмотреть систему с несколькими параллельными каналами, в которой клиенты переходят туда и обратно между несколькими очередями; к сожалению, в данном случае трудно продвинуться далеко в аналитическом исследовании. Проиллюстрируем это явление на примере двухканальной системы с пуассоновским входящим потоком с параметром λ и экспоненциальным временем обслуживания с параметрами μ_1 и μ_2 соответственно. Предположим, что эти две очереди всегда имеют одну и ту же длину или же число клиентов, находящихся в них, отличается на единицу. Как только разница составит более одного клиента, последний клиент из более длинной очереди немедленно перемещается в более короткую очередь. Вновь прибывающий клиент всегда становится в более короткую очередь, если очереди не равны. Если очереди равны, он становится в очередь к обслуживающему устройству с более коротким временем обслуживания, если он располагает подобной информацией, или с равной вероятностью занимает место в любой очереди при отсутствии такой информации. Рассмотрим второй случай.

Пусть $P(n_1, n-n_1, t)$ — вероятность того, что в момент времени t в первом канале системы находится n_1 клиентов (включая и обслуживаемого), а во втором канале — $(n-n_1)$ клиентов, где n — общее число клиентов, находящихся в системе. Имеем:

$$P(0,\ 0,\ t+\Delta t) = (1-\lambda\Delta t) P(0,\ 0,\ t) + \mu_1 \Delta t P(1,\ 0,\ t) + \mu_2 \Delta t P(0,\ 1,\ t),$$

$$P(1,\ 0,\ t+\Delta t) = [1-(\lambda+\mu_1)\Delta t] P(1,\ 0,\ \Delta t) + \frac{\lambda}{2}\Delta t P(0,\ 0,\ t) + \mu_2 \Delta t P(1,\ 1,\ t),$$

$$P(0,\ 1,\ t+\Delta t) = [1-(\lambda+\mu_2)\Delta t] P(0,\ 1,\ \Delta t) + \frac{\lambda}{2}\Delta t P(0,\ 0,\ t) + \mu_1 \Delta t P(1,\ 1,\ t).$$

При $n > 2$, если $n = 2n_1$, имеем

$$P(n_1, n-n_1, t+\Delta t) = [1-(\lambda+\mu_1+\mu_2)\Delta t] P(n_1, n-n_1, t) + \\ + \frac{\lambda\Delta t}{2}[P(n_1-1, n-n_1, t)+P(n_1, n-n_1-1, t)] + \\ + \mu_1\Delta t P(n_1+1, n-n_1, t)+\mu_2\Delta t P(n_1, n-n_1+1, t).$$

Если $n = 2n_1+1$, т. е. если во втором канале на одного клиента больше, чем в первом, имеем

$$P(n_1, n-n_1, t+\Delta t) = [1-(\lambda+\mu_1+\mu_2)\Delta t] P(n_1, n-n_1, t) + \\ + \frac{\lambda}{2}\Delta t P(n_1, n-n_1-1, t)+\mu_1\Delta t P(n_1+1, n-n_1, t).$$

С другой стороны, если в первом канале больше на одного клиента, т. е.

$$n = 2n_1 - 1,$$

то имеем

$$P(n_1, n-n_1, t+\Delta t) = [1-(\lambda+\mu_1+\mu_2)\Delta t] P(n_1, n-n_1, t) + \\ + \frac{\lambda}{2}\Delta t P(n_1-1, n-n_1, t)+\mu_2\Delta t P(n_1, n-n_1+1, t).$$

13.7. Переменное число параллельных каналов

Во многих практических ситуациях в системах с параллельными каналами свободные обслуживающие устройства могут выполнять другую работу. Примером этого может служить выдача инструментов рабочим. Когда у раздаточного окошка никого нет или находится всего несколько человек, служащие инструментальной кладовой могут вести подготовительные или другие несрочные работы. Когда же число ожидающих увеличивается до некоторого определенного уровня, второй или третий служащий (канал) приступает к обслуживанию ожидающих рабочих. В подобных ситуациях, которые наблюдаются также в банках, магазинах, ремонтных мастерских и при выполнении других операций, число действующих каналов является случайной величиной.

Для стационарного состояния одноканальной системы с пуассоновским входящим потоком с параметром λ и одинаковым экспоненциальным распределением времени обслуживания с параметром μ Романи [735] получил вероятности, описывающие переменное число каналов, минимальное число которых равно единице и растет с увеличением числа ожидающих требований. Число каналов увеличивается только после того, как достигается максимально допустимая длина очереди; т. е. как только число требований, находящихся в очереди, становится больше N, новый канал приступает к обслуживанию требования, стоящего в очереди первым. Если ожидающих требований нет, то по окончании обслуживания число каналов уменьшается, и только один канал остается открытым все время. Пусть $P_n(m, t)$ — вероятность того, что в момент времени t в очереди находится n требований и m требований обслуживается (число их равно числу действующих каналов). Тогда $p_n(m)$ — соответствующая вероятность для стационарного состояния. Имеем:

$$\begin{gathered}
P_n(m, t+\Delta t) = (1-\lambda\Delta t)(1-m\mu\Delta t)P_n(m, t) + \\
+\lambda\Delta t P_{n-1}(m, t) + m\mu\Delta t P_{n+1}(m, t), \\
P_N(m, t+\Delta t) = (1-\lambda\Delta t)(1-m\mu\Delta t)P_N(m, t) + \\
+\lambda\Delta t P_N(m-1, t) + \lambda\Delta t P_{N-1}(m, t), \\
P_0(m, t+\Delta t) = (1-\lambda\Delta t)(1-m\mu\Delta t)P_0(m, t) + \\
+m\mu\Delta t P_1(m, t) + (m+1)\mu\Delta t P_0(m+1, t), \\
P_0(1, t+\Delta t) = (1-\lambda\Delta t)(1-\mu\Delta t)P_0(1, t) + \lambda\Delta t P_0(0, t) + \\
+\mu\Delta t P_1(1, t) + 2\mu\Delta t P_0(2, t), \\
P_0(0, t+\Delta t) = (1-\lambda\Delta t)P_0(0, t) + \mu\Delta t P_0(1, t).
\end{gathered} \tag{13.70}$$

Когда действует m каналов, то вероятность того, что за время Δt будет закончено обслуживание одного требования, равна $m\mu\Delta t$. Имеем начальные условия

$$P_n(m, 0) = \delta_{ni}^{mj} \equiv \begin{cases} 1, & n=i, \quad m=j, \\ 0, & n\neq i, \quad m\neq j. \end{cases} \tag{13.71}$$

Уравнения для стационарного состояния, выраженные через $p_n(m)$, находятся обычным способом. Решение этих уравнений имеет вид

$$p_n(1) = \frac{\left(\frac{\rho}{1!}\right)\sum\limits_{k=0}^{N-n} \rho^{-k}}{(N+1)e^{\rho} + \sum\limits_{k=1}^{N}(N-k+1)\rho^{-k}}, \tag{13.72}$$

$$p_n(m) = \frac{\frac{\rho^m}{m!}}{(N+1)e^{\rho} + \sum\limits_{k=1}^{N}(N-k+1)\rho^{-k}}, \tag{13.73}$$

где $\rho = \frac{\lambda}{\mu}$ (а не $\rho = \frac{\lambda}{c\mu}$, как в случае многоканальной системы с постоянным числом каналов).

Читатель может проверить справедливость этих формул для системы с бесконечным числом каналов, которую получим, положив $N=0$ (второй член знаменателя отбрасывается). В этом случае

$$p_n(m) = \frac{\rho^m}{m!}e^{-\rho}. \tag{13.74}$$

Если допустить, что N стремится к бесконечности, то в результате получим одноканальную систему, и можно показать, что

$$\lim_{N\to\infty} p_{n-1}(1)=\rho^n(1-\rho). \tag{13.75}$$

Можно также показать, что среднее число требований, находящихся на обслуживании, равно

$$\sum_{m=0}^{\infty} mp_n(m)=\rho. \tag{13.76}$$

Вариант модели Ромáни изучал Филлипс [643]. В этом случае число каналов ограничено максимальным числом M. Когда действует M каналов, то очередь может возрастать безгранично. Во всем остальном процесс аналогичен описанному, за исключением разницы в обозначениях. В этой модели «точкой переключения» считается наличие в очереди N, а не $N+1$ требований. Для описания процесса необходимо восемь уравнений для стационарного состояния, которые для нескольких частных случаев решаются рекуррентным способом. Было получено общее решение в следующем виде:

$$p_n(1)=\frac{\rho^{n+1}(1-\rho^{N-n})}{1-\rho^N}p_0(0),\quad 0\leqslant n\leqslant N+1,\quad m=1, \tag{13.77а}$$

$$p_n(m)=\frac{\rho^{N+m-1}(1-\rho)}{m!\,(1-\rho^N)}p_0(0),\quad \begin{matrix}0\leqslant n\leqslant N-1,\\ 2\leqslant m\leqslant M-1,\end{matrix} \tag{13.77б}$$

$$p_n(M)=\frac{\rho^{N+M-1}(1-\rho)(M^{n+1}-\rho^{n+1})}{M^nM!\,(1-\rho^N)(M-\rho)}p_0(0),$$
$$0\leqslant n\leqslant N-2,\quad m=M, \tag{13.77в}$$

$$p_n(M)=\frac{\rho^{n+M}(1-\rho)(M^N-\rho^N)}{M^nM!\,(1-\rho^N)(M-\rho)}p_0(0),\quad \begin{matrix}N-1\leqslant n<\infty,\\ m=M,\end{matrix} \tag{13.77г}$$

где

$$p_0(0)=\left\{1+\frac{\rho}{1-\rho}-\frac{N\rho^{N-1}}{1-\rho^N}+\frac{N(1-\rho)\rho^{N-1}}{1-\rho^N}\times\right.$$
$$\left.\times\left[\sum_{m=0}^{M-1}\frac{\rho^m}{m!}+\frac{\rho^M}{(M-1)!\,(M-\rho)}\right]\right\}^{-1}. \tag{13.78}$$

Можно показать, что при $N=1$ эти уравнения имеют то же решение, что и уравнения для многоканальной системы с фиксированным числом каналов.

Интересно сравнить приведенное выше уравнение для $p_n(m)$ с соответствующим уравнением, полученным Романи. В обоих случаях выражения не зависят от n. Это означает, что независимо от того, какое заданное число каналов (за исключением единицы или M) действует, любое число требований в пределах от нуля до точки переключения равновероятно. Для системы с переменным числом каналов Филлипс ввел новый показатель эффективности — среднюю частоту прерывания. Этот показатель характеризует, как часто прекращается процесс в тех обслуживающих устройствах, которые выполняют вспомогательные работы. Этот показатель определяется с помощью времени, в течение которого система испытывает прерывание. В случае предельного числа каналов эта доля времени равна

$$\sum_{m=1}^{M-1} p_{N-1}(m).$$

Чтобы найти этот показатель, нужно умножить эту величину на λ и разделить на число обслуживающих каналов $M-1$.

$$\frac{\lambda}{M-1} \sum_{m=1}^{M-1} p_{N-1}(m) = \frac{\lambda \rho^{N-1}(1-\rho) p_0(0)}{(M-1)(1-\rho^N)} \sum_{m=1}^{M-1} \frac{\rho^m}{m!}. \qquad (13.79)$$

Данный показатель характеризует число прерываний на один канал в единицу времени.

Упражнение 8. При каких условиях для двух моделей системы с переменным числом каналов, описанных в этом параграфе, получим примерно одинаковый результат? (Ответ: Когда M и N велики, а загрузка ρ относительно мала.)

Упражнение 9. Каков смысл времени простоя в процессах с переменным числом каналов? [Ответ: Так как в системе с переменным числом каналов всегда предусматриваются вспомогательные работы, то простаивающие каналы могут быть только в системе с фиксированным числом каналов. Показателем времени простоя является $p_0(0)$.]

13.8. Объединение параллельных каналов

Фаген и Риордан [231] исследовали c-канальную систему с пуассоновским входящим потоком с параметром λ в стационарном состоянии при обслуживании требований в порядке поступления, когда все каналы могут объединяться для обслуживания многих потребностей клиента.

Рассматриваются следующие распределения времени обслуживания.

1. Эрланговское распределение:

$$B(t) = \int_0^t k \frac{(kx)^{k-1}}{(k-1)!} e^{-kx} dx, \quad t \geqslant 0, \quad k = 1, 2, \ldots$$

При $k=1$ имеем экспоненциальное распределение, а при $k=\infty$ — вырожденное распределение $B(t)=\begin{cases}0, & t<1,\\ 1, & t\geqslant 1.\end{cases}$

2. Равномерное распределение:

$$B(t)=\begin{cases}0, & t<1-a,\\ \dfrac{t-1+a}{2a}, & 1-a\leqslant t\leqslant 1+a, \quad 0<a<1,\\ 1, & t>1+a,\end{cases}$$

которое при $a\to 0$ стремится к вырожденному распределению времени обслуживания.

Как только заканчивается обслуживание в канале с наименьшим временем обслуживания, следующее требование обслуживается всеми каналами, и, следовательно, его время обслуживания является наименьшим. Каналы имеют независимые одинаково распределенные длительности обслуживания с функцией распределения $B(t)$. Вероятность того, что в момент времени t в каком-либо канале не окончится обслуживание требования, равна $1-B(t)$, и вследствие того, что для всех каналов эти события взаимно независимы, $[1-B(t)]^c$ есть вероятность того, что в момент времени t ни в одном канале обслуживание не закончится. Поэтому $1-[1-B(t)]^c=H(t)$ представляет собой функцию распределения времени обслуживания одного требования. Теперь можно рассмотреть задачу для одноканальной системы с приведенным выше распределением времени обслуживания.

Обозначим через μ интенсивность обслуживания, $\rho=\dfrac{\lambda}{\mu}$. Пусть $P(w)$ — вероятность того, что время ожидания в очереди не меньше w; введем преобразования Лапласа—Стилтьеса:

$$\gamma(s)=\int_0^\infty e^{-sw}dP(w),$$

$$\tilde{H}(s)=\int_0^\infty e^{-st}dH(t).$$

Тогда, применив уравнение (9.9), получим

$$\gamma(s)=(1-\rho)\left\{1-\frac{\mu\rho}{s}[(1-\tilde{H}(s)]\right\}^{-1}. \qquad (13.80)$$

Заметим, что вероятность того, что время пребывания в системе меньше w, есть распределение суммы двух случайных величин, первой из которых является время, проведенное в очереди, а второй — время обслуживания. Преобразование Лапласа—Стилтьеса функции распределения этой суммы (которое является сверткой) равно произведению исходных преобразований, т. е. $\gamma(s)\tilde{H}(s)$.

Для случая $k=1$ имеем $H(t)=1-e^{-ct}$ и без труда получаем

$$\gamma(s)\tilde{H}(s)=(c-\lambda)(c-\lambda+s)^{-1}. \qquad (13.81)$$

Следовательно, вероятность того, что время пребывания в системе меньше w, равна

$$1-e^{-(c-\lambda)w}. \qquad (13.82)$$

Отсюда j-й момент распределения равен $\mu_j=j!\ (c-\lambda)^{-j}$. Особый интерес представляет случай $j=1$, который дает среднее время ожидания, равное $1/(c-\lambda)$. Время ожидания вычислено также для эрланговского распределения при $k=2$.

Среднее время обслуживания одного требования в многоканальной системе с пуассоновским входящим потоком и экспоненциальным временем обслуживания, в которой каналы обслуживают одновременно различные категории требований, равно

$$1+\frac{C(c,\lambda)}{c-\lambda},$$

где

$$C(c,\ \lambda)=\frac{\lambda_n}{(c-\lambda)(c-1)!}\times$$

$$\times\left[1+\lambda+\ \ldots\ +\frac{\lambda^{c-1}}{(c-1)!}+\frac{\lambda^c}{(c-\lambda)(c-1)!}\right]^{-1}. \qquad (13.83)$$

Произведем сравнение среднего времени ожидания, взяв отношение приведенного выше времени ожидания к $(c-\lambda)^{-1}$. Среднее время ожидания для раздельного процесса, т. е. когда несколько каналов действуют независимо, больше, чем среднее время ожидания в том случае, когда несколько каналов производят выполнение многих операций обслуживания одного требования. При $k=2$ это сравнение провести трудно, если значение λ превосходит критическое, но на практике при таких значениях λ предпочтительно иметь раздельное обслуживание. Для значений λ, меньших критического, более желательно выполнение нескольких операций. Это доказательство можно распространить на другие случаи эрланговского распределения, а также и для равномерного распределения. При выполнении нескольких операций наименее желательным является постоянное время обслуживания, наибольший же выигрыш получается в случае экспоненциального или равномерного распределения времени обслуживания при $a=1$. Чем больше вероятность того, что время обслуживания мало, тем выгоднее применять обслуживание одного требования несколькими каналами.

13.9. Система со специальным обслуживанием

Кёнигсберг [452] исследовал систему, в которой из c параллельных каналов выделяется подгруппа c_1, которая может производить специальный вид обслуживания. Поступающие требования либо нуждаются в специальном обслуживании в c_1 каналах, либо им требуется обычное обслуживание в любом из c каналов. Поступление на обслуживание происходит по принципу «первым пришел — первым обслужен». Не нуждающееся в специальном обслуживании требование, которое стоит первым в очереди, может поступить в один из остающихся $c-c_1$ каналов, если последние свободны, даже в том случае, если требование, которое нуждается в специальном обслуживании, может ожидать поступления на обслуживание в один из c_1 каналов. Разумеется, требование, не нуждающееся в специальном обслуживании, может поступить в канал, производящий специальное обслуживание, если оно является первым в очереди и такой канал свободен. Заметим, что в данном случае требования не обладают приоритетами.

Эти положения можно обобщить на случай, когда производится несколько видов специального обслуживания и для этого выделяется некоторое число каналов. В настоящем параграфе эти положения не получают развития, достаточного для того, чтобы составить полное представление о проблеме. Однако читатель без труда сможет написать уравнения для системы с пуассоновским входящим потоком с интенсивностью поступления требований, нуждающихся в специальном обслуживании, равной $r\lambda=\lambda_1$, и интенсивностью поступления требований, не нуждающихся в специальном обслуживании, равной $(1-r)\lambda=\lambda_2$, и экспоненциальным временем обслуживания. Здесь можно воспользоваться исследованиями уравнений для системы с приоритетом, прерывающим обслуживание. Кёнигсберг также предпринял попытку сформулировать эту задачу для системы с постоянным временем обслуживания.

13.10. Параллельные каналы с различным временем обслуживания

Допустим, что в системе с несколькими параллельными каналами, пуассоновским входящим потоком с параметром λ при обслуживании требований в порядке поступления каждый канал имеет различное экспоненциальное время обслуживания. Для простоты допустим, что имеется два канала с параметрами μ_1 и μ_2 соответственно для первого и второго каналов. Пусть $P_n(t)$ — вероятность того, что в момент времени t в системе находится n требований при $n\geqslant 2$. При $n=1$ требование может находиться в первом или втором обслуживающем канале. Обозначим через $P_1(1, 0, t)$ вероятность того, что если в момент времени t в системе имеется одно требование, то оно находится в первом канале, а во втором канале требований нет. Аналогично $P_1(0, 1, t)$ показывает наличие требования во втором канале, когда первый канал свободен. Когда оба канала свободны, то поступающее требование с вероятностью $\frac{\mu_1}{\mu_1+\mu_2}$ направляется в первый канал и с вероятностью $\frac{\mu_2}{\mu_1+\mu_2}$ — во второй. Если свободен только один канал, то требование немедленно направляется туда на обслуживание. Другое допущение заключается в том, что когда оба канала свободны, требование поступает в любой канал с вероятностью, равной $^1/_2$. Оно может также всегда поступать в первый канал. Уравнения, описывающие систему, имеют вид:

$$P_0'(t) = -\lambda P_0(t) + \mu_1 P_1(1,\ 0;\ t) + \mu_2 P_1(0,\ 1;\ t),$$

$$P_1'(1,\ 0;\ t) = -(\lambda+\mu_1) P_1(1,\ 0;\ t) + \frac{\mu_1}{\mu_1+\mu_2}\lambda P_0(t) + \mu_2 P_2(t),$$

$$P_1'(0,\ 1;\ t) = -(\lambda+\mu_2) P_1(0,\ 1;\ t) + \frac{\mu_1}{\mu_1+\mu_2}\lambda P_0(t) + \mu_1 P_2(t),$$

$$P_2'(t) = -(\lambda+\mu_1+\mu_2) P_2(t) + \lambda[P_1(1,\ 0;\ t) + P_1(0,\ 1;\ t)] + \\ + (\mu_1+\mu_2) P_3(t), \tag{13.84}$$

$$P_n'(t) = -(\lambda+\mu_1+\mu_2) P_n(t) + \lambda P_{n-1}(t) + (\mu_1+\mu_2) P_{n+1}(t),\ n>2.$$

Положим $\mu_1=\mu_2=\mu$ и воспользуемся соотношением $P_1(t) = P_1(1,\ 0;\ t) + P_1(0,\ 1;\ t)$. Получим классические уравнения для системы с двумя параллельными каналами.

Уравнения для стационарного состояния найдем, приравняв нулю производные по времени. Эти уравнения справедливы, если выполняется соотношение

$$\rho = \frac{\lambda}{\mu_1+\mu_2} < 1.$$

Если длительность промежутка времени T стремится к бесконечности, то среднее число обслуженных требований будет сближаться со средним числом λT требований, поступивших за время T. Средняя длительность промежутка, в течение которого первый канал действует, равна

$$T[1-p_0-p_1(0,\ 1)]. \tag{13.85}$$

Среднее число требований, обслуженных за время T, составляет

$$\mu_1 T[1-p_0-p_1(0,\ 1)]. \tag{13.86}$$

Для второго канала это число равно

$$\mu_2 T[1-p_0-p_1(1,\ 0)]. \tag{13.87}$$

На основании изложенного ранее имеем

$$\lambda = (\mu_1+\mu_2)(1-p_0) - \mu_1 p_1(0,\ 1) + \mu_2 p_1(1,\ 0). \tag{13.88}$$

Данное уравнение совместно с первым уравнением системы, полученной для стационарного состояния, приводит к выражению

$$p_1 = (1-\rho)(1-p_0),$$

где $p_1 = p_1(1,\ 0) + p_1(0,\ 1)$.

Другие уравнения для стационарного состояния имеют вид $p_n = \rho p_{n-1}$. После решения их получим $p_n = (1-\rho)\rho^{n-1}(1-p_0)$. Среднее число требований, находящихся в системе, равно $L = \frac{1-p_0}{1-\rho}$, где p_0 определяется обычным способом.

13.11. Перемещение обслуживающего канала

Макмиллан и Риордан [557] рассмотрели следующую задачу. Обслуживающее устройство перемещается совместно со сборочной линией, которая сама движется с постоянной скоростью вместе с изделиями, размещенными на ней для обслуживания. Как только обслуживающее устройство заканчивает обслуживание изделия, оно без потери времени возвращается к следующему изделию. Линия имеет барьер, в котором обслуживающее устройство «поглощается», т. е. обслуживание прекращается, как только устройство перемещается за этот барьер. Поэтому желательно, чтобы оно без потери времени возвращалось в начальную точку. Допустим, что время обслуживания имеет экспоненциальное распределение с параметром μ. Обозначим через $p(k,\ T)$ вероятность того, что закончено обслуживание k изделий, если обслуживание первого изделия началось в тот момент, когда обслуживающее устройство находилось на удалении T временны́х единиц от барьера.

Карлин, Миллер и Прабху [413] показали, что эта задача аналогична задаче для одноканальной системы с пуассоновским входящим потоком и распределением времени обслуживания $F_s(t) = B(t)$. Заметим, что $B(t)$ соответствует распределению промежутков времени между изделиями. В системе с одним движущимся обслуживающим устройством время, протекающее до поглощения, эквивалентно длительности периода занятости в обычной системе, в которой распределение времени обслуживания требования, стоящего в очереди первым, равно $F_T(t) = \begin{cases} 1, & t > T, \\ 0, & t \leqslant T. \end{cases}$ Число требований, обслуженных до поглощения, на единицу меньше числа требований, обслуженных за период занятости, в котором первое требование обслуживается в соответствии с функцией распределения $F_T(t)$. Введем обозначение

$$P(x,\ T) = \sum_{k=0}^{\infty} p(k,\ T)\, x^k. \tag{13.89}$$

Пусть f_j — вероятность того, что j требований будут обслужены за период занятости, в котором первое требование имеет распределение времени обслуживания F_s, и пусть

$$F(x)=\sum_{j=1}^{\infty} f_j x^j. \qquad (13.90)$$

Если за время обслуживания первого требования T поступит n требований, то распределение вероятностей того, что эти требования будут обслужены за оставшуюся часть периода занятости, есть n-кратная свертка $\{f_j\}$. Таким образом,

$$P(x, T)=\sum_{n=0}^{\infty} \frac{e^{-\mu T}(\mu T)^n}{n!}[F(x)]^n = e^{-\mu T[1-F(x)]}. \qquad (13.91)$$

Из рассмотрения в гл. 8 распределения длительности периода занятости известно, что при $|x|\leqslant 1$ функция $F(x)$ является единственным аналитическим решением интегрального уравнения

$$F(x)=x\int_0^{\infty} e^{-\mu t[1-F(x)]}\,dB(t) \qquad (13.92)$$

при $F(0)=0$; поэтому $P(x, t)$ является единственным решением уравнения

$$P(x, T)=\exp\left\{-\mu T\left[1-x\int_0^{\infty} P(x, t)\,dB(t)\right]\right\} \qquad (13.93)$$

при $|P(x, T)|\leqslant 1$, $|x|\leqslant 1$ для всех t. Для входящего потока с постоянными промежутками времени между моментами поступления требований, равными ε временны́м единицам,

$$B(t)=\begin{cases} 1, & t>\varepsilon, \\ 0, & t\leqslant\varepsilon, \end{cases}$$

положив в последнем уравнении $T=\varepsilon$, получим

$$P(x, \varepsilon)=e^{-\mu\varepsilon[1-xP(x, \varepsilon)]}. \qquad (13.94)$$

Отсюда Макмиллан и Риордан получили разложение Лагранжа

$$P(x, T)=e^{-\mu T}+\sum_{k=1}^{\infty}\frac{T(T+k\varepsilon)^{k-1}}{k!}e^{-\mu T}(\mu e^{-\mu\varepsilon})^k x^k. \qquad (13.95)$$

Следовательно, мы имеем $p(k, T)$.

Математическое ожидание равно

$$L(T)=\sum kp(k, T)=\mu T(1-\mu\varepsilon)^{-1},\ \mu\varepsilon<1, \qquad (13.96)$$

а дисперсия —

$$\sigma^2(T)=\mu T(1-\mu\varepsilon)^{-3}. \qquad (13.97)$$

Если $B(t) = 1 - e^{-\lambda t}$, то $P(x, T)$ найдем, интегрируя обе части уравнения по $dB(T)$ и вычисляя $\int\limits_0^\infty P(x, T)\,dB(t)$; это выражение равно

$$\frac{\mu + \lambda - \sqrt{(\mu + \lambda)^2 - 4\mu\lambda x}}{2\mu x}. \tag{13.98}$$

Таким образом,

$$P(x, T) = \exp\left[-\left(\frac{T}{2}\right)(\mu - \lambda) + \sqrt{(\mu + \lambda)^2 - 4\mu\lambda x}\right]. \tag{13.99}$$

В этом случае математическое ожидание равно

$$L(T) = \frac{\mu T}{1 - \frac{\mu}{\lambda}}, \tag{13.100}$$

а дисперсия —

$$\sigma^2(T) = \frac{\left[1 - \left(\frac{\mu}{\lambda}\right)^2\right]\mu T}{\left(1 - \frac{\mu}{\lambda}\right)^3}. \tag{13.101}$$

Пусть $G(u, T)$ — вероятность того, что обслуживающее устройство поглощается до момента времени u. Карлин и др. получили решение в виде преобразования Лапласа—Стилтьеса этой функции:

$$\tilde{G}(s, T) = \exp\left\{-sT - \mu T\left[1 - \int\limits_0^\infty \tilde{G}(s, t)\,dB(t)\right]\right\}. \tag{13.102}$$

Для первого случая, приведенного выше, эта функция имеет вид

$$e^{-(s+\mu)T} + \sum_{k=1}^{\infty} \frac{T(T + k\varepsilon)^{k-1}}{k!} e^{-\mu T}(\mu e^{-\mu\varepsilon})\, e^{-s[T+k\varepsilon]} \tag{13.103}$$

и для второго случая —

$$\exp\left\{-\frac{T}{2}\left[s + \mu - \lambda + \sqrt{(s + \mu + \lambda)^2 - 4\mu\lambda}\right]\right\}. \tag{13.104}$$

13.12. Циклические системы массового обслуживания

Кёнигсберг [453] исследовал добычу угля, когда забой угольной шахты разрабатывается врубовой машиной, отбойными молотками, бригадой взрывников и т. д. После окончания работы в таком же порядке возобновляется работа в другом забое, а после окончания работы во всех N забоях вновь начинается работа в первом забое. В этой задаче можно рассматривать ряд последовательных систем массового обслуживания, обслуживающих N требований (забоев), при обслуживании в порядке поступления, когда требование, покидающее последнюю фазу, ожидает обслуживания в первой фазе, поступая последовательно (в очереди и на обслуживание) в каждую фазу; отсюда и название — циклические системы.

Эти системы исследовал Финч [243]. При этом он принял допущение, что требование после окончания обслуживания в последней фазе с определенной вероятностью p_j возвращается в j-ю фазу. Рассматривается еще один случай, когда существует вероятность $p_j<1$ того, что сразу же после окончания обслуживания в j-й фазе требование поступает в очередь этой фазы. Однако из определения p_j видно, что существует вероятность того, что требование может поступить и в следующую фазу.

1. Циклические системы без обратной связи

Пусть n_j — число требований, ожидающих в очереди и находящихся на обслуживании в j-й фазе, если общее число фаз равно k. Тогда $\sum_{j=1}^{k} n_j = N$.

Распределение времени обслуживания в j-й фазе, которая состоит из одного обслуживающего канала, является экспоненциальным с интенсивностью μ_j; таким образом, $\mu_j \Delta t$ — вероятность того, что в промежутке времени $(t,\ t+\Delta t)$ будет закончено обслуживание в j-й фазе. Уравнения, описывающие систему через вероятности того, что в момент времени t в j-й фазе $(j=1,\ \ldots,\ k)$ находится n_j требований, имеют вид

$$\begin{aligned} P(n_1,\ \ldots,\ n_k;\ t+\Delta t) &= [1-(\mu_1+\ldots+\mu_k)\Delta t]\times \\ &\times P(n_1, \ldots, n_k;\ t)+\mu_1 \Delta t P(n_1+1,\ n_2-1, n_3, \ldots,\ n_k;\ t)+ \\ &+\mu_2 \Delta t P(n_1,\ n_2+1,\ n_3-1,\ n_4, \ldots,\ n_k;\ t)+\ldots+ \\ &+\mu_k \Delta t P(n_1-1,\ n_2, \ldots,\ n_k+1;\ t). \end{aligned} \qquad (13.105)$$

После упрощения имеем

$$\begin{aligned} P'(n_1, \ldots,\ n_k;\ t) &= -\sum_{j=1}^{k} \mu_j P(n_1, \ldots,\ n_k;\ t)+ \\ &+\sum_{j=1}^{k} \mu_j P(n_1, \ldots, n_j+1,\ n_{j+1}-1, \ldots,\ n_k;\ t). \end{aligned} \qquad (13.106)$$

Если для некоторых m имеем $n_m=0$, то m-й член не появляется, т. е. он равен нулю, и если $n_{m+1}-1<0$, то m-й член также равен нулю. Заметим, что k-я фаза связана с первой фазой, т. е. возможны переходы из состояния $(n_1-1,\ \ldots,\ n_k+1)$ в состояние $(n_1,\ \ldots,\ n_k)$. Поскольку вся совокупность распределена между фазами, то не имеет смысла говорить о поступлении требований. Уравнения для стационарного состояния находятся из приведенных выше уравнений обычным путем. Имеем следующее решение этой системы уравнений:

$$p(n_1, \ldots, n_k) = \frac{\mu_1^{N-n}}{\mu_2^{n_2} \ldots \mu_k^{n_k}} p(N, 0, \ldots, 0). \tag{13.107}$$

Последнюю величину в правой части этого выражения найдем, просуммировав левую часть по всем разбиениям $N(n_1, \ldots, n_k)$ и приравняв эту сумму единице. Если не будет оговорено противное, то встречающиеся далее суммы понимаются, как распространенные на все N. Заметим, что первоначально вся совокупность ожидает обслуживания в первой фазе. Заметим также, что ту долю времени D_j в течение которого свободно обслуживающее устройство в j-й фазе, найдем, полагая в $p(n_1, \ldots, n_k)$ значение n_j равным нулю и суммируя по всем разбиениям N, в которых $n_j = 0$. После вычитания из единицы получим вероятность того, что обслуживающее устройство j-й фазы занято. Таким образом, $1 - D_j$ есть загрузка обслуживающего устройства j-й фазы.

Введем обозначение $x_j = \frac{\mu_1}{\mu_j}$, тогда $x_1 = 1$; получим

$$p(N, 0, \ldots, 0) = \frac{1}{\sum x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}} = \left[\sum_{j=1}^{k} \frac{x_j^{N+k-1}}{\prod\limits_{\substack{m=1 \\ m \neq j}}^{k} (x_j - x_m)} \right]^{-1}. \tag{13.108}$$

Кёнигсберг вычислил обратное значение последней величины в правой части (13.108) при $k = 1, \ldots, 5$. Среднее число требований, находящихся в j-й фазе (включая обслуживаемое), равно

$$\begin{aligned} \bar{n}_j &= \sum n_j p(n_1, \ldots, n_j, \ldots, n_k) = \\ &= \frac{x_j}{\sum x_1^{n_1} \ldots x_k^{n_k}} \frac{d}{dx_j} \sum x_1^{n_1} \ldots x_k^{n_k}. \end{aligned} \tag{13.109}$$

Среднее число требований, ожидающих обслуживания в j-й фазе, равно

$$\sum (n_j - 1) p(n_1, \ldots, n_k) = (\bar{n}_j - 1 + D_j). \tag{13.110}$$

Здесь сумма берется по всем разбиениям N, в которых $n_j \geqslant 1$. Имеем

$$D_j = \sum p(n_1, n_2, \ldots, n_{j-1}, 0, n_{j+1}, \ldots, n_k). \tag{13.111}$$

Если $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$, то $x_j = 1$ и $p(N, 0, \ldots, 0) = \frac{(N+k-1)!}{(k-1)!\,N!}$, $D_j = \frac{k-1}{N+k-1}$ (не зависит от j и, следовательно, имеет одно и то же значение для всех j), $\bar{n}_j=\frac{N}{k}$ (не зависит от j и, следовательно, имеет одно и то же значение для всех j). Среднее число требований, ожидающих в j-й фазе, равно $\frac{N(N-1)}{k(N+k-1)}$.

Среднее время ожидания в j-й фазе найдем, разделив последнюю величину на μ. Время, необходимое для прохождения всех фаз, т. е. для завершения всего цикла, равно

$$\frac{1}{\mu}\left[k+\frac{N(N-1)}{N+k-1}\right]. \qquad (13.112)$$

Та доля времени, в течение которой занято обслуживающее устройство j-й фазы (т. е. вероятность того, что оно занято), составляет

$$(1-D_j)\mu_j=\frac{N}{N+k-1}\mu. \qquad (13.113)$$

Если μ_j — интенсивность выходящего потока и общее время действия H известно, то общее число выходящих требований можно найти, умножив последнее выражение на H.

Чтобы проиллюстрировать частный случай при $m=j$, в выражении для $p(N, 0, \ldots, 0)$ положим $x_1=x_2=x_3$ при $k=4$. Тогда, рассуждая аналогично, при равных интенсивностях обслуживания получим

$$p(N, 0, 0, 0)=\left[\sum_{n_4=0}^{N} x_4^{n_4}\frac{(N+k-n_4-2)!}{(k-2)!(N-n_4)!}\right]^{-1}. \qquad (13.114)$$

Если, например, j-я фаза включает в себя два параллельных канала с одинаковыми значениями μ_j, то μ_j заменяется на $2\mu_j$ при $n\geqslant 2$, в противном случае уравнения не изменяются. Таким образом,

$$p(n_1, \ldots, n_k)=x_1^{n_1}\ldots\frac{x_j^{n_j}}{2!\,2^{n_j-2}}\ldots x_k^{n_k}p(N, 0, \ldots, 0). \qquad (13.115)$$

При вычислении $p(N, 0, \ldots, 0)$ в предыдущей формуле x_j заменяется на $\frac{x_j}{2}$.

2. Циклические системы с обратной связью

Конечная обратная связь. Случаи, к которым мы обратимся, аналогичны тем, которые были изложены выше [243]. Пусть $p_j<1\,(j=1, \ldots, k)$ — вероятность того, что после завершения обслуживания в последней, k-й, фазе требование возвращается в j-ю фазу, $p=\sum_{j=1}^{k}p_j$, а $q=1-p$ — вероятность того, что требование не возвращается в систему. Такой процесс обслуживания называется обслуживанием с конечной обратной связью. Пусть

$$\varepsilon_j = \begin{cases} 1, & n_j > 0, \\ 0, & n_j = 0, \end{cases} \quad a = \begin{cases} 1, & \sum\limits_j n_j < N, \\ 0, & \sum\limits_j n_j = N. \end{cases} \tag{13.116}$$

При $n_j = 0, \ldots, N$ и $\sum\limits_{j=1}^{k} n_j \leqslant N$, уравнения для вероятностей состояния (в стационарном случае) можно записать в следующем виде:

$$\begin{gathered} 0 = -\left(\lambda a + \sum_{j=1}^{k} \mu_j \varepsilon_j\right) p(n_1, \ldots, n_k) + \\ + \lambda \varepsilon_1 p(n_1 - 1, n_2, \ldots, n_k) + \\ + \sum_{j=1}^{k-1} \mu_k p_j \varepsilon_j p(n_1, \ldots, n_{j-1}, n_j - 1, n_{j+1}, \ldots, n_{k-1}, n_k + 1) + \\ + \mu_k p_k \varepsilon_k p(n_1, \ldots, n_k) + \\ + \sum_{j=1}^{k-1} \mu_j \varepsilon_j p(n_1, \ldots, n_{j-1}, n_j + 1, n_{j+1} - 1, \ldots, n_{k-1}, n_k) + \\ + (1 - p) \mu_k a p(n_1, \ldots, n_{k-1}, n_k + 1). \end{gathered} \tag{13.117}$$

Эта система уравнений имеет единственное решение

$$p(n_1, \ldots, n_k) = x_1^{n_1} \ldots x_k^{n_k} p(0, \ldots, 0) \tag{13.118}$$

при

$$x_j = \frac{\lambda(1 - p + p_1 + \ldots + p_j)}{\mu_j(1 - p)}, \quad j = 1, \ldots, k. \tag{13.119}$$

Снова находим $p(0, \ldots, 0)$, приравнивая единице сумму вероятностей по $n_j = 1, \ldots, N$; $\sum n_j \leqslant N$.

Если $p_j = 0$ для всех j, $\frac{\lambda}{\mu_j} < 1$ и $N \to \infty$, то получим решение, которое приводит Р. Джексон для системы с последовательными каналами. Если $x_j < 1$ $(j = 1, \ldots, k)$, то при $N \to \infty$ значение $p(0, \ldots, 0)$ стремится к конечному пределу, отличному от нуля; в противном случае оно стремится к нулю, и, следовательно, $p(n_1, \ldots, n_k)$ также стремится к нулю. В последнем случае имеем

$$v(0, \ldots, 0) = \prod_{j=1}^{k} (1 - x_j)$$

и

$$p(n_1, \ldots, n_k) = \prod_{j=1}^{k} x_j^{n_j}(1 - x_j). \tag{13.120}$$

Одноканальная система с обратной связью. Здесь через $p_j < 1$ обозначена вероятность того, что после окончания обслуживания в j-й фазе требование возвращается в очередь j-й фазы [введем также $q_j=1-p_j$ — вероятность того, что требование переходит в $(j+1)$-ю фазу или покидает систему, если $j=1$]. Тогда уравнения для равновесного состояния будут иметь вид

$$0=-\left(\lambda a+\sum_{j=1}^{k}\mu_j q_j \varepsilon_j\right)p(n_1,\ldots,n_k)+\lambda\varepsilon_1 p(n_1-1,n_2,\ldots,n_k)+$$
$$+\sum_{j=1}^{k-1}\mu_j q_j \varepsilon_{j+1} p(n_1,\ldots,n_j+1,n_{j+1}-1,\ldots,n_k)+$$
$$+\mu_k q_k a p(n_1,\ldots,n_k+1). \qquad (13.121)$$

Имеем единственное решение

$$p(n_1,\ldots,n_k)=x_1^{n_1}\ldots x_k^{n_k}p(0,\ldots,0), \qquad (13.122)$$

где

$$x_j=\frac{\lambda}{\mu_j q_j}. \qquad (13.123)$$

Если $x_j<1$, то при $N\to\infty$ рассуждаем, как и в предыдущем случае. Полученные ранее уравнения справедливы и при использовании только что выведенного выражения для x_j. В случае одноканальной системы оба рассмотренные выше случая совпадают.

13.13. Циклическое обслуживание в системе с несколькими очередями

Рассмотрим случай, когда имеется N пронумерованных очередей $(1, 2, \ldots, N)$ и одно обслуживающее устройство; это устройство вначале обслуживает все требования, которые к моменту его прибытия находились в первой очереди, а затем перемещается ко второй очереди и т. д. до N-й очереди, после чего снова возвращается к первой очереди. Все требования, которые поступают в очередь после того, как обслуживающее устройство уже начало обслуживать эту очередь, должны ожидать обслуживания не в этом цикле, а в следующем. Таким образом, только те требования, которые находятся в очереди к моменту прибытия обслуживающего устройства, обслуживаются в течение этого цикла. Допускается, что имеется запаздывание во времени между окончанием обслуживания одной очереди и началом обслуживания следующей. Это запаздывание между двумя последовательными очередями описывается с помощью N случайных величин, которые имеют одинаковое распределение с плотностью $w(t)$, определяющей это время запаздывания или перемещения. Допустим также, что если $N=1$, то имеется запаздывание в обслуживании уже имеющихся требований и тех, которые поступают во время обслуживания.

Пусть потоки, поступающие во все очереди, являются пуассоновскими с одинаковым параметром λ, и пусть время обслуживания всех требований во всех очередях имеет одинаковое распределение, описываемое функцией плотности $b(t)$. Лейбович [500] вычислил для стационарного состояния вероятность π_n того, что в момент прибытия обслуживающего устройства к этой очереди в ней будет находиться n требований.

Эти положения применимы к сообщениям, которые поступают на выходные устройства вычислительной машины и хранятся (ожидают) в запоминающем устройстве, а затем обслуживаются (сообщения передаются для обработки на вычислительной машине) в описанном выше порядке. Данная постановка задачи применима также к сортировке по ячейкам писем на почтамте. Поступление писем в каждую ячейку имеет пуассоновское распределение. Служащий берет письма, которые находятся в этой ячейке, и, прочитав адрес, опускает их в соответствующий мешок для транспортировки. Покончив с этой ячейкой, служащий переходит к следующей, не обращая внимания на то, что в ячейку, которую он только что обслужил, могут поступить новые письма. Эти письма будут рассортированы во время следующего цикла.

Положим $N=1$. Пусть r_{in} — вероятность того, что после прибытия обслуживающего устройства в очереди находится n требований при условии, что при предыдущем прибытии было i требований. Тогда

$$\pi_n = \sum_{i=0}^{\infty} r_{in}\pi_i. \qquad (13.124)$$

Заметим, что плотность вероятности времени обслуживания i требований есть n-кратная свертка функции $b(t)$. Обозначим ее через $b_i(t)$. Плотность суммы времени обслуживания i требований и временной задержки перед обслуживанием следующей группы требований есть свертка функций $b_i(t)$ и $\omega(t)$, что мы обозначим символом $\operatorname{conv}[b_i(t), \omega(t)]$. Как и для системы с групповым поступлением требований и групповым обслуживанием, имеем

$$r_{in} = \int_0^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \operatorname{conv}[b_i(t), \omega(t)]\, dt. \qquad (13.125)$$

Положим

$$P(z) = \sum_{n=0}^{\infty} \pi_n z^n \qquad (13.126)$$

и обозначим через $b^*(s)$ и $\omega^*(s)$ преобразования Лапласа этих функций. Затем подставляя (13.124) и (13.125) в (13.126), изменяя порядок суммирования и упрощая, получим

$$P(z) = \omega^*[\lambda(1-z)]\, P\{b^*[\lambda(1-z)]\}. \qquad (13.127)$$

Дифференцируя $P(z)$ и положив $z=1$, можно найти математическое ожидание и дисперсию распределения π_n. Например, математическое ожидание равно $\frac{\lambda\bar{\omega}}{\left(\frac{1-\lambda}{\mu}\right)}$, где $\bar{\omega}$ — среднее время перемещения, а $\frac{1}{\mu}$ — среднее время обслуживания. Заметим, что вычисления упрощаются, если выражение (13.127) прологарифмировать. Для существования стационарного состояния необходимо, чтобы выполнялось условие $\frac{\lambda}{\mu} < 1$.

Упражнение 10. Вычислите дисперсию распределения π_n.

Если $N>1$, то для нахождения π_n для конкретной очереди необходимо знать вероятность того, что во второй очереди находится n_2 требований, в третьей — n_3 и т. д. Таким образом, в данном случае π_n не является стационарной вероятностью для марковского процесса. Рассмотрение этого случая с помощью матриц перехода приводит к сложным результатам. Приближенно можно считать, что одна и та же вероятность π_n описывает состояние каждой очереди в момент прибытия обслуживающего устройства. Это допущение приводит к поправке по отношению к математическому ожиданию точной задачи и к исправленному разложению π_n в ряд по степеням $\frac{\lambda}{\mu}$ до члена, содержащего $\left(\frac{\lambda}{\mu}\right)^3$.

В этом случае правая часть (13.127) возводится в N-ю степень, при этом $P(z)$ определяется для нового значения π_n. Имеем

$$P(z) = (\omega^* [\lambda(1-z)] P\{b^* [\lambda(1-z)]\}^N. \tag{13.128}$$

Основанием для этого является то, что за время цикла число требований n, ставших в рассматриваемую очередь, равно сумме

$$n = \delta_l + \ldots + \delta_N,$$

где δ_k — число поступлений в данную очередь за время обслуживания требований k-й очереди. Так как (13.127) есть производящая функция распределения вероятностей появления случайной величины δ_k, то для нахождения (13.128) ее нужно возвести в N-ю степень. В этом случае математическое ожидание равно $\frac{N\lambda\bar{\omega}}{\left[1-N\left(\frac{\lambda}{\mu}\right)\right]}$, и для стационарного процесса должно выполняться условие $\frac{N\lambda}{\mu} < 1$.

Упражнение 11. Вычислите дисперсию.

При $z=1$ получим разложение в ряд с точностью до членов порядка $O(\lambda^3)$, $(\mu=1)$:

$$\log P(z) = \frac{-N\lambda}{1-N\lambda}(1-z) + \left\{\frac{N^2\lambda^2\omega(t)\,\overline{b^2}}{1-N\lambda} + N\lambda^2\sigma^2[\omega(t)]\right\}\frac{(1-z)^2}{2} - $$
$$- N\lambda^3(\overline{\overline{\omega^3}} - 3\bar{\omega}\overline{\omega^2} + 2\bar{\omega}^3)\frac{(1-z)^3}{6}, \tag{13.129}$$

где $\overline{\omega^j}$ — j-й начальный момент для $\omega(t)$, а $\overline{b^2}$ — второй начальный момент для $b(t)$.

Для нахождения π_n показательная функция (13.129) дифференцируется по z. При $n>3$ имеем $\pi_n=0$.

Заметим, что при больших N для сохранения стационарного состояния значение $\frac{\lambda}{\mu}$ должно быть мало. При $\mu=1$, $\alpha=N\lambda$ получим

$$\log P(z)=\frac{\alpha\overline{\omega}}{1-\alpha}(1-z)+\frac{1}{N}\left\{\frac{\alpha^2\overline{b^2}\overline{\omega}}{1-\alpha}+\alpha^2\sigma^2[\omega(t)]\right\}\times$$
$$\times\frac{(1-z)^2}{2}+\ldots \qquad (13.130)$$

Отсюда, разлагая $\log P(z)$ в ряд по степеням z, беря π_n как коэффициент при z^n $\left[\text{до членов, содержащих } O\left(\frac{1}{N}\right)\right]$ и обозначая через f коэффициент при $(1-z)$ в формуле (13.130), а через g — коэффициент в фигурных скобках, имеем

$$\pi_0=e^{-f}\left(1+\frac{g}{2N}\right),$$

$$\pi_1=e^{-f}\left(f+\frac{gf}{2N}-\frac{g}{N}\right),$$

$$\pi_n=e^{-f}\frac{f^n}{n!}+\frac{g}{N}e^{-f}\frac{f^{n-2}}{(n-2)!}\left[\frac{f^2}{2}-nf+\frac{n(n-1)}{2}\right],\quad n\geqslant 2. \qquad (13.131)$$

При $N\to\infty$ вероятность π_n имеет пуассоновское распределение с параметром λ.

Точное решение для $N=2$ находится следующим образом. Обозначим через π_{nm} вероятность того, что в момент прибытия обслуживающего устройства в первую очередь имеется n требований в первой очереди и m требований во второй. Обозначим через $r(n, m\mid i, j)$ вероятность совместного появления этих двух событий при условии, что в момент предыдущего прибытия обслуживающего устройства в первую очередь в ней находилось i требований, а во второй очереди было j требований. Тогда

$$\pi_{nm}=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}r(n, m\mid i, j)\,\pi_{ij}. \qquad (13.132)$$

Имеем

$$r(n, m\mid i, j)=\sum_{k=j}^{\infty}\int_0^{\infty}\int_0^{\infty}\left\{\operatorname{conv}[b_i(t_1), \omega(t_1)]\,e^{-\lambda t_1}\frac{(\lambda t_1)^{k-j}}{(k-j)!}\times\right.$$
$$\left.\times\operatorname{conv}[b_k(t_2), \omega(t_2)]\right\}e^{-\lambda(t_1+t_2)}\frac{[\lambda(t_1+t_2)]^n}{n!}e^{-\lambda t_2}\frac{(\lambda t_2)^m}{m!}\,dt_1dt_2, \qquad (13.133)$$

где величина в фигурных скобках — плотность вероятности того, что если в первой очереди находится i требований, а во второй j требований в момент, когда обслуживающее устройство начинает обслуживание первой очереди, то необходимо время t_1 для того, чтобы закончить обслуживание в первой очереди и перейти ко второй, в которой нужно будет обслужить k требований, и необходимо время t_2 для обслуживания второй очереди и возвращения к первой.

Для решения (13.132) вначале вводится производящая функция

$$Q(y, z) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \pi_{nm} x^n y^m. \qquad (13.134)$$

После некоторых преобразований приходим к уравнению

$$P(x, y) = \omega^*(u)\,\omega^*(v)\,P[b^*(u), b^*(v)], \qquad (13.135)$$

где

$$u = \lambda(2 - x - y) \text{ и } v = \lambda(1 - x) + \lambda + \lambda b^*(u).$$

Логарифмируя, дифференцируя и подставляя $x=1$ и $y=1$, находим, что математическое ожидание равно $\frac{2\bar{\omega}\lambda}{(1-2\lambda)}$.

Упражнение 12. Проверьте формулу (13.135) и выражение для математического ожидания. Найдите дисперсию.

ПРИЛОЖЕНИЯ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

14.1. Введение

В этой главе мы прежде всего рассмотрим приложения теории массового обслуживания к телефонным системам. Это даст возможность исследовать системы с потерями, так как до сих пор мы уделяли главное внимание системам с ожиданием. Затем рассмотрим приложения теории к движению автомобильного и воздушного транспорта. Иногда такие приложения, например, в § 14.4, носят общий характер и мало используют собственно теорию массового обслуживания. Приложение теории к обслуживанию станков позволяет исследовать случай, когда входящий поток образуется конечной совокупностью. С помощью теории массового обслуживания исследуются также вопросы обеспечения запасными частями. В другом параграфе приводится пример приложения теории массового обслуживания к управлению запасами.

Для инженеров представляет интерес проектирование системы плотин и создание хранилищ. Здесь рассматриваются и эти вопросы, так как многие важные положения этой области связаны с теорией массового обслуживания. Приводятся общие замечания относительно работы больниц и дается описание очередей, возникающих при оказании медицинской помощи. Подобным же образом рассматриваются вопросы, связанные с планировкой кафе. Затем показано приложение теории массового обслуживания к работе угольной шахты, где потери рабочего времени должны сводиться к минимуму. Приводится также модель системы массового обслуживания, описывающая шумы в полупроводниках. В конце главы рассматриваются приложения теории в других различных реальных и возможных случаях.

14.2. Приложения теории к телефонным системам

Кратко рассмотрим некоторые важные положения теории телефонной связи, предварительное описание которых дано в гл. 1. Заметим, что обычно в системах с ожиданием и системах с потерями существует полная доступность линий связи всем абонентам.

Однако, когда рассматривается ступенчатое включение и вводятся различные виды соединений, то полной доступности нет, и важное значение приобретает реально существующая структура системы.

Рассмотрим вначале полнодоступные системы с потерями, поскольку системы массового обслуживания с ожиданием уже рассматривались. Мы приведем известную формулу Эрланга для системы с потерями, затем расскажем о работе Воло, посвященной относительной интенсивности потока, а затем дадим краткое изложение работ Коэна. Случай повторных вызовов, исследованный Коэном, рассматривается в задаче, приведенной в конце главы.

1. Полнодоступная система

Формула Эрланга для системы с потерями. Пусть $P_n(t)$ — вероятность того, что занято ровно n из c каналов, если в момент времени $t=0$ все каналы были свободны. Пусть входящий поток исходит из неограниченного источника и следует закону Пуассона с параметром λ на единицу времени, а распределение времени обслуживания во всех каналах является экспоненциальным с одним и тем же параметром μ. Тогда обычным способом найдем систему уравнений

$$P'_n(t) = -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t),$$

$$1 \leqslant n \leqslant c-1,$$

$$P'_0(t) = -\lambda P_0(t) + \mu P_1(t),$$

$$P'_c(t) = -(\lambda + c\mu) P_c(t) + \lambda P_{c-1}(t).$$

Исследовав характеристические корни матрицы коэффициентов, можно показать, что существует решение для стационарного состояния. Находим, что оно имеет один нулевой корень. Так как решение является линейной комбинацией показательных функций, то один член будет постоянным; когда $t \to \infty$, он является единственным членом, который сохраняется; этот член является вероятностью стационарного состояния. Решение уравнений для стационарного состояния, полученных из приведенной выше системы, имеет вид

$$p_n = \frac{\dfrac{\rho^n}{n!}}{1 + \dfrac{\rho}{1!} + \dfrac{\rho^2}{2!} + \ldots + \dfrac{\rho^c}{c!}}, \quad \rho = \frac{\lambda}{\mu}. \qquad (14.1)$$

Это выражение является формулой Эрланга для системы с потерями. При $n=c$ данная формула описывает состояние перегрузки через временны́е зависимости. Вероятность того, что вызов теряется, равна вероятности того, что все каналы заняты (т. е. выражение, описывающее состояние полной занятости через временны́е зависимости, совпадает с выражением, описывающим состояние полной занятости через число вызовов в системе, что справедливо только в случае пуассоновского входящего потока). При $c \to \infty$ имеем закон Пуассона

$$p_n = \frac{\rho^n}{n!} e^{-\rho}.$$

Упражнение. Читателю предлагается написать уравнения для стационарного состояния и получить формулу Эрланга для системы с потерями.

Воло, Полячек, Пальм, Костен, Форте и Б. А. Севастьянов приводят различные доказательства того важного положения, что *вид формулы Эрланга не зависит от распределения времени обслуживания, если входящий поток является пуассоновским*[1].

В своей статье, посвященной обобщению формул Энгсета (см. ниже), Коэн приводит выражения, аналогичные формуле Эрланга для системы с потерями, с той лишь особенностью, что при этом он рассматривает конечную совокупность и произвольное распределение времени обслуживания, которое в действительности может быть различным для каждого источника.

Когда входящий поток поступает из конечной совокупности, то каждый элемент этой совокупности может рассматриваться как возможный источник входящего потока [обычно к источнику применяются допущения, вытекающие из закона Пуассона, т. е. вероятность того, что вызов появится в промежутке времени $(t, t+\Delta t)$ равна $\lambda\Delta t$, а вероятность того, что два вызова поступят одновременно, пренебрежимо мала].

Распределения Энгсета—О'Делла. Предположим, что в c-канальную систему поступает пуассоновский поток и что отношение интенсивности входящего потока к интенсивности обслуживания зависит от числа свободных каналов. Таким образом, y_0— значение указанной величины, когда нет занятых каналов, y_1 — та же величина в случае, когда занят один канал, и т. д. до y_c; здесь $y_i = \frac{\lambda_i}{\mu}$.

Пусть p_n $(n=0, \ldots, c)$ — вероятность того, что занято n каналов. Тогда для системы с одинаковым экспоненциальным временем обслуживания с параметром μ получим следующие уравнения:

$$\begin{gathered} y_0 p_0 = p_1, \\ (y_i + i) p_i = (i+1) p_{i+1} + y_{i-1} p_{i-1}, \quad i = 1, \ldots, c-1, \\ c p_c = y_{c-1} p_{c-1}. \end{gathered} \tag{14.2}$$

Эта система имеет решение

$$p_i = p_0 \frac{y_0 \ldots y_{i-1}}{i!}, \tag{14.3}$$

где p_0 находится из соотношения $\sum_{i=0}^{c} p_i = 1$.

---

[1] Из перечисленных авторов наиболее полное и математически безупречное доказательство принадлежит Б. А. Севастьянову. — *Прим. ред.*

Число потерянных вызовов в единицу времени составляет $y_c p_c$. Общее число вызовов (поступлений) в единицу времени равно $\sum_{i=0}^{c} y_i p_i = \sum_{i=1}^{c} i p_i + y_c p_c$. Доля потерянных вызовов равна частному от деления потерянных вызовов на общее число вызовов:

$$\frac{\frac{(y_1 \ldots y_c)}{c!}}{1 + y_1 + \frac{y_1 y_2}{2!} + \ldots + \frac{y_1 \ldots y_c}{c!}} \tag{14.4}$$

и не зависит от y_0. Положив $y_i = y$ $(i = 1, \ldots, c)$, получим формулу Эрланга.

Принимая, что число источников N конечно, и положив $y_i = (N - i)\lambda$, получим важное приложение этой формулы. В зависимости от того, зависит ли λ от потерь или же является постоянной величиной, получим распределение Энгсета или распределение О'Делла. Знаменатель конечного выражения можно упростить, выразив его для удобства вычисления через неполную бета-функцию, после чего приходим к выражению

$$(1+\lambda)^{N-1} \frac{\int\limits_{\frac{\lambda}{\lambda+1}}^{1} w^c (1-w)^{N-c-2} dw}{\int\limits_{0}^{1} w^c (1-w)^{N-c-2} dw}.$$

Если $N \to \infty$ и при этом $\lambda N = \mathrm{const}$, то формулы О'Делла превращаются в формулы Эрланга. Это положение справедливо и для формулы Энгсета.

Системы смешанного типа с приоритетами. Коэн [138] рассмотрел полнодоступную c-канальную систему, в которую поступают два пуассоновских потока с параметрами λ_1 и λ_2. Каналы имеют одинаковое экспоненциальное время обслуживания с одним и тем же параметром μ. Если все каналы заняты, то вызовы из первого источника могут ожидать, а вызовы из второго источника теряются.

Пусть p_n при $n \leqslant c$ — вероятность того, что занято n каналов, а при $n > c$ — вероятность того, что ожидает $n - c$ требований из первого источника. Тогда обычным способом при $\rho_1 \equiv \frac{\lambda_1}{\mu} < c$ и $\rho_2 \equiv \frac{\lambda_2}{\mu} < c$ получим уравнения для стационарного распределения:

$$\begin{gathered}(n+1)p_{n+1} + (\rho_1 + \rho_2)p_{n-1} - (\rho_1 + \rho_2 + n)p_n = 0, \quad 0 \leqslant n \leqslant c, \\ c p_{c+1} + (\rho_1 + \rho_2)p_{c-1} - (\rho_1 + c)p_c = 0, \\ c p_{n+1} + \rho_1 p_{n-1} - (\rho_1 + c)p_n = 0, \quad c < n. \end{gathered} \tag{14.5}$$

Кроме того,

$$\sum_{n=0}^{\infty} p_n = 1.$$

Решение этих уравнений имеет вид

$$p_n = \frac{c-\rho_1}{c-\rho_1+\rho_1 E_c(\rho_1+\rho_2)} \frac{\frac{(\rho_1+\rho_2)^n}{n!}}{N_c(\rho_1+\rho_2)}, \quad 0 \leqslant n \leqslant c,$$

$$p_n = \frac{(c-\rho_1)\, E_c(\rho_1+\rho_2)}{c-\rho_1+\rho_1 E_c(\rho_1+\rho_2)} \left(\frac{\rho_1}{c}\right)^{n-c}, \quad c \leqslant n, \tag{14.6}$$

где

$$N_c(\rho) = \sum_{i=0}^{c} \frac{\rho^i}{i!}$$

и

$$E_c(\rho) = \frac{\frac{\rho^c}{c!}}{N_c(\rho)}.$$

Вероятность того, что вызов теряется, найдем, просуммировав p_n по всем $n \geqslant c$:

$$P(>0) = \frac{cE_c(\rho_1+\rho_2)}{c-\rho_1+\rho_1 E_c(\rho_1+\rho_2)}. \tag{14.7}$$

Число потерянных вызовов из второго источника равно $\rho_2 P(>0)$. Вероятность того, что время ожидания вызова из первого источника больше t, равна

$$P(>t) = \int_t^\infty \sum_{n=c}^{\infty} p_n e^{-\mu c t} \frac{(\mu c t)^{n-c}}{(n-c)!} c\mu\, dt = P(>0)\, e^{-(c-\rho_1)\mu t}. \tag{14.8}$$

Среднее время ожидания всех вызовов из первого источника равно

$$\frac{P(>0)}{\mu(c-\rho_1)}, \tag{14.9}$$

а среднее время ожидания для задержанных вызовов —

$$\frac{1}{\mu(c-\rho_1)} \tag{14.10}$$

и не зависит от ρ_2.

Коэн рассмотрел также систему, в которой вызовы из первого источника имеют преимущество над вызовами из второго источника, и вызовы из обоих источников могут ожидать. Средние длительности ожидания для всех вызовов из первого и второго источников составляют соответственно

$$\frac{P(>0)}{\mu(c-\rho_1)} \tag{14.11}$$

и

$$\frac{cP(>0)}{\mu(c-\rho_1)(c-\rho_1-\rho_2)}. \tag{14.12}$$

Среднее время ожидания для задержанных вызовов с первым приоритетом равно

$$\frac{1}{\mu(c-\rho_1)}, \tag{14.13}$$

а для задержанных вызовов со вторым приоритетом —

$$\frac{c}{\mu(c-\rho_1)(c-\rho_1-\rho_2)}. \tag{14.14}$$

Исследовалась также одноканальная система с экспоненциальным временем обслуживания, различным для каждого из двух источников, с параметрами μ_1 и μ_2, когда вызовы из первого источника могут ожидать, а вызовы из второго источника теряются. Здесь $p(n, 1)$ — вероятность состояния $(n, 1)$, когда канал занят требованием из второго источника, а n требований из другого источника ожидают обслуживания, и $p(n, 0)$ — вероятность состояния $(n, 0)$, когда канал занят требованием из первого источника, а $n-1$ требований из этого же источника ожидают.

Заметим, что

$$\sum_{n=0}^{\infty}[p(n, 1)+p(n, 0)]=1.$$

Если

$$\alpha=\frac{\mu_2}{\mu_1}, \quad \rho_1=\frac{\lambda_1}{\mu_1}<1 \text{ и } \rho_2=\frac{\lambda_2}{\mu_2},$$

то уравнения для стационарного состояния имеют вид:

$$p(1, 0)+\alpha p(0, 1)-(\rho_1+\alpha\rho_2)p(0, 0)=0, \tag{14.15а}$$

$$\rho_2 p(0, 0)-\left(1-\frac{\rho_1}{\alpha}\right)p(0, 1)=0, \tag{14.15б}$$

$$p(n+1, 0)+\rho_1 p(n-1, 0)+\alpha p(n, 1)-(\rho_1+1)p(n, 0)=0, \quad n>0, \tag{14.15в}$$

$$\frac{\rho_1}{\alpha}p(n-1, 1)-\left(\frac{\rho_1}{\alpha}+1\right)p(n, 1)=0, \quad n>0. \tag{14.15г}$$

Кроме того, очевидно,

$$\lim_{n\to\infty}p(n, 1)=0, \quad \lim_{n\to\infty}p(n, 0)=0.$$

Последнее уравнение этой системы используется для упрощения предыдущего уравнения, которое суммируется по n с учетом первых двух уравнений. Получаем

$$p(n, 0)=\rho_1^n\left\{1+\frac{\alpha\rho_2}{\rho_1+\alpha-1}\left[1-\left(\frac{1}{\rho_1+\alpha}\right)^n\right]\right\}p(0, 0). \tag{14.16}$$

Так как сумма вероятностей должна равняться единице, то при $\rho_1<1$ находим

$$p(0, 0)=\frac{1-\rho_1}{1+\rho_2} \tag{14.17}$$

и

$$p(n, 1)=\frac{\alpha\rho_2}{(\rho_1+\alpha)}\frac{(1-\rho_1)}{(1+\rho_2)}\left(\frac{\rho_1}{\rho_1+\alpha}\right)^n. \tag{14.18}$$

Далее

$$P(>0)=1-p(0, 0). \tag{14.19}$$

Это вероятность того, что требование, поступающее из первого источника, должно ожидать, или вероятность того, что требование, поступающее из второго источника, теряется.

Среднее число ожидающих требований равно

$$L=\sum_{n=1}^{\infty}(n-1)p(n, 0)+\sum_{n=0}^{\infty}np(n, 1)=\frac{\rho_1\rho_2}{\alpha(1+\rho_2)}+\frac{\rho_1^2}{1-\rho_1}. \tag{14.20}$$

Среднее время ожидания требований из первого источника равно $\frac{L}{\lambda_1}$.

Рассмотрим случай, когда могут ожидать вызовы, поступающие из обоих источников, а вызовы из первого источника имеют преимущество над вызовами из второго источника. Обозначим через $p_1(n_1, n_2)$ $(n_1>0, n_2\geqslant 0)$ вероятность того, что канал занят требованием из первого источника, а n_1-1 требований из первого источника и n_2 требований из второго источника ожидают. Аналогично, через $p_2(n_1, n_2)$ $(n_1\geqslant 0, n_2>0)$ обозначим вероятность того, что в канале находится требование из второго источника, а n_1 требований из первого источника и n_2-1 требований из второго источника ожидают, и через $p(0, 0)$ — вероятность того, что канал свободен. Можно написать вероятности стационарного состояния.

Среднее число ожидающих требований из первого источника равно

$$\sum_{n_2=0}^{\infty}\sum_{n_1=1}^{\infty}(n_1-1)p_1(n_1, n_2)+\sum_{n_2=1}^{\infty}\sum_{n_1=0}^{\infty}n_2p_2(n_1, n_2)=$$
$$=\frac{\rho_1\left(\rho_1+\frac{\rho_2}{\alpha}\right)}{1-\rho_1}. \tag{14.21}$$

Для второго источника аналогично имеем

$$\sum_{n_2=0}^{\infty}\sum_{n_1=1}^{\infty}n_1p_1(n_1, n_2)+\sum_{n_2=1}^{\infty}\sum_{n_1=0}^{\infty}(n_2-1)p_2(n_1, n_2)=$$
$$=\frac{\rho_2(\alpha\rho_1+\rho_2)}{(1-\rho_1)(1-\rho_1-\rho_2)}. \tag{14.22}$$

Общее среднее число ожидающих требований равно сумме двух последних выражений. Среднее время ожидания найдем, разделив первое выражение на λ_1, а второе — на λ_2. Чтобы найти среднее время ожидания для задержанных требований из каждого источника, нужно разделить среднее общее время ожидания для требований из этого источника на $\rho_1+\rho_2$.

Упражнение. Составьте уравнения и дайте вывод формул (14.21) и (14.22).

Если требования не обладают приоритетами, то для нахождения среднего числа ожидающих требований можно использовать формулу Полячека—Хинчина. Таким образом,

$$L=\frac{\rho^2}{2(1-\rho)}\int\limits_0^\infty \frac{t^2}{h^2}\,b(t)\,dt=\frac{(\rho_1+\alpha\rho_2)\left(\rho_1+\frac{\rho_2}{\alpha}\right)}{1-\rho_1-\rho_2}, \tag{14.23}$$

где

$$\rho=\rho_1+\rho_2;\quad h=\frac{\rho_1+\rho_2}{\lambda_1+\lambda_2},$$

а $b(t)$ — плотность распределения времени обслуживания, т. е.

$$b(t)=\mu_1\frac{\lambda_1}{\lambda_1+\lambda_2}\,e^{-\mu_1 t}+\mu_2\frac{\lambda_2}{\lambda_1+\lambda_2}\,e^{-\mu_2 t}.$$

В этом случае среднее число ожидающих требований из первого источника равно

$$\frac{\rho_1\left(\rho_1+\frac{\rho_2}{\alpha}\right)}{1-\rho_1-\rho_2} \tag{14.24}$$

и из второго источника —

$$\frac{\rho_2(\rho_2+\alpha\rho_1)}{1-\rho_1-\rho_2}. \tag{14.25}$$

Среднее время ожидания найдём, разделив первую величину на λ_1, а вторую — на λ_2.

Сравнение L (общего числа ожидающих требований) для этого случая с L для предыдущего случая показывает, что если дается преимущество вызовам из источника с более коротким временем обслуживания, то при $\alpha<1$ в такой системе ожидает меньшее число требований, чем в системе с обслуживанием в порядке поступления.

2. Системы связи

Комбинаторный метод. Как указывалось в гл. 1, в системе связи множество источников может иметь ограниченный доступ к множеству абонентов. Системы связи исследовались с помощью уравнений состояний (весьма сложных). Якобеус исследовал их с помощью комбинаторных методов, дающих хорошие приближения. Допустим, например, что занято p линий, а $m-p$ линий свободно (общее число линий равно m). Обозначим через $G(p)$ вероятность того, что занято p линий, а через $H(m-p)$ — вероятность того, что заняты остальные $m-p$ линий. Произведение $H(m-p)G(p)$ есть вероятность того, что телефонный вызов теряется при условии, что занято p линий. Общую вероятность того, что вызов теряется, найдем, просуммировав по всем p. Получим

$$\sum_{p=0}^{m} H(m-p)G(p) \equiv E. \tag{14.26}$$

Чтобы найти $H(m-p)$, допустим, что j вызовов поступают случайным образом и распределяются среди m линий также случайным образом. Предположим, что i вызовов направляется в $m-p$ свободных линий, а $j-i$ вызовов — в p занятых линий. Вероятность этого события описывается гипергеометрическим распределением

$$\frac{\binom{p}{j-i}\binom{m-p}{i}}{\binom{m}{j}}. \tag{14.27}$$

Если число $i=m-p$ является единственно возможным, то получим

$$H(m-p)=\sum_{j=m-p}^{m} \frac{\binom{p}{j-m+p}}{\binom{m}{j}} F(j), \tag{14.28}$$

где $F(j)$ — функция распределения входящего потока.

Если, например,

$$F(j)=\binom{m}{j} b^{j}(1-b)^{m-j}, \tag{14.29}$$

то

$$H(m-p)=b^{m-p}.$$

Если

$$G(p)=\binom{m}{p} c^{p}(1-c)^{m-p},$$

то

$$E=(b+c-bc)^{m}. \tag{14.30}$$

Справедливость этой формулы можно проверить интуитивно. Рассмотрим m пар линий. Если b — вероятность того, что занята первая линия пары, а c — вероятность того, что занята вторая линия пары, то $(b+c-bc)^{m}$ — вероятность того, что занято m линий.

Управление входящим потоком в системе с потерями. В системе связи этого типа имеется прибор, управляющий входящим потоком. Поступающий вызов теряется, если заняты все c каналов или занят управляющий прибор. Вначале каждый вызов поступает к оператору, управляющему входящим потоком (как бы обращаясь за разрешением на телефонный разговор), а затем поступает в канал. Пусть P_j — вероятность того, что занято j линий и управляющий прибор свободен, а Q_j — вероятность того, что занято j линий и занят управляющий прибор. Тогда, если требования образуют пуассоновский поток с параметром λ, а время обслуживания в каналах и управляющем приборе является экспоненциальным с параметрами μ и η соответственно, то

$$\begin{aligned}(\lambda+j\mu)P_j&=\eta Q_{j-1}+(j+1)\mu P_{j+1},\\(\eta+j\mu)Q_j&=\lambda P_j+(j+1)\mu Q_{j+1}.\end{aligned}\tag{14.31}$$

Заметим, что, к примеру, переходы из состояния $(j, 1)$ и в состояние $(j, 1)$, когда линии заняты, а управляющий прибор свободен, происходят: 1) когда система находится в состоянии $(j, 1)$ и остается в нем; 2) когда система переходит из состояния $(j+1, 1)$ и состояния $(j-1, 0)$ (если управляющий прибор занят и занято $j-1$ каналов); 3) когда система переходит из состояния $(j, 1)$ в состояние $(j, 0)$, если поступает одно требование. В приведенных уравнениях можно вычислить P_j или Q_j и получить уравнение четвертого порядка. Вероятность того, что система занята, равна сумме вероятностей того, что заняты все обслуживающие устройства и занят управляющий прибор:

$$E=\frac{\left(\frac{\lambda}{\mu}\right)(r+K)}{\left(\frac{\lambda}{\mu}\right)r+\left(\frac{\lambda}{\mu}+r\right)K},\quad r=\frac{\eta}{\mu}.\tag{14.32}$$

Если число каналов равно c, то

$$K=\sum_{i=0}^{c-1}\frac{c!}{(c-i-1)!}\,\frac{\Gamma\left(\frac{\lambda}{\mu}+r+1+i\right)}{\Gamma(\mu+r+1)}\,\frac{1}{\left[\left(\frac{\lambda}{\mu}\right)r\right]^i}.\tag{14.33}$$

Эта формула была получена Воло и Леруа и уже рассматривалась в гл. 13 для случая ориентирования поступающих вызовов.

Управление входящим потоком в системе с ожиданием. Рассмотрим описанную выше схему, но для системы с ожиданием. Пусть $P(i, 1)$ — вероятность того, что в очереди находится i ожидающих требований и управляющий прибор свободен, а $Q(i, 0)$ — вероятность того, что имеется i ожидающих требований и управляющий прибор занят. Пусть обслуживание производится одним каналом. $P(0, 0)$ — вероятность отсутствия требований в системе. Имеем [795]

$$\begin{aligned}(\lambda+\mu)P(i, 1)&=\lambda P(i-1, 1)+\eta Q(i, 0), & i&\geqslant 1,\\ \lambda P(0, 0)&=\mu P(0, 1), & i&=0,\\ (\lambda+\eta)Q(i, 0)&=\lambda Q(i-1, 0)+\mu P(i+1, 1), & i&\geqslant 1,\\ (\lambda+\eta)Q(0, 0)&=\lambda P(0, 0)+\mu P(1, 1), & i&=0,\\ \sum P(i, 1)&+Q(i, 0)=1.\end{aligned}\tag{14.34}$$

Эта система уравнений имеет решение

$$P(i,1)=\frac{\lambda(1-\rho)}{\mu(\alpha_2-\alpha_1)}\left(\alpha_2^{i+1}-\alpha_1^{i+1}\right),\quad i\geqslant 0,$$

$$P(0,0)=1-\rho,$$

$$Q(i,0)=\frac{\lambda}{\mu\eta}\frac{1-\rho}{\alpha_2-\alpha_1}\left[(\lambda+\mu)\left(\alpha_1^{i+1}-\alpha_2^{i+1}\right)-\lambda\left(\alpha_2^{i}-\alpha_1^{i}\right)\right],\quad i\geqslant 0. \quad (14.35)$$

Здесь α_1 и α_2 — корни уравнений

$$\mu\eta\alpha^2-\lambda(\lambda+\mu+\eta)\alpha+\lambda^2=0,$$

$$\rho=\lambda\left(\frac{1}{\eta}+\frac{1}{\mu}\right).$$

При $\eta\to\infty$ получим геометрическое распределение, которое является решением уравнений для одноканальной системы. (Проверьте это.)

Для бесконечного числа каналов имеем

$$P(j,0)=\left(1-\frac{\lambda}{\eta}\right)e^{-\frac{\lambda}{\mu}}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j,\quad j=0,1,\ldots,$$

$$Q(j,i)=\left(1-\frac{\lambda}{\eta}\right)\left(\frac{\lambda}{\eta}\right)^{i+1}e^{-\frac{\lambda}{\mu}}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j,\quad \begin{matrix} i=0,1,\ldots, \\ j=0,1,\ldots, \end{matrix} \quad (14.36)$$

где $P(0,j)$ — вероятность того, что занято j линий, а управляющий прибор свободен, и, следовательно, нет ожидающих требований, а $Q(i,j)$ — вероятность того, что j линий занято и i требований ожидает, так как занят управляющий прибор. Для бесконечного числа каналов имеем три уравнения, для c каналов — девять уравнений, а для рассмотренной выше одноканальной системы — четыре уравнения. Рассмотрение этого случая будет продолжено в задачах.

3. Неполнодоступные системы

Беркли рассмотрел системы со ступенчатым включением. Он сформулировал так называемый принцип переполнения. На основании этого принципа принимается, что поток вызовов, не принятых частью системы со ступенчатым включением, имеет пуассоновское распределение с соответствующим средним значением, а для вычисления вероятности потерь применяется формула Эрланга. Этот способ является приближенным, но на практике дает хорошие результаты.

Недавно Вилкинсон в своей теории эквивалентных случайных величин распространил этот способ на случай, когда рассматриваются два параметра (математическое ожидание и дисперсия). Эллдин и в последнее время Экберг для решения такой задачи использовали уравнения состояний, но этот метод является очень сложным и требует применения вычислительных машин.

14.3. Приложения теории к движению автомобильного транспорта

Рассмотрим вначале задачу, которую решил Борель [79]. Придерживаясь формального решения этой задачи, получим приложение ее к движению автомобильного транспорта.

Бесконечная последовательность точек $\{A_i\}$ $(i=1, 2, \ldots)$, расположенных на полупрямой OX, имеет пуассоновское распределение с математическим ожиданием $\lambda=1$, т. е. единица масштаба равна интенсивности потока. Таким образом, вероятность того, что на отрезке длиной L находится n точек, равна $L^n \frac{e^{-L}}{n!}$.

Пусть ρ — фиксированная длина. Распределение точек A_i можно выразить через ρ следующим образом. Пусть A_1 — первая точка, расположенная справа от начала отсчета. Выберем точку B_1 справа от точки A_1 таким образом, чтобы длина интервала A_1B_1 равнялась ρ. Если в этом интервале нет новых точек A_i, то говорят, что этот интервал образует последовательность, состоящую из одного интервала. С другой стороны, если этот интервал содержит α_1 точек A_i, то вправо от точки B_1 откладываем новый отрезок длиной $\alpha_1\rho$, заканчивающийся в точке B_2. Если в интервале B_1B_2 не содержится новых точек A_i, то говорят, что этот интервал образует последовательность, состоящую из $1+\alpha_1$ интервалов. Если же в нем содержится α_2 точек A_i, то отмеряем вправо от B_2 интервал длиной $\alpha_2\rho$ и т. д. Этот процесс продолжается до бесконечности.

Последовательность может окончиться, если будет достигнут последний интервал $B_{n-1}B_n$, который не содержит новых точек A_i. За начало новой последовательности примем первую точку последовательности A_i, расположенную справа от B_n. Теперь вычислим p_n (вероятность того, что последовательность состоит ровно из n интервалов длиной ρ) и найдем p_∞ (вероятность того, что число таких интервалов ρ бесконечно).

Вероятность того, что последовательность содержит данное число интервалов

$$n=1+\alpha_1+\ldots+\alpha_k,$$

равна вероятности того, что в первом интервале A_1B_1 содержится α_1 точек, равной

$$\frac{\rho^{\alpha_1}e^{-\rho}}{\alpha_1!},$$

умноженной на вероятность того, что во втором интервале B_1B_2 (длиной $e^{-\alpha_1\rho}$) содержится α_2 точек, т. е. на $\frac{(\alpha_1\rho)^{\alpha_1}e^{-\alpha_1\rho}}{\alpha_2!}$, и т. д. до последнего интервала, в котором нет точек $\left[\text{вероятность этого события равна } \frac{(\alpha_k\rho)^0 e^{-\alpha_k\rho}}{0!}\right]$. Это произведение равно

$$\frac{\alpha_1^{\alpha_2}\alpha_2^{\alpha_3}\ldots\alpha_{k-1}^{\alpha_k}\rho^{n-1}e^{-n\rho}}{\alpha_1!\,\alpha_2!\ldots\alpha_k!}. \tag{14.37}$$

Чтобы найти p_n, необходимо рассмотреть все возможные разложения $n=1+\sum_{i=1}^{k}\alpha_i$. Для каждого такого разложения получим выражение, аналогичное (14.37), которое в общем виде можно записать как

$$p_n=a_n\rho^{n-1}e^{-n}. \tag{14.38}$$

Если $\rho<1$, то $p_\infty=0$; из сходимости ряда, составленного из p_n, вытекает соотношение $\sum_{n=1}^{\infty}p_n=1$ и мы имеем

$$\rho=\sum_{n=1}^{\infty}a_nz^n, \tag{14.39}$$

где

$$z=\rho e^{-\rho}. \tag{14.40}$$

Это уравнение имеет два нуля: $\rho_1<1$ и $\rho_2>1$ для всех z, лежащих между нулем и e^{-1}. Заметим, что при возрастании ρ от нуля до бесконечности z возрастает до e^{-1} при $\rho=1$, а затем убывает до нуля.

Упражнение. Проверьте эти положения.

Имеем

$$a_n=\frac{1}{2\pi i}\int\limits_C\frac{\rho dz}{z^{n+1}}=\frac{1}{2\pi i}\int\limits_C\frac{(1-\rho)e^{n\rho}}{\rho^n}d\rho. \tag{14.41}$$

Разложив подынтегральное выражение в ряд и выбрав коэффициенты при ρ^{-1}, получим

$$a_n=\frac{n^{n-2}}{(n-1)!}.$$

Поэтому

$$p_n=\frac{n^{n-2}}{(n-1)!}\rho^{n-1}e^{-n\rho}. \tag{14.42}$$

Этот результат справедлив также и при $\rho\gg1$. В соотношении (14.39) ρ_1 выражается как функция от z. Однако (14.39) справедливо и при $\rho=1$, т. е. при $z=e^{-1}$; следовательно, $p_\infty=0$. При $\rho>1$ положим $\rho=\rho_2$. В этом случае сумма вероятностей равна $e^{\rho_1-\rho_2}$ и $p_\infty=1-e^{\rho_1-\rho_2}$. При $\rho<1$ среднее число интервалов равно

$$\sum np_n=\frac{1}{1-\rho}, \tag{14.43}$$

а вероятная длина последовательности, к которой принадлежит точка A_i, составляет

$$\frac{\sum n^2 p_n}{\sum n p_n} = \frac{1}{(1-\rho)^2}. \qquad (14.44)$$

Среднее число последовательностей интервалов, содержащихся в отрезке длиной A, равно $A(1-\rho)$.

Формула (14.42) дает вероятность того, что в последовательности содержится n точек, включая и первую точку этой последовательности, если интервал A_1B_1 равен ρ. Если же интервал A_1B_1 равен $r\rho$, то $n=r+\alpha_1+\ldots+\alpha_k$, и, как и ранее, получим

$$p_n = r\frac{n^{n-r-1}}{(n-r)!}\rho^{n-r}e^{-n\rho}, \quad n=r, \ r+1, \ldots. \qquad (14.45)$$

Читатель может проверить это выражение, вычислив выражение, соответствующее (14.41).

Как показал Хейт [321], эта формула находит важное применение. Пусть проезд автомобилей мимо светофора разрешается через равные промежутки времени T. Допустим, что поток транспорта является пуассоновским с параметром λ, равным одной временной единице. Пусть $\rho=\lambda T$. Предположим, что в момент появления красного сигнала в очереди находится Z автомобилей, а в момент предыдущего появления зеленого сигнала число ожидавших автомобилей равнялось X. Чему равна вероятность того, что в очереди находится Z автомобилей при заданном X и при условии, что число автомашин, пропущенных при зеленом свете за промежуток времени α есть постоянная величина N? Значение ρ приближенно равно $\frac{\alpha\lambda}{N}$.

Пусть $f(z;\, x) \equiv P(Z=z \mid X=x)$, тогда

$$f(z;\, x) = \frac{(\alpha\lambda)^{z-x+N}e^{-\alpha\lambda}}{(z-x+N)!}, \quad x>N,$$

$$f(0;\, x) = \sum_{j=x}^{N} R(j;\, x), \quad z=0, \quad x \leqslant N,$$

где

$$R(n;\, r) = r\left[\frac{n^{n-r-1}}{(n-r)!}\right]e^{-n\rho}\rho^{n-r} \qquad (n=r, \ r+1, \ \ldots);$$

$$f(z;\, x) = e^{\rho z}\left[R(N+z;\, x) - \sum_{j=1}^{z-1} R(z;\, j)\, f(j;\, x)\right],$$

$$z>0, \quad x \leqslant N.$$

Первое уравнение показывает, что при $x>N$ число ожидающих автомобилей будет зависеть только от числа автомобилей, прибывших при зеленом свете. Второе уравнение элементарно. Третье уравнение находим следующим образом. При фиксированных значениях $x \leqslant N$ и $z>0$ число ожидающих автомобилей будет равно z, и после того как зажигается красный свет, в течение времени, необходимого для проезда ожидающих автомобилей, новые автомобили не прибывают. Вероятность этого события равна $f(z;\ x)\,e^{-\rho z}$. Это вероятность того, что впервые очередь исчезнет после проезда $N+z$ автомобилей. Вероятность последнего события равна $R(N+z;\ x)$, и, таким образом,

$$[R(N+z;\ x)-f(z;\ x)\,e^{-\rho z}]=\sum_{j=1}^{z=1} f(j;\ x)\,R(z;\ j)$$

есть вероятность того, что очередь, в которой находится $N+z$ автомобилей, исчезнет, а число ожидающих автомобилей будет равно j. Получим третье уравнение. Хейт приводит таблицы, в которых вычислены эти выражения. Он показывает, что при фиксированном x выполняется соотношение $\sum_{z} f(z;\ x)=1$. Кроме того, он рассматривает транспортный поток из боковых улиц и его влияние на изменение продолжительности светового сигнала.

Таннер [829] исследовал задачу взаимного влияния двух транспортных потоков, направляющихся к однорядному проезду с противоположных сторон. Он воспользовался изложенными выше идеями Бореля, но его результаты и анализ более детальны. Из-за недостатка места мы не можем останавливаться на них подробно. Но вследствие важности взаимного влияния очередей мы кратко рассмотрим это явление, хотя полученные результаты довольно ограничены.

Рассмотрим дорогу, определенный участок которой длиной AB имеет ширину, достаточную для проезда только в одном направлении, а за пределами этого участка существует свободный проезд в обоих направлениях. Автомобили V_1 случайным образом (по закону Пуассона) с интенсивностью q_1 машин в единицу времени прибывают в пункт A, намереваясь проехать из пункта A в пункт B. Аналогично, автомобили V_2 случайным образом с интенсивностью q_2 машин в единицу времени прибывают в пункт B, намереваясь проехать из пункта B в пункт A. Прибыв в пункт A, автомобиль V_1 сразу же направится на участок AB, если на этом участке нет встречных автомобилей и если в предыдущем промежутке времени β_1 ни один автомобиль V_1 не направился на участок AB. (Конкретный смысл величины β_1 выяснится позже.) В противном случае автомобиль V_1 будет ожидать до тех пор, пока не будут выполнены оба эти условия. При аналогичных условиях в предыдущем промежутке времени β_2 автомобиль V_2 направился на участок AB. Предположим далее, что каждому автомобилю V_1 для проезда через участок AB потребуется ровно α_1, а автомобилю V_2 — α_2 временны́х единиц. Допустим, что 1) все автомобили передвигаются с постоянной скоростью и 2) время, на протяжении которого автомобиль набирает скорость или останавливается, пренебрежимо мало, хотя можно допустить, что оно имеет некоторую продолжительность β_1 и β_2. Эти величины можно включить в α_1 и α_2 ($\alpha_1 > \beta_1$, $\alpha_2 > \beta_2$).

Обозначим через VB_1 промежуток времени, когда участок AB занят автомобилями V_1, ему предшествует и за ним следует промежуток времени, когда это условие не выполняется. Аналогично определяется VB_2. VB_0 — промежуток времени, когда участок AB свободен. Длительность периодов VB_0, VB_1 и VB_2 равна соответственно t_0, t_1 и t_2. Длительность каждого периода t_i ($i=0$, 1, 2) имеет распределение, моментная производящая функция которого $E(e^{t_i T})$ по аргументу T равна $M_i(T)$. Число автомобилей V_1 и V_2, ожидающих в начале периодов VB_1 и VB_2, равно соответственно r_1 и r_2. Запишем $M_1(-q_1)=M_1$ и $M_2(-q_2)=M_2$. Пусть n — число требований, которые будут обслужены (т. е. число автомобилей, которые пройдут через этот участок) до того момента, когда больше не будет обслуживаемых требований. Будем рассматривать $n\beta_1$ как время, необходимое для проезда автомобилей в одном направлении, если в момент, когда начинается проезд через участок, в очереди ожидает r автомобилей. Это время на величину β_1 больше промежутка времени, который пройдет до того момента, когда в очереди больше не будет ожидающих автомобилей V_1.

Пусть $\psi(r)$ — средняя общая длительность ожидания в очереди для автомобилей V_1 за время проезда через участок всех автомобилей, ожидающих в очереди, в которой первоначально было r автомобилей. Средняя общая длительность ожидания в очереди за время проезда через участок автомобилей, ожидающих в очереди, в которой первоначально был один автомобиль, равна $\left[e^{(\alpha_1-\beta_1)q_1} - 1\right]\psi(1)$, где α_1 — время, необходимое для проезда через участок автомобиля, ожидающего в этой очереди. Для автомобилей, прибывающих в тот момент, когда последний автомобиль из ожидавших в очереди уже вступил на участок, и для автомобилей, прибывающих в последнем промежутке $\alpha_1 - \beta_1$, время ожидания равно нулю. Таким образом, среднее время ожидания равно сумме этих величин:

$$\psi(r) + \left[e^{(\alpha_1-\beta_1)q_1} - 1\right]\psi(1). \tag{14.46}$$

После усреднения по r получим среднее общее время ожидания.

Для нахождения $\psi(r)$ используем метод Бореля. Рассмотрим группы автомобилей, проезжающих через участок. Пусть за то время, когда проезжает начальное число r автомобилей, прибывает a машин, а за время их проезда — еще b автомобилей и т. д. Вероятность того, что число автомобилей в каждой группе равно r, a, b, $c, \ldots$, составляет

$$\frac{e^{-\rho_1 r}(\rho_1 r)^a}{a!}\,\frac{e^{-\rho_1 a}(\rho_1 a)^b}{b!}\,\frac{e^{-\rho_1 b}(\rho_1 b)^c}{c!}\ldots, \tag{14.47}$$

где $\rho_1 = \beta_1 q_1$.

Заметим, что если ρ_2 — соответствующая величина для другой очереди (определяемая аналогично), то для того чтобы существовало решение для стационарного состояния, должно выполняться соотношение $\rho_1+\rho_2<1$.

Средняя общая длительность ожидания при таком законе распределения равна

$$\frac{1}{2}\beta_1\,[r(r-1)+a(a-1)+b(b-1)+\ldots]+$$
$$+\frac{1}{2}\beta_1\,[ra+ab+bc+\ldots]. \qquad (14.48)$$

Выражение в первых квадратных скобках есть время задержки автомобилей вследствие проезда через участок автомобилей своей группы, а выражение во вторых квадратных скобках — время задержки автомобилей вследствие проезда через участок автомобилей предыдущих групп.

После некоторых преобразований получаем выражение

$$\psi(r)=\frac{1}{2}\beta_1\left[\frac{r(2\rho_1-1)}{(1-\rho_1)^2}+\frac{r^2}{1-\rho_1}\right].$$

После усреднения по r находим

$$\overline{\psi(r)}=\frac{1}{2}\beta_1\left[\frac{2\rho_1-1}{(1-\rho_1)^2}E(r)+\frac{E(r^2)}{1-\rho_1}\right].$$

Выражение для среднего времени ожидания автомобилей V_1 имеет вид

$$W_1=\frac{1}{2(1-\rho_1)}\left[\beta_1\rho_1+\frac{\rho_1E\left(t_2^2\right)(1-\rho_1-\rho_2)}{\gamma+\varepsilon_1\delta_2+\varepsilon_2\delta_1}\right], \qquad (14.49)$$

где

$$\delta_1=q_1+q_2-q_1M_1;\quad \delta_2=q_1+q_2-q_2M_2;$$
$$\gamma=M_1+M_2-M_1M_2;$$
$$\varepsilon_1=\frac{e^{(\alpha_1-\beta_1)q_1}-1}{q_1};$$
$$\varepsilon_2=\frac{e^{(\alpha_2-\beta_2)q_1}-1}{q_2},$$

$M(\theta)$ — моментная производящая функция случайной величины t_1 (времени, в течение которого автомобили первой очереди занимают проезд) по аргументу θ.

То же самое относится к M_2 и t_2.

В свою очередь, среднее время ожидания для автомобилей V_2 равно

$$W_2 = \frac{1}{2(1-\rho_2)}\left[\beta_2\rho_2 + \frac{\rho_2 E\left(t_1^2\right)(1-\rho_1-\rho_2)}{\gamma+\varepsilon_1\delta_2+\varepsilon_2\delta_1}\right]. \qquad (14.50)$$

При $\beta_1 = \beta_2 = 0$ имеем

$$W_1 = \frac{e^{\alpha_1 q_2}\left(q_2 + q_1 e^{\alpha_2(q_1+q_2)}\right)}{q_2 e^{\alpha_1 q_2}\left(e^{(\alpha_1+\alpha_2)q_1} - e^{\alpha_2 q_1} + 1\right) + q_1 e^{\alpha_2 q_1}\left(e^{(\alpha_1+\alpha_2)q_2} - e^{\alpha_1 q_2} + 1\right)} \times$$

$$\times \frac{e^{\alpha_2 q_2} - \alpha_2 q_2 - 1}{q_2}. \qquad (14.51)$$

Чтобы найти W_2, заменим α_1 на α_2, q_1 на q_2 и q_2 на q_1.

При $\alpha_2 = \beta_2 = 0$ имеем

$$M_2 = 1,\ \delta_2 = q_1,\ \varepsilon_2 = 0,\ \gamma = 1,$$

$$W_1 = \frac{\rho_1^2}{2q_1(1-\rho_1)}, \qquad (14.52)$$

$$W_2 = \frac{1}{2(1-\rho_1)^2 q_1 q_2} q_2 \left[2\rho_1 - 3\rho_1^2 + 2\rho_1^3 + \right.$$

$$\left. + 2q_1\varepsilon_1(1-\rho_1) - 2q_1\alpha_1(1-\rho_1^2)\right] = \frac{e^{(\alpha_1-\beta_1)q_1}}{q_1(1-\rho_1)} - \alpha_1 -$$

$$-\frac{1}{q_1} + \frac{2\rho_1^3 - \rho_1^2}{2q_1(1-\rho_1)^2}. \qquad (14.53)$$

При $\alpha_2 = \beta_1 = \beta_2 = 0$ имеем

$$W_1 = 0,\quad W_2 = \frac{e^{\alpha_1 q_1} - \alpha_1 q_1 - 1}{q_1}.$$

Упражнение. Проверьте формулы, полученные для частных случаев.

Эти положения применимы и к пересекающимся потокам пешеходов при одностороннем движении, например, в узких проходах, при входе в помещения, в проходах между рядами, на пешеходных мостах и т. д. В этом случае более применимо упрощение $\beta_1 = \beta_2 = 0$.

Если два транспортных потока пересекаются под прямыми углами, то среднее время ожидания при $\beta_1 = 0$ или $\beta_2 = 0$ или при $\beta_1 = \beta_2 = 0$ то же, что и приведенное ранее.

14.4. Остановки транспорта в туннеле

Для оказания помощи автомобилям, потерявшим управление, и вывода их из туннеля в нью-йоркских туннелях Порт-Оторити были оборудованы аварийные гаражи, снабженные специальными тягачами. Когда туннель Линкольна имел один проезд, то он был оборудован одним гаражом, современные туннели с двумя проездами (Голландский и Линкольна) оборудованы двумя гаражами. При проектировании третьего проезда для туннеля Линкольна предусматривалось строительство третьего гаража. Используя теорию вероятностей для определения числа автомобилей, останавливающихся в туннелях, Эдай [195] показал, что оборудование третьего гаража будет использоваться слишком редко, и большие вложения капитала в строительство, а также затраты на оборудование и обслуживание дополнительного гаража не будут оправданы.

Было определено, что остановка одного автомобиля может задержать 200 машин. В 1951 г. в туннеле Линкольна с двумя проездами произошло 2714 таких остановок средней продолжительностью 5 мин. В семи случаях произошла остановка двух автомобилей одновременно. Допуская, что остановки автомобилей имеют пуассоновское распределение, можно найти вероятность того, что произойдет остановка сразу двух автомобилей в течение любого пятиминутного промежутка времени:

$$P(2)=\left(\frac{1}{2!}\right)e^{-m}m^2=\frac{1}{2}\left(e^{-0{,}0129}\right)(0{,}0129)^2=82{,}5\cdot10^{-6},$$

где m — среднее число остановок продолжительностью 5 мин., приходящееся на один проезд.

В году содержится 105 120 пятиминутных промежутков. Поэтому среднее число одновременных остановок двух автомобилей, приходящееся на один проезд, равно

$$nP(2)=(105\,120)(82{,}5\cdot10^{-6})=8{,}7,$$

а для двух проездов —

$$2nP(2)=17{,}4.$$

Полученный результат не согласовался с опытными данными. Он значительно превышает имевшее место в действительности число остановок, равное 7. Однако после тщательной проверки оказалось, что в действительности была 31 одновременная остановка двух машин. Но и это число отличается от вычисленного значения. Причина ошибки вычисления обнаружилась, когда было рассмотрено почасовое распределение остановок. Примерно три четверти остановок происходило в течение одной трети суток. Пусть N — число остановок в течение наиболее оживленного восьмичасового периода, n — число остановок в течение остальных 16 часов, а M — среднее число остановок, приходящееся на один промежуток длительностью 5 мин для наиболее оживленного восьмичасового периода. В году содержится около 35 000 таких промежутков. Вероятность того, что в любом из этих промежутков остановится один автомобиль, равна

$$P=\frac{1}{35\,000}=28{,}6\cdot10^{-6}.$$

Среднее число остановок одного автомобиля, приходящееся на один проезд, равно 1357. Таким образом,

$$N = \frac{3}{4} 1357 = 1020,$$

и

$$M = NP = 1020 (28,6 \cdot 10^{-6}) = 29,2 \cdot 10^{-3},$$

$$P(2) = \frac{1}{2!} e^{-0,0292} (0,0292)^2 = 405 \cdot 10^{-6}.$$

Среднее число остановок двух автомобилей одновременно, происходящих в течение наиболее оживленного восьмичасового периода, равно

$$35\,000 (405 \cdot 10^{-6}) = 14,2.$$

Аналогично найдем, что среднее число случаев одновременных остановок двух автомобилей, происходящих в течение остальных 16 часов, равно 0,8. Таким образом, общее число остановок двух автомобилей одновременно равно $2(14,2+0,8)=30$. Этот результат хорошо согласуется с числом таких остановок, имевшим место в действительности, равным 31. Выполнив аналогичные выкладки, получим, что для Голландского туннеля среднее число таких остановок равно 140, в действительности было 149 остановок.

Для вычисления числа остановок трех автомобилей одновременно используется тот же метод.

Таким же способом была произведена оценка работы туннеля с тремя проездами. Было получено, что в этом случае за год произойдет 2533 остановки одного автомобиля, 57 остановок двух автомобилей одновременно и одна остановка трех автомобилей одновременно. Таким образом, 114 автомобилей должны обслуживаться при одновременной остановке двух машин. Опыт показал, что в Голландском туннеле для обслуживания 149 таких автомобилей было достаточно двух гаражей.

Была дана рекомендация не производить постройку третьего гаража для обслуживания третьего проезда. В результате вычислений была обеспечена экономия примерно 130 000 долларов. Затрачено же было около пяти человеко-дней.

14.5. Приложения теории к движению воздушного транспорта

С некоторыми понятиями, связанными с управлением движением воздушного транспорта, мы познакомились в иллюстративном примере первой главы. Пирси [641] рассмотрел приложения некоторых идей теории массового обслуживания к организации посадки самолетов. В данном случае обычно представляет интерес сокращение времени посадки. Вычислим вначале вероятность того, что один за другим $n-1$ самолетов ожидают приземления.

Допустим, что самолеты приближаются к зоне управления со случайных направлений через случайные промежутки времени, распределенные по экспоненциальному закону, с постоянной интенсивностью прибытия, которая принимается равной одной единице. Следовательно, e^{-t} — распределение промежутков времени между моментами прибытия. Самолет, который прибывает через промежуток времени, меньший, чем минимальное время, необходимое для безопасного приземления предыдущего самолета, задерживается на минимальное время. Отношение минимального времени, необходимого для безопасной посадки, к среднему значению промежутка времени между прибывающими самолетами обозначается через T (для простоты будем считать, что для данного аэропорта эта величина является постоянной). Обычно представляет интерес случай, когда $T<1$. Вероятность того, что прибывший самолет не задерживается, равна

$$\int_{T}^{\infty} e^{-t} dt = e^{-T}. \tag{14.54}$$

Вероятность того, что будет задержан один самолет, найдем, рассмотрев все задержки одиночных самолетов между двумя незадерживаемыми самолетами. Самолет, который будет задержан, должен прибыть через промежуток времени $t_1<T$ после прибытия незадерживаемого самолета, непосредственно предшествующего ему, а незадерживаемый самолет, непосредственно следующий за ним, должен прибыть через промежуток времени $t>2T-t_1$. Таким образом, искомая вероятность совместного появления этих двух событий равна

$$\int_{0}^{T} e^{-t_1} dt_1 \int_{2T-t_1}^{\infty} e^{-t} dt = Te^{-2T}.$$

Вероятность того, что будет задержано два самолета, находится аналогично (рассматриваются два задерживаемых самолета между двумя незадерживаемыми) путем вычисления вероятности совместного появления событий:

$t_1 < T$ — для первого задерживаемого самолета, следующего за незадерживаемым;

$t_2 < 2T - t_1$ — для второго задерживаемого самолета, следующего за первым задерживаемым;

$t > 3T - t_1 - t_2$ — для незадерживаемого самолета, следующего непосредственно за двумя задерживаемыми.

В результате для двух задерживаемых самолетов получим

$$\int_{0}^{T} e^{-t_1} dt_1 \int_{0}^{2T-t_1} e^{-t_2} dt_2 \int_{3T-t_1-t_2}^{\infty} e^{-t} dt = \frac{3T^2}{2} e^{-3T}. \tag{14.55}$$

Общее выражение для вероятности того, что задерживается $n-1$ самолетов, имеет вид $\alpha_n T^{n-1} e^{-nT}$, где α_n — коэффициент, зависящий только от n. Очевидно, что должно выполняться соотношение

$$\sum_{n=1}^{\infty} \alpha_n T^{n-1} e^{-nT} = 1 \tag{14.56}$$

или

$$\sum_{n=1}^{\infty} \alpha_n u^n = T, \tag{14.57}$$

где величина $u \equiv Te^{-T}$ для малых T определяется однозначно, следовательно, T можно выразить как функцию от u:

$$T(u) = \sum_{n=1}^{\infty} \alpha_n u^n. \tag{14.58}$$

Используя то обстоятельство, что начало координат — кратный полюс, имеем

$$\alpha_n = \frac{1}{2\pi i} \oint \frac{T(u)}{u^{n+1}}\, du = \frac{1}{2\pi i} \oint \frac{e^{nT}}{T^n} (1-T)\, dT = \\ = \frac{n^{n-1}}{(n-1)!} - \frac{n^{n-2}}{(n-2)!} = \frac{n^{n-1}}{n!}. \tag{14.59}$$

Следовательно, разложив подынтегральное выражение в ряд и собирая коэффициенты при T^{-1}, можно найти вычет.

Вероятность того, что один за другим задерживаются $n-1$ самолетов, равна

$$\frac{n^{n-1}}{n!} T^{n-1} e^{-nT} \approx \frac{n^{n-1}}{n^n e^{-n} \sqrt{2\pi n}} T^{n-1} e^{nT} = \frac{T^{n-1} e^{n(1-T)}}{n\sqrt{2\pi n}}. \tag{14.60}$$

Используя формулу Стирлинга для $n!$, Пирси приводит ряд кривых для этого распределения.

Среднее число самолетов, находящихся в системе (с учетом первого самолета, совершающего посадку без ожидания), равно

$$L = \sum_{n=1}^{\infty} n\alpha_n T^{n-1} e^{-nT} = \frac{1}{1-T}. \tag{14.61}$$

Это выражение можно легко найти, дифференцируя (14.56) по T и производя упрощения. (Заметим, что при $T=1$ задерживаются все самолеты.) Аналогично находим второй начальный момент, он равен $\frac{1}{(1-T)^3}$.

Доля задерживаемых самолетов определяется как отношение среднего числа самолетов, находящихся в системе, без учета самолета, совершающего посадку, к среднему числу самолетов:

$$\frac{L-1}{L}=T.$$

Распределение длительности посадки найдем путем следующих рассуждений. Все промежутки времени длительностью $t<T$ имеют нулевую частоту; промежутки времени длительностью $t=T$ появляются с частотой T; доля задерживаемых самолетов, т. е. доля промежутков времени длительностью $t>T$, появляется с частотой $1-T$ появления незадерживаемых самолетов, умноженной на вероятность их прибытия, т. е. на $e^{-(t+T)}$. Используем единичную функцию $H(T-t)$ (которая равна единице для положительных значений аргумента и равна нулю для отрицательных; ее производная есть дельта-функция) и дельта-функцию $\delta(T-t)$ для того, чтобы представить это распределение в виде

$$(1-T)\,e^{-(t-T)}H(T-t)+T\delta(T-t).$$

Теперь, используя интегральное уравнение Линдли, можно получить распределение времени ожидания. Посредством детального анализа Пирси находит выражение для распределения в промежутке времени t, $mT<t<(m+1)T$;

$$Q_T(t)\,dt=(1-T)\,e^{t-T}dt\sum_{r=0}^{\infty}\frac{[(m+r+1)\,T-t]^{m+r}}{(m+r+1)!}\times[t+$$

$$+(m+r+1)(1-T)]\,e^{-(m+r)T},$$

откуда после интегрирования по t $(0\leqslant t\leqslant\infty)$ он определяет T как долю задерживаемых самолетов. Заметим, что при суммировании по m необходимо рассматривать интервалы $[mT,\ (m+1)T]$. Отсюда найдем также среднее время ожидания

$$\frac{T^2}{2(1-T)}.$$

Заметим, что время ожидания увеличивается с ростом T. Приведенное выше распределение дает критерии для определения необходимой пропускной способности аэропорта. Более важным критерием является такой, который устанавливает, что только небольшое число самолетов должно ожидать больше определенного промежутка времени.

Белл [39] проанализировал различные наблюдения и измерения, производимые для изучения управления воздушным транспортом, вывел формулы для стационарных состояний одноканальной системы с пуассоновским входящим потоком и постоянным временем обслуживания.

В статье Геллихера и Уилера [273] основное внимание уделяется вычислительной стороне задачи, связанной с посадкой самолетов.

Интенсивность прибытия самолетов в район Нью-Йорка рассматривается как периодически изменяющаяся с периодом, равным одним суткам. Предполагается, что в течение данных суток время приземления постоянно, но принимает одно из двух различных значений (одна и две минуты). Сутки были разделены на промежутки, равные постоянному времени приземления (которое медленно изменяется с течением времени). В i-м промежутке средняя интенсивность прибытия λ_i постоянна, и распределение числа самолетов, прибывающих в этом промежутке времени, является пуассоновским:

$$\frac{\lambda_i^n e^{-\lambda_i}}{n!} \equiv a_n^i, \quad n=0,\ 1,\ \ldots .$$

Самолеты обслуживаются в порядке прибытия; число каналов, имеющихся в распоряжении до $(i-1)$-го промежутка времени для использования в i-м промежутке, равно c_i. Для вычисления стационарных вероятностей p_n^i того, что в конце i-го промежутка времени в системе находится n самолетов, выраженных через эти же вероятности для $(i-1)$-го промежутка, используется формула Кроммелина.

Когда вероятности p_n^i известны, то остальные вероятности для всех значений i, следующих за этим промежутком, можно вычислить рекуррентным способом. Для обозначения номера промежутка используется верхний индекс. Уравнения имеют вид

$$\begin{aligned} p_0^i &= \alpha_0^i P_{c_i}^{i-1}, \\ p_1^i &= \alpha_1^i P_{c_i}^{i-1} + \alpha_0^i p_{c_i+1}^{i-1}, \\ p_n^i &= \alpha_n^i P_{c_i}^{i-1} + \sum_{j=1}^{n} \alpha_{n-j}^i p_{c_i+j}^{i-1}, \end{aligned} \tag{14.62}$$

где

$$P_n^i = \sum_{j=0}^{n} p_j^i.$$

Так как вероятности p_n^i вычисляются в конце промежутков времени (т. е. для вложенной цепи Маркова), то промежуточные значения вероятностей находим с помощью интерполяции. Обычным способом можно вывести также различные показатели эффективности. Чтобы произвести обработку данных, полученных в течение суток, включая определение среднего числа ожидающих самолетов и среднего числа самолетов, находящихся в системе, вычислительному устройству потребуется немного времени (всего семь минут). Эти данные используются для вычисления числа ожидающих самолетов и определения других показателей.

Кратко остановимся на выдающейся работе Ли и его сотрудников [494, 495], посвященной организации процесса прибытия самолетов в Лондонский аэропорт, погрузки, принятия пассажиров на борт и вылета. Пассажир, прибывающий не позже заданного момента времени накануне вылета самолета, проходит обычную систему регистрации. Если он прибывает в последнюю минуту, то становится к окну с более короткой очередью, чтобы ускорить оформление документов.

Работа аэропорта в большой степени зависит от того, насколько быстро и экономично производится прием и передача информации. Контроль за передачей информации осуществляет центральный пост, который помещается в той части аэропорта, где происходит погрузка и обслуживание самолетов. Сообщения о состоянии готовности самолетов к вылету передаются непрерывно с помощью подвижного приемопередатчика, размещенного в машине. Один оператор одновременно следит за подготовкой двух-трех самолетов. Данные о готовности к полету накапливаются на центральном посту и высвечиваются на табло. Если поступают сообщения о каких-либо задержках, то из центрального поста по телефону наводится справка у ответственного за определенные операции (например, за погрузку багажа или доставку продуктов питания для самолетного буфета) о причинах помех и выясняется время задержки. После того как станет известно время задержки, центральный пост доводит до всеобщего сведения, в каком из проходов задерживается посадка пассажиров.

Одной из причин, вызывавших задержку отправления самолетов, являлась перегрузка каналов связи. Раньше из подвижных передатчиков сообщали о завершении каждой операции в процессе подготовки самолета к вылету. Это приводило к перегрузкам в работе центрального поста и, следовательно, к задержкам отправления самолетов. После того как эта проблема была изучена, было дано указание передавать сообщения с помощью подвижных радиопередатчиков только в случае наличия каких-либо задержек. С помощью теории массового обслуживания были выработаны сообщения, которые поступают на центральный пост и принимаются на посту в зависимости от наличия приоритетов. Это помогло значительно упростить работу центрального поста. Затруднения, которые возникали вследствие перегруженности линий связи, сократились до минимума.

Впоследствии, также с помощью методов теории массового обслуживания, было проведено исследование порядка регистрации пассажиров. Цель исследования — добиться относительной экономии при выполнении этого вида обслуживания. Нужно было определить численность штата работников для успешной работы различных вариантов системы регистрации. Было предложено несколько вариантов, каждый из которых был проверен на практике. Была принята такая система регистрации, которая обеспечивала наименьшее число обслуживающего персонала и не приводила к неоправданному ожиданию в очереди. Для Лондонского аэропорта была принята смешанная система регистрации, при которой опаздывающие пассажиры проходят ускоренное обслуживание у последнего окошка. Однако для других аэропортов полученная схема может не подойти. Как ни странно, оказалось, что для некоторых аэропортов общая система регистрации является непрактичной.

14.6. Приложение понятия о конечном источнике входящего потока к обслуживанию станков

В этом параграфе рассматривается поток станков, поступающих на обслуживание. Общее число станков m конечно. Станки выходят из строя, т. е. требуют обслуживания, по пуассоновскому закону с параметром λ. Таким образом, станки являются источниками входящего потока. (Заметим, что λ является характеристикой одного станка.) Среднее время работы станка равно $\frac{1}{\lambda}$. Обслуживание станков производят c мастеров. Здесь рассматривается время обслуживания, имеющее экспоненциальное распределение с параметром μ, и время обслуживания, постоянное для всех мастеров и всех станков. (Такач рассматривает произвольное время обслуживания.) Допустим, что все станки одинаковы и в каждом из них в единицу рабочего времени появляется одно и то же среднее число неисправностей. Временем простоя станка является время, в течение которого станок остановлен и ожидает обслуживания.

Важной задачей, которую нужно рассмотреть, является определение оптимального числа станков (основанное на анализе стоимости рабочего времени мастера и стоимости простоя станка), которое должен обслуживать один мастер. Нередко на практике, когда станок останавливается, оказывается, что мастер занят, так как в это время он либо обслуживает другие станки, либо выполняет другие работы, скажем, заготавливает материалы, необходимые для работы станков, как, например, на текстильной фабрике. Наша задача состоит не в том, чтобы дать исчерпывающее изложение вопросов, связанных с обслуживанием станков, а в том, чтобы показать применение теории массового обслуживания к этому вопросу.

1. Пуассоновский входящий поток, произвольное распределение времени обслуживания

А. Я. Хинчин, Крониг и Мондриа (1943), Пальм (1947), Ашкрофт (1950) и Такач (1957) исследовали систему с пуассоновским входящим потоком, поступающим из конечной совокупности, и произвольным распределением времени обслуживания применительно к обслуживанию станков одним мастером.

Аналогичные идеи встречаются также при рассмотрении телефонной системы, в которой вызов теряется, если все линии заняты. Эту задачу, о которой уже упоминалось в первом параграфе настоящей главы, исследовали Эрланг (1918), Фрай (1928), Пальм (1943), Полячек (1953), Такач (1956) и Коэн (1957).

Рассматривая обслуживание m станков одним мастером, Такач определил p_n $(0 < n < m)$ — вероятность того, что в любой момент времени t имеется n станков, нуждающихся в обслуживании.

Вероятность того, что за время $(t, t+\Delta t)$, поступит на обслуживание один станок, равна $\lambda t+o(\Delta t)$; функция распределения времени обслуживания любого станка равна $B(t)$; обслуживание станков производится в порядке поступления. Вначале Такач нашел π_n $(0 \leqslant n \leqslant m)$ — вероятность того, что непосредственно перед окончанием обслуживания данного станка работает n станков, а затем определил p_n. Такой путь доказательства приемлем для стационарного состояния, которое существует, так как этот процесс определяет апериодическую цепь Маркова с конечным числом состояний.

Вероятность r_{ij} того, что непосредственно перед окончанием обслуживания имеется j работающих станков при условии, что непосредственно перед окончанием обслуживания предыдущего станка было i работающих станков, равна

$$r_{ij}=\int\limits_0^\infty \binom{i+1}{j} e^{-j\lambda t}(1-e^{-\lambda t})^{i+1-j}dB(t), \quad i=1, \ \ldots, \ m-2,$$

$$r_{m-1, j}=r_{m-2, j}, \quad i=m-1.$$

Заметим, что не имеет смысла рассматривать случай, когда непосредственно перед окончанием обслуживания действует m станков, так как имеется всего m станков. Заметим также, что $j \leqslant i+1$, коль скоро предполагается, что предыдущий станок был обслужен и новых заявок на обслуживание не поступило.

Для нахождения этого интеграла рассматривается вероятность того, что станок требует обслуживания, т. е. $p \equiv 1-e^{-\lambda t}$ (во время обслуживания этого станка ожидают j других станков), и вероятность того, что станок не нуждается в обслуживании, т. е. $q \equiv e^{-\lambda t}$. С помощью этих двух формул получим выражение

$$\binom{i+1}{j} e^{-j\lambda t}(1-e^{-\lambda t})^{i+1-j}.$$

После того как закончится обслуживание предыдущего станка, число действующих станков станет равным $i+1$, а вероятность того, что j из них работают в течение времени t, и есть приведенное выражение. Так как t задается распределением вероятностей, то, умножая полученное выражение на функцию распределения для t и интегрируя по всем значениям t, получаем приведенное ранее выражение для r_{ij}.

Имеем

$$\sum_{j=0}^{m-1} \pi_j = 1$$

и

$$\pi_j = \sum_{i=j-1}^{m-1} r_{ij}\pi_i, \ j = 1, \ldots, m-1.$$

Второе выражение есть 1) вероятность того, что непосредственно перед окончанием обслуживания предыдущего станка имеется $j-1$ действующих станков, обслуженный станок присоединяется к действующим и до конца обслуживания предыдущего станка не поступает новых заявок на обслуживание; или 2) вероятность того, что непосредственно перед окончанием обслуживания предыдущего станка имеется j действующих станков и этот станок присоединяется к ним, но непосредственно перед окончанием обслуживания данного станка один станок потребует обслуживания.

Чтобы найти π_j (из второго выражения), умножим второе выражение на z^j, подставим значение r_{ij} и просуммируем по j, $(0 \leqslant j \leqslant m-1)$. Пусть

$$P(z) = \sum_{j=0}^{m-1} \pi_j z^j.$$

Тогда

$$P(z) = \int_0^\infty (1 - e^{-\lambda t} + z e^{-\lambda t}) P(1 - e^{-\lambda t} + z e^{-\lambda t})\, dB(t) +$$

$$+ (1-z)\pi_{m-1} \int_0^\infty e^{-\lambda t} (1 - e^{-\lambda t} + z e^{-\lambda t})^{m-1}\, dB(t).$$

Введем обозначение

$$P(z) = \sum_{j=0}^{m-1} B_j (z-1)^j;$$

тогда

$$B_j = \frac{1}{j!} \frac{d^j P(z)}{dz^j}\bigg|_{z=1}, \ B_0 = 1.$$

Биномиальные моменты вероятностей π_i имеют вид

$$B_j = \sum_{i=j}^{m-1} \binom{i}{j} \pi_i.$$

Следовательно,

$$\pi_j = \sum_{n=j}^{m-1} (-1)^{n-j} \binom{n}{j} B_j.$$

По определению,

$$B_j = C_j \frac{\sum\limits_{i=j}^{m-1} \binom{m-1}{i} \left(\frac{1}{C_i}\right)}{\sum\limits_{i=0}^{m-1} \binom{m-1}{i} \left(\frac{1}{C_i}\right)},$$

где

$$C_0 = 1; \ C_i = \frac{\beta(\lambda)\beta(2\lambda)\ldots\beta(i\lambda)}{[1-\beta(\lambda)][1-\beta(2\lambda)]\ldots[1-\beta(i\lambda)]}.$$

Здесь β — преобразование Лапласа—Стилтьеса функции распределения, времени обслуживания.

Вероятности p_n находятся аналогично: вместо r_{ij} используются вероятности перехода p_{ij}, определяемые как и r_{ij}, с той лишь особенностью, что в подынтегральном выражении вместо $dB(t)$ будет $\mu[1-B(t)]\,dt$ — оставшееся время обслуживания. Вводятся также биномиальные моменты для p_n, которые можно связать с биномиальными моментами для π_n. В результате получим следующее выражение:

$$p_n = \frac{m\pi_{n-1}}{n\left(\frac{m\lambda}{\mu} + \pi_{n-1}\right)}, \ n = 1, \ldots, m,$$

где μ — интенсивность обслуживания.

Значение вероятности p_0 находим по формуле

$$p_0 = 1 - \sum_{n=1}^{m} p_n.$$

Среднее число станков, находящихся в системе, равно

$$\frac{1-p_m}{\frac{\lambda}{\mu}} = \frac{m}{\frac{m\lambda}{\mu} + \pi_{m-1}}.$$

Среднее время ожидания равно

$$\frac{m-1}{\mu} + (1-\pi_{m-1})\left[\frac{\sigma^2 + \left(\frac{1}{\mu}\right)^2}{\frac{2}{\mu}} - \frac{\frac{1}{\mu}}{1-\beta(\lambda)}\right],$$

где σ^2 — дисперсия времени обслуживания.

2. Экспоненциальное время обслуживания, число мастеров равно c

Если p_n — вероятность того, что в стационарном состоянии в системе находится n станков, как обслуживаемых, так и ожидающих обслуживания, то легко найдем уравнения для стационарного состояния:

$$m\lambda p_0 = \mu p_1,$$

$$[(m-n)\lambda + n\mu]\, p_n = (m-n+1)\,\lambda p_{n-1} + \\ + (n+1)\,\mu p_{n+1},\quad 1 \leqslant n < c, \tag{14.63}$$

$$[(m-n)\lambda + c\mu]\, p_n = (m-n+1)\,\lambda p_{n-1} + c\mu p_{n+1},\quad c \leqslant n \leqslant m.$$

Подставив первое соотношение во второе и т. д., найдем

$$(n+1)\,\mu p_{n+1} = (m-n)\,\lambda p_n,\quad n < c,$$

$$c\mu p_{n+1} = (m-n)\,\lambda p_n,\quad c \leqslant n \leqslant m. \tag{14.64}$$

Пальм [236] получил следующие выражения:

$$p_n = p_0 (c\rho)^n \binom{m}{n},\quad 0 \leqslant n < c,$$

$$p_n = p_0 (c\rho)^n \binom{m}{n} \frac{n!}{c!\, c^{n-c}},\quad c \leqslant n \leqslant m,$$

$$p_0 = 1 - \sum_{n=1}^{m} p_n = \left(\sum_{n=0}^{m} \frac{p_n}{p_0} \right)^{-1}, \tag{14.65}$$

где $\rho = \dfrac{\lambda}{c\mu}$.

Упражнение. Решите приведенные уравнения.

Если принять обозначения, которые ввел Наор [601]:

$$p(k,\ \xi) = \frac{\xi^k e^{-\xi}}{k!},$$

$$P(k,\ \xi) = \sum_{i=k}^{\infty} p(i,\ \xi),$$

$$S\left(m,\ c,\ \frac{\mu}{\lambda}\right) = \sum_{i=0}^{c-1} \left[\frac{c^i}{i!} - \frac{c^{i-1}}{(i-1)!} \right] [1 - P(m-i+1,\ \rho^{-1})] = \\ = \sum_{i=0}^{c-1} \frac{c^i}{i!} p(m-i,\ \rho^{-1}) + \frac{c^{c-1}}{(c-1)!} [1 - P(m-c+1,\ \rho^{-1})],$$

то

$$p_n = \frac{c^n}{n!} \frac{p(m-n,\ \rho^{-1})}{S\left(m,\ c, \frac{\mu}{\lambda}\right)},\quad n < c, \tag{14.66}$$

$$p_n = \frac{c^{c-1}}{(c-1)!} \frac{p(m-n,\ \rho^{-1})}{S\left(m,\ c, \frac{\mu}{\lambda}\right)},\quad c \leqslant n \leqslant m.$$

Эти выражения можно проверить, произведя подстановку.

Среднее число обслуживаемых станков равно

$$c\frac{S\left(m-1,\ c,\frac{\mu}{\lambda}\right)}{S\left(m,\ c,\frac{\mu}{\lambda}\right)}. \tag{14.67}$$

Таким образом, если a — среднее число действующих станков, b — число обслуживаемых станков, а L_q — число станков, ожидающих в очереди, то

$$a+b+L_q=m,\quad \frac{a}{b}=\frac{\mu}{\lambda},$$

$$b=\sum_{n=0}^{c-1}np_n+c\sum_{n=c}^{m}p_n=c-\sum_{n=0}^{c-1}(c-n)p_n=c\frac{S\left(m-1,\ c,\frac{\mu}{\lambda}\right)}{S\left(m,\ c,\frac{\mu}{\lambda}\right)}.$$

Среднее число незанятых мастеров равно $c-b$, коэффициент занятости мастеров равен $\frac{b}{c}$, коэффициент потерь вследствие обслуживания равен $\frac{b}{m}$, а коэффициент потерь вследствие простоя станков (так как на обслуживании находятся другие станки) равен $\frac{L_q}{m}$. Общий коэффициент потерь равен $\frac{(b+L_q)}{m}$. Коэффициент полезного действия станков равен $\frac{a}{m}$. Наор приводит также иллюстративные примеры. Вероятность ожидания равна

$$P(>0)=\sum_{n=c}^{m}p_n. \tag{14.68}$$

Среднее время ожидания ремонта равно

$$W_q=\frac{1}{\lambda}\sum_{n=c+1}^{m}(n-c)=\frac{L_q}{\lambda}, \tag{14.69}$$

где L_q — среднее число станков, находящихся в очереди.

Среднее число станков, ожидающих обслуживания, равно

$$\frac{W_q}{P(>0)}. \tag{14.70}$$

Результат, который получил Ашкрофт [15] для случая, когда обслуживание станков производит один мастер и время обслуживания имеет экспоненциальное распределение, найдем, положив $c=1$.

Наор [602] получил нормальные аппроксимации, облегчающие вычисление p_n. Рассмотрим распределение

$$p_c(j)=\frac{(c-1)!}{j!\,c^{c-j}}(c-j),\quad j=0,\ 1,\ \ldots,\ c-1,$$

первый и второй моменты которого соответственно имеют вид

$$E_c\{j\}=\frac{1}{p(c,\ c)}\{cp(c,\ c)-[1-P(c,\ c)]\},\tag{14.71}$$

$$E_c\{j^2\}=\frac{1}{p(c,\ c)}\{c^2p(c,\ c)-$$
$$-[1-P(c,\ c)]-2c[1-P(c-1,\ c)]\}.\tag{14.72}$$

Образуем свертку

$$\{p(k,\ \xi)\}*\{p_c(j)\}=v(i,\ c,\ \xi)=$$
$$=\sum_{n=0}^{i}\frac{(c-1)!}{n!\,c^{c-n}}(c-n)p(i-n,\ \xi),\ i\leqslant c-1.$$

Здесь суммирование производится до $c-1$ при $i\geqslant c-1$, где $i=j+k$. Если ввести обозначения $u(i,\ c,\ \zeta)=v(i,\ c,\ \xi)$ и $\zeta=\frac{\xi}{c}$, то

$$U(i,\ c,\ \zeta)=\sum_{s=0}^{i}u(s,\ c,\ \zeta)=\sum_{n=0}^{c-1}\frac{(c-1)!}{n!\,c^{c-n}}(c-n)[1-P(i-n+1,\ c\zeta)].$$

Это выражение справедливо при $i\geqslant c-1$, что нас и интересует.

Легко показать, что

$$S(i,\ c,\ \zeta)=\frac{c^{c-1}}{(c-1)!}U(i,\ c,\ \zeta).$$

Так как u является сверткой, то при больших c математическое ожидание и дисперсия имеют вид

$$E_u\{i\}=c\zeta+E_c\{j\}\approx c\zeta+c-\frac{1}{2}\sqrt{2\pi c}+\frac{1}{3},\tag{14.73}$$

$$V_u\{i\}=c\zeta+E_c\{j^2\}-E_c^2\{j^2\}\approx c\zeta+c\left(2-\frac{\pi}{2}\right)-\frac{\sqrt{2\pi c}}{6}+\frac{2}{9}.\tag{14.74}$$

При достаточно больших значениях $c\zeta$ — математического ожидания распределения $p_c(j)$ — с помощью центральной предельной теоремы находим следующее приближенное выражение:

$$U(i,\ c,\ \zeta)\approx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{y^2}{2}}dy,\tag{14.75}$$

где

$$z=\frac{i+\frac{1}{2}-c\zeta-E_c\{j\}}{\left(c\zeta+E_c\{j^2\}-E_c^2\{j\}\right)^{\frac{1}{2}}}.$$

Здесь используется поправка на непрерывность. Заметим, что при стандартной форме записи нормального распределения $y=\frac{(x-\mu)}{\sigma}$, где μ — математическое ожидание, а σ — среднее квадратическое отклонение. Здесь x заменено на $i+\frac{1}{2}$, исправленное значение i. Это приближенное выражение упрощает вычисление p_n.

3. Обслуживание станков одним мастером при постоянном времени обслуживания

Эту задачу решали Бенсон и Кокс [59], Ашкрофт [15] и Вестгарт [874]. Чтобы получить результат Бенсона и Кокса, предположим, что обслуживание станков осуществляется в k этапов, каждый из которых имеет экспоненциальное распределение с математическим ожиданием $\frac{1}{\mu}$. Общее время обслуживания имеет распределение хи-квадрат с $2k$ степенями свободы и математическим ожиданием $\frac{k}{\mu}$. Если величины k и μ неограниченно возрастают и при этом их отношение $\frac{k}{\mu}$ остается постоянным, то получим решение для случая постоянного времени обслуживания. При распределении времени обслуживания по закону хи-квадрат вероятность того, что n станков из общего числа m требует ремонта, равна

$$p_n=\sum_{i=1}^{n} i\binom{m-n+i}{i}\nabla^{i-1}\left\{\frac{\left[1+\left(\frac{\lambda}{\mu}\right)(m-n)\right]^k-1}{m-n}\right\}p_{n-i}, \qquad (14.76)$$

где ∇ — обратный разностный оператор ($\nabla f_i=f_i-f_{i-1}$).

В случае постоянного времени обслуживания эта вероятность равна

$$p_n=\sum_{i=1}^{n} i\binom{m-n+i}{i}\nabla^{i-1}\left\{\frac{\exp\left[(m-n)\left(\frac{k\lambda}{\mu}\right)\right]-1}{m-n}\right\}p_{n-i}, \qquad (14.77)$$

при этом должно выполняться условие

$$\sum_{n=0}^{m} p_n=1.$$

Ашкрофт приводит таблицы для определения среднего числа действующих станков.

4. Обслуживание станков одним мастером при появлении двух различных типов остановок

Предположим, что происходят два различных типа остановок станков, каждый из которых появляется случайным образом со средней интенсивностью λ_1 и λ_2 соответственно; остановки первого типа обладают приоритетом. Общее число станков равно m. Допустим, что время обслуживания станков одним мастером имеет экспоненциальное распределение с параметром μ_1 при остановке первого типа и с параметром μ_2 при остановке второго типа. Обозначив через $p_r(n_1, n_2)$ вероятность того, что в стационарном состоянии произошло n_1 остановок первого типа и n_2 остановок второго типа и что происшедшая остановка относится к r-му типу ($r = 1, 2$), получим следующую систему уравнений:

$$[\mu_1 + (m - n_1 - n_2)(\lambda_1 + \lambda_2)]\, p_1(n_1, n_2) =$$

$$= (m - n_1 - n_2 + 1)\lambda_1 p_1(n_1 - 1, n_2) + (m - n_1 - n_2 + 1)\lambda_2 \times$$

$$\times p_1(n_1, n_2 - 1) + \mu_1 p_1(n_1 + 1, n_2) + \mu_2 p_2(n_1, n_2 + 1), \quad n_1, n_2 \neq 0, m,$$

$$[\mu_1 + (m - n_1)(\lambda_1 + \lambda_2)]\, p_1(n_1, 0) = (m - n_1 + 1)\lambda_1 p_1(n - 1, 0) +$$

$$+ \mu_1 p_1(n_1 + 1, 0) + \mu_2 p_2(n_1, 1), \quad n_1 \neq m,$$

$$\mu_1 p_1(m, 0) = \lambda_1 p_1(m - 1, 0);$$

$$[\mu_2 + (m - n_1 - n_2)(\lambda_1 + \lambda_2)]\, p_2(n_1, n_2) =$$

$$= (m - n_1 - n_2 + 1)\lambda_1 p_2(n_1 - 1, n_2) +$$

$$+ (m - n_1 - n_2 + 1)\lambda_2 p_2(n_1, n_2 - 1), \quad n_1, n_2 \neq 0, m,$$

$$[\mu_2 + (m - n_2)(\lambda_1 + \lambda_2)]\, p_2(0, n_2) = (m - n_2 + 1)\lambda_2 p_2(0, n_2 - 1) +$$

$$+ \mu_1 p_1(1, n_2) + \mu_2 p_2(0, n_2 + 1), \quad n_2 \neq m,$$

$$\mu_2 p_2(0, m) = \lambda_2 p_2(0, m - 1),$$

$$m(\lambda_1 + \lambda_2)\, p(0, 0) = \mu_1 p_1(1, 0) + \mu_2 p_2(0, 1),$$

$$p(0, 0) + \sum_{n_1=1}^{m} \sum_{n_2=0}^{m-n_1} p_1(n_1, n_2) + \sum_{n_2=1}^{m} \sum_{n_1=0}^{m-n_2} p_2(n_1, n_2) = 1.$$

Представляет интерес решение этих уравнений относительно p_n (вероятности того, что число остановившихся станков равно n). Решение задачи в общем виде не приводится.

5. Каждый из c мастеров обслуживает остановки одного типа; всего имеется c типов остановок

Общее число станков равно m. Допустим, что каждый из различных типов остановок имеет пуассоновское распределение с параметром λ_i, соответствующим остановке i-го типа. Каждый из c мастеров обслуживает станок при остановке определенного типа. Распределение времени обслуживания для остановки i-го типа является экспоненциальным с параметром μ_i. Обозначим через $p(n_1, \ldots, n_c)$ вероятность того, что в стационарном состоянии имеет место n_1 остановок первого типа, n_2 остановок второго типа и т. д. Имеем систему уравнений:

$$\left[(m-n)\sum_{i=1}^{c}\lambda_i + \sum_{i=1}^{c}\mu_i\right] p(n_1, \ldots, n_c) =$$

$$= \sum_{i=1}^{c}(m-n+1)\lambda_i p(n_1, \ldots, n_i - 1, \ldots, n_c) +$$

$$+ \sum_{i=1}^{c}\mu_i p(n_1, \ldots, n_i + 1, \ldots, n_c), \quad n_i = 1, \ldots, m-1, \quad n \leqslant m-1,$$

$$\sum_{i=1}^{c}\mu_i p(n_1, \ldots, n_c) = \sum_{i=1}^{c}\lambda_i p(n_1, \ldots, n_i - 1, \ldots, n_c),$$

$$n_i = 1, \ldots, m-1, \; n = m,$$

$$\left[(m-n)\sum_{i=1}^{c}\lambda_i + \sum_{j}\mu_i\right] p(n_1, \ldots, n_i, \ldots, 0, \ldots) =$$

$$= \sum_{j}(m-n+1)\lambda_i p(n_1, \ldots, n_i - 1, \ldots, 0, \ldots) +$$

$$+ \sum_{j}\mu_i p(n_1, \ldots, n_i + 1, \ldots, 0, \ldots) +$$

$$+ \sum_{n-j}\mu_i p(n_1, \ldots, n_i, \ldots, 1, \ldots),$$

$$\mu_i p(0, \ldots, m, \ldots, 0) = \lambda_i p(0, \ldots, m-1, \ldots, 0),$$

$$m\sum_{i=1}^{c}\mu_i p(0, \ldots, 0, \ldots, 0) = \sum_{i=1}^{c}\mu_i p(0, \ldots, 1, \ldots, 0),$$

где $n = \sum_{i=1}^{c} n_i$. Сумма $\sum_{j}$ берется по таким значениям i, для которых $n_i \neq 0$, а сумма $\sum_{n-j}$ — по i, для которых $n_i = 0$.

Решение этой системы уравнений имеет вид

$$p(n_1, \ldots, n_c) = \frac{m!}{(m-n)!}\prod_{i=1}^{c}\rho_i^{n_i} p(0, \ldots, 0),$$

где $\rho = \frac{\lambda_i}{\mu_i}$.

Сумма вероятностей, распространенная на все значения n_1, n_2 и т. д., должна равняться единице. Из этого соотношения определяется

$$p(0, \ldots, 0) = \sum_{i=1}^{c}\frac{\rho_i^{c-1}F(\rho_i, m)}{\prod\limits_{\substack{i=1 \\ j\neq i}}(\rho_i - \rho_j)},$$

где $\rho_i \neq \rho_j$ при $i \neq j$ и

$$F(\rho_i, m) = \sum_{n=0}^{m} \frac{m!}{(m-n)!} \rho_i^n = 1 + m\rho_i + \ldots + m!\rho_i^m.$$

Среднее число действующих станков равно

$$E[m-n] = m \sum_{i=1}^{c} \frac{\rho_i^{c-1} F(\rho_i, m-1)}{\prod\limits_{\substack{i=1 \\ j \neq i}}^{c} (\rho_i - \rho_j)} p(0, \ldots, 0).$$

Вероятность исправности станка составляет

$$\frac{E[m-n]}{m} [p(0, \ldots, 0)]^{-c}.$$

В качестве упражнения определите последнюю величину при $c=2$.

Если доля L рабочего времени мастера затрачивается на вспомогательные работы, а остальное время — на обслуживание станков, то можно определить такое число станков m, при котором доля этого времени равна $1-L$ и, таким образом, с помощью m можно найти действительный коэффициент полезного действия. Максимальная интенсивность обслуживания равна $\frac{1-L}{\rho}$. Минимальное число станков, необходимое для получения этого значения, можно найти указанным способом. Бенсон и Кокс приводят также таблицы для вычислений подобного рода.

Бенсон [60] рассматривает несколько измененную ситуацию, когда на мастера возложены и другие обязанности, например, он выполняет вспомогательные работы.

6. Постоянное время обслуживания, мастер переходит от станка к станку

Мэк, Мёрфи и Уэбб [543] исследовали коэффициент использования m станков, которые выходят из строя в случайные моменты времени с интенсивностью λ и обслуживаются одним мастером, который переходит в определенном порядке от одного станка к другому, а затем снова возвращается к первому станку. Время перехода t_i от $(i-1)$-го к i-му станку постоянно. Мастер обслуживает станок в течение постоянного промежутка времени R. Пусть p_n — вероятность того, что во время осмотра всех станков n из них потребуют обслуживания. Вероятность того, что станок, с которого был начат осмотр, продолжает работать и в конце осмотра, равна $e^{-\lambda(mT+nR)}$, где $mT = \sum_{i=1}^{m} t_i$. Уравнения имеют вид

$$p_n e^{-\lambda(mT+nR)} + p_{n+1} e^{-\lambda[mT+(n+1)R]} = p_n. \tag{14.78}$$

Решение их записывается как

$$p_n = (a^{n-1}b - 1)(a^{n-2}b - 1)\ldots(ab - 1)(b - 1)p_0, \qquad (14.79)$$

где $a = e^{\lambda R}$, $b = e^{\lambda m T}$, а p_0 находится из соотношения

$$\sum_{n=0}^{m} \binom{m}{n} p_n = 1. \qquad (14.80)$$

В работе показано, что это решение является единственным. Затем вычисляются различные величины, которые уже встречались нам в других примерах.

14.7. Обеспечение запасными частями и агрегатами

Тейлор и Р. Джексон [832] исследовали задачу обеспечения запасными авиационными двигателями. Предположим, что непрерывно работает m авиационных двигателей, и как только один из них выходит из строя, он немедленно заменяется запасным при условии, что такой запасной двигатель имеется, а неисправный двигатель сразу же направляется в ремонтную мастерскую. Эти m двигателей выходят из строя случайным образом с постоянной интенсивностью, равной λ двигателей в единицу времени. Время, необходимое для обслуживания одного двигателя, имеет экспоненциальное распределение, и его математическое ожидание равно $\frac{1}{\mu}$. Одновременно может обслуживаться c двигателей. Пусть общее число запасных двигателей равно n. До тех пор пока ремонтируется не более n двигателей, число действующих двигателей будет равно m. Как только число двигателей, поступивших в ремонтную мастерскую, станет равно $n+1$, число действующих двигателей станет меньше m. В этом случае возможны два варианта:

1. Работа всех двигателей прекращается до тех пор, пока из мастерской не поступит замена.

2. Продолжает работать число двигателей, меньшее m.

Рассмотрим первый случай. Вероятность того, что в любой момент времени в мастерской будет находиться i неисправных двигателей, равна $P_i(t)$. Система уравнений для этого случая имеет вид:

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t), \quad i = 0,$$

$$P_i'(t) = \lambda P_{i-1}(t) - (\lambda + i\mu) P_i(t) + (i+1)\mu P_{i+1}(t), \quad 0 < i < c, \qquad (14.81)$$

$$P_i'(t) = \lambda P_{i-1}(t) - (\lambda + c\mu) P_i(t) + c\mu P_{i+1}(t), \quad c \leqslant i < n+1,$$

$$P_{n+1}'(t) = \lambda P_n(t) - c\mu P_{n+1}(t), \quad i = n+1.$$

Представляет интерес стационарное решение, которое записывается в виде

$$p_i = \frac{(c\rho)^i}{i!} p_0, \quad i \leqslant c,$$

$$p_i = \rho^{i-c} \frac{(c\rho)^c}{c!} p_0, \ c \leqslant i \leqslant n+1, \tag{14.82}$$

где $\rho = \frac{\lambda}{c\mu}$; из соотношения $\sum_{i=0}^{n+1} p_i = 1$ находим

$$p_0 = \left[e^{c\rho} - \sum_{i=c+1}^{\infty} \left(\frac{c\rho}{i!}\right)^i + \rho \left(\frac{c\rho}{c!}\right)^c \frac{1-\rho^{n+1-c}}{1-\rho} \right]^{-1}.$$

Среднее число неисправных двигателей равно

$$U = \sum_{i=0}^{n+1} i p_i. \tag{14.83}$$

Среднее число двигателей, находящихся на обслуживании, составляет

$$S = \sum_{i=0}^{c} i p_i + c \sum_{c+1}^{n+1} p_i. \tag{14.84}$$

Среднее число двигателей, ожидающих обслуживания, равно

$$L_q = \sum_{i=c+1}^{n+1} (i-c) p_i. \tag{14.85}$$

Рассмотрим случай, когда $\rho = 1$, т. е. когда возможная интенсивность обслуживания равна интенсивности появления неисправных двигателей. Из решения, полученного для стационарного состояния, видно, что распределение вероятностей числа неисправных двигателей есть монотонно возрастающая последовательность. С помощью рассуждений, аналогичных описанным ниже для случая, когда $\rho < 1$, можно показать, что увеличение числа запасных двигателей не приводит к существенному уменьшению вероятности появления аварийной обстановки (p_{n+1}). Кроме того, увеличение числа запасных двигателей приводит к тому, что число двигателей, ожидающих обслуживания, возрастает по линейному закону.

Рассмотрим наиболее часто встречающийся на практике случай, когда $\rho < 1$.

При $c \leqslant i \leqslant n+1$ имеем следующее соотношение:

$$p_i < \frac{1-\rho}{1-\rho^{n-c+2}} \rho^{i-c}, \tag{14.86}$$

правая часть которого является убывающей функцией от n и i. Среднее число двигателей, ожидающих обслуживания, имеет следующую верхнюю границу:

$$L_q < \frac{\rho}{1-\rho}. \tag{14.87}$$

Это означает, что среднее число двигателей, ожидающих обслуживания, не зависит от n; следовательно, в этом случае увеличение числа запасных двигателей не приводит к увеличению числа двигателей, ожидающих ремонта, а увеличивает лишь эксплуатационную эффективность.

14.8. Приложения теории к управлению запасами

Приведем пример применения теории массового обслуживания к решению задачи управления запасами. Заметим, что хотя в таких задачах налицо вероятностные процессы, все же лишь немногие из этих задач решаются методами теории массового обслуживания. В теории управления запасами рассматривается поток поступающих товаров, их хранение, пополнение запаса и, наконец, сбыт товаров. Очередь может состоять из ожидающих удовлетворения заявок на товары, но чаще рассматривается очередь товаров, ожидающих использования.

Каруш [418] рассмотрел задачу, в которой дается некоторое углубление этих положений. Пусть общее число запасенных товаров равно n, из них n_0 предметов имеется в наличии, а $n_1 = n - n_0$ предметов находится в процессе пополнения запаса. Каждый из этих предметов проходит цикл пополнения запаса с данной функцией распределения (одинаковой для всех предметов). Поступающие заявки (или прибытие клиентов) образуют поток Пуассона с параметром λ. Как только один из n_0 запасенных товаров ($n_0 > 0$) оказывается проданным, посылается заявка на пополнение проданного товара. Если $n_0 = 0$, продажа прекращается, покупатель возвращается без покупки и очередь товаров, ожидающих пополнения запаса, не устанавливается. В этом случае имеем задачу массового обслуживания при отсутствии очереди. Требуется вычислить вероятность того, что в стационарном состоянии нет товаров, подготовленных к продаже (отношение числа потерянных клиентов к общему числу заявок), как функцию от n. Более общим выражением является вероятность $p(n_0 \mid n)$ того, что в наличии имеется n_0 предметов при условии, что общее число запасенных предметов равно n.

Процесс пополнения запаса может происходить r различными путями, каждый из которых имеет вероятность $\alpha_i (i = 1, \ldots, r)$. Каждый путь состоит из k_i фаз $(i = 1, \ldots, r)$, а время, затрачиваемое в каждой фазе, имеет экспоненциальное распределение. Как только заканчивается обслуживание в данной фазе, немедленно начинается обслуживание в следующей фазе.

Среднее время, необходимое для перемещения товара из j-й фазы i-го пути, равно $\frac{1}{\mu_{ij}}$. В любой момент времени n_1 предметов, находящихся в процессе пополнения запаса, распределены некоторым образом между фазами, при этом в j-й фазе i-го пути имеется n_{ij} предметов. Определим состояние системы в произвольный момент времени с помощью вектора

$$N=(n_0,\ n_{11},\ \ldots,\ n_{1k},\ n_{21},\ \ldots,\ n_{rk_r}).$$

Обозначим через $P(N)$ стационарную вероятность для состояния N и одношаговые переходы запишем в виде:

$$T_{ij}(N)\to N,$$

$$T_{ij}(N)\equiv(n_0,\ \ldots,\ n_{ij}+1,\ n_{i,\,j+1}-1,\ \ldots),$$

$$T_{ik_i}(N)\equiv(n_0-1,\ \ldots,\ n_{ik_i}+1,\ \ldots),$$

$$T_{0i}(N)\equiv(n_0+1,\ \ldots,\ n_{i1}-1,\ \ldots).$$

Уравнения для стационарного состояния имеют вид

$$\Big(\lambda+\sum_{i,j}\mu_{ij}n_{ij}\Big)p(N)=\lambda\sum_i\alpha_i p\,[T_{0i}(N)]+$$
$$+\sum_{i,j}\mu_{ij}(n_{ij}+1)p\,[T_{ij}(N)],\quad n_0>0. \qquad (14.88)$$

При $n_0=0$ в левой части λ опускается.

Теперь $p(n_0\,|\,n)=\sum p(N)$. Суммирование производится по всем векторам N, для которых

$$\sum_{i,j}n_{ij}=n-n_0.$$

Если ввести однородный полином n_1-й степени, записанный как

$$P(n_0)\equiv\sum_N p(N)\,x_{11}^{n_{11}}x_{12}^{n_{12}}\ldots x_{ij}^{n_{ij}}\ldots,\quad \sum_{i,j}n_{ij}=n_1, \qquad (14.89)$$

то приведенная выше система уравнений принимает вид

$$\lambda P(n_0)+\sum_{i,\,j}\mu_{ij}x_{ij}\frac{\partial P(n_0)}{\partial x_{ij}}=\lambda\Big(\sum_i\alpha_i x_{i1}\Big)P(n_0+1)+$$
$$+\sum_{j\neq k_i}\mu_{ij}x_{i,\,j+1}\frac{\partial P(n_0)}{\partial x_{ij}}+\sum_i\mu_i k_i\frac{\partial P(n_0-1)}{\partial x_{ik_i}},\quad n_0=1,\ldots,n. \qquad (14.90)$$

При $n_0=0$ опускаются первый член в левой части этого выражения и последний член в правой части.

Решение (единственность которого показал Каруш при помощи несложного приема) имеет вид

$$P(n_0)=\frac{\lambda^{n_1}}{n_1!}\left(\frac{\sum\limits_{i,j}a_i x_{ij}}{\mu_{ij}}\right)^{n_1},\quad n_0=0,\ 1,\ \ldots,\ n, \tag{14.91}$$

при $P(n+1)\equiv 0$. Вероятность того, что в j-й фазе i-го пути имеется n_{ij} предметов, равная

$$p(N)=\frac{\lambda^{n_1}}{C}\prod_{i,j}\frac{a_i^{n_{ij}}}{n_{ij}!\mu_{ij}^{n_{ij}}}, \tag{14.92}$$

определяется как общий коэффициент полинома $P(n_0)$. Нормирующий множитель C находится из выражения

$$C=1+\rho+\frac{\rho^2}{2!}+\ldots+\frac{\rho^n}{n!}, \tag{14.93}$$

где $\rho=\frac{\lambda}{\mu}$, а $\frac{1}{\mu}$ — среднее время пополнения запаса.

Кроме того, используя зависимость между $p(n_0\,|\,n)$ и $p(N)$, находим

$$p(n_0\,|\,n)=\frac{1}{C}\frac{\rho^{n_1}}{n_1!},\quad n=n_0+n_1, \tag{14.94}$$

что и является искомым выражением.

Вероятность того, что товаров в наличии нет, равна $p(0\,|\,n)$. Чтобы найти эту величину, в формуле (14.94), заменим n_1 на n, так как $n=n_0+n_1$, где $n_0=0$.

Обобщая эти результаты на случай, когда параметр λ является функцией от n_0, получим

$$p(n_0\,|\,n)=\frac{1}{Cn_1!}\frac{\lambda(n)}{\mu}\frac{\lambda(n-1)}{\mu}\frac{\lambda(n_0+1)}{\mu},\quad n=0,\ 1,\ldots,n-1, \tag{14.95}$$

Здесь для определения C используется то обстоятельство, что сумма вероятностей равна единице.

14.9. Плотины и накапливающие системы

В этом параграфе не приводится решение задач, здесь главным образом рассматривается построение моделей. Даунтон [190], Гани [276—279], Кендалл [436], Моран [582—585] и Прабху [685] исследовали теорию плотин и накапливающих систем. Некоторые положения этой теории связаны с теорией массового обслуживания и ее приложениями. Обычно рассматривается распределение количества воды, остающейся в водохранилище после использования, в непрерывном или дискретном времени (например, за год). В последнем случае количество воды, поступающей в водохранилище, будет изменяться из года в год в соответствии с некоторым распределением. Вода накапливается (за исключением возможных случаев переполнения, когда происходит сброс воды) для использования в засушливое время года. В течение этого времени вода расходуется с постоянной скоростью. Можно предположить, что общее количество расходуемой воды меняется каждый год. Для простоты будем считать, что это количество равно постоянной величине M.

Пусть $\{X(t)\}$ — семейство попарно независимых и одинаково распределенных случайных величин, то есть, если $t_1<t_2$, то X_{t_1} не зависит от X_{t_2}. Через X_t обозначается поток, поступающий в водохранилище в момент времени t. Если же значение t дискретно и измеряется годами, то X_t — поток в начале t-го года. Обозначим через Z_t количество воды, накопленной к моменту времени t, перед поступлением потока X_t. Пусть Y_t — количество вытекшей воды. Если через t обозначено время в годах, то Y_t — количество воды, израсходованной в засушливое время t-го года. Рассмотрим эту задачу для дискретного и непрерывного распределения вероятностей и для дискретного и непрерывного времени. Кроме того, можно принять, что водохранилище имеет конечную емкость K, или же его емкость бесконечна.

Заметим, что сброс воды через водослив будет происходить в том случае, если

$$W_t = X_t + Z_t - K$$

величина положительная. В противном случае вода будет продолжать накапливаться. Распределение величины $W(t)$ можно определить с помощью распределения величин X_t и Z_t, если последние известны.

1. Дискретное время и дискретное распределение вероятностей

Рассмотрим случай дискретного времени (измеряемого годами). Сформулируем задачу для случая дискретного распределения вероятностей, не приводя решения, которое можно получить только численными методами. Вначале допустим, что $Y_t = M$. Таким образом, количество расходуемой воды равно $\min(X_t + Z_t, M)$. Предположим, что в момент времени t величина X_t принимает значения 0, 1, ... с вероятностями $p_0, p_1, \ldots$, а величина Z_t принимает значения 0, 1, ... в момент времени t с вероятностями $P_0, P_1, \ldots, P_k$, а в момент времени $t+1$

с вероятностями $P'_0, \ldots, P'_k$; кроме того, $K \geqslant M$. Получим систему уравнений:

$$
\begin{aligned}
P'_0 &= P(p_0 + p_1 + \ldots + p_M) + P_1(p_0 + \ldots + p_{M-1}) + \\
&\quad + \ldots + P_M p_0, \\
P'_1 &= P_0 p_{M+1} + P_1 p_M + \ldots + P_{M+1} p_0, \qquad (14.97)\\
P'_{K-M} &= P_0(p_k + \ldots) + P_1(p_{k-1} + \ldots) + \\
&\quad + \ldots + P_{K-M}(p_M + \ldots), \\
P'_{K-M-1} &= \ldots = P'_K = 0.
\end{aligned}
$$

Решив их, найдем распределение величины Z_{t+1} в зависимости от распределения величины Z_t и т. д. до Z_{t+n}. Эти величины образуют цепь Маркова. При $n \to \infty$ переходим к стационарному решению, коль скоро $p_i > 0$ $(i = 0, \ldots, K)$.

Чтобы найти решение для стационарного случая, положим $P'_i = P_i$ $(i = 0, \ldots, K - M)$ и, кроме того, используем соотношение $\sum\limits_{i=1}^{K-M} P_i = 1$.

Эту задачу можно решить различными численными методами, методом Монте-Карло и матричными методами.

2. Дискретное время и непрерывное распределение вероятностей

Приступая к рассмотрению случая непрерывного распределения, предположим вначале, что при дискретном распределении вероятностей поступающий поток воды и содержимое водохранилища в сумме принимают значения $Z_t + X_t = i$ с вероятностью R_i; кроме того, допустим, что стационарный режим устанавливается. В этом случае легче получить распределение величины Z_t. Так, если известны распределения случайной величины $Z_t + X_t$ и случайной величины X_t, то можно определить распределение Z_t. В стационарном случае имеем следующую систему уравнений, описывающую распределение случайной величины R_i:

$$
\begin{aligned}
R_0 &= p_0 R_0 + p_0 R_1 + \ldots + p_0 R_M, \\
R_1 &= p_1 R_0 + p_1 R_1 + \ldots + p_1 R_M + p_0 R_{M+1}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
R_{K-M-1} &= p_{K-M-1} R_0 + \ldots + p_{K-M-1} R_M + p_{K-M-2} R_{M+1} + \\
&\quad + \ldots + p_0 R_{K-1}, \qquad (14.98)
\end{aligned}
$$

$$R_{K-M}=p_{K-M}R_0+\ldots+p_{K-M}R_M+p_{K-M-1}R_{M+1}+$$
$$+\ldots+p_1(R_K+R_{K+1}+\ldots),$$

. . . .

$$R_{K-M+s}=p_{K-M+s}R_0+\ldots+p_{K-M+s}R_M+$$
$$+p_{K-M+s-1}R_{M+1}+\ldots+p_s(R_K+\ldots),$$

. . . .

Заметим, что величина M показывает расходуемое количество воды. Читатель без труда сможет показать справедливость этих уравнений.

Имея в виду, что

$$P_0=R_0+\ldots+R_M,\quad P_{K-M}=R_K+R_{K+1}+\ldots,$$

и $P_{K-M+s}=0$ (при $s>0$), из этой системы уравнений получим уравнения для стационарного состояния.

Случай непрерывного распределения вероятностей аналогичен описанному выше. Если плотность вероятности случайной величины X_t равна $f(x)$ при $X\geqslant 0$ и нулю при $X<0$, а плотность X_t+Z_t равна $g(x)$, то получим следующее уравнение:

$$g(x)=f(x)\int_0^M g(y)\,dy+\int_M g(y)f(x+M-y)\,dy+$$
$$+f(x-K+M)\int_K^\infty g(y)\,dy. \tag{14.99}$$

Три члена правой части уравнения соответствуют условиям:

$0\leqslant X_t+Z_t\leqslant M$ —	вода находится ниже требуемого уровня;
$M<X_t+Z_t\leqslant K$ —	сброса через водослив нет, вода накапливается;
$K<X_t+Z_t$ —	сброс воды через водослив; переполнение водохранилища определяется разностью $K-(X_t+Z_t)$.

Заметим, что $p_0=\int_0^M g(y)\,dy$ и $p_1=\int_K^\infty g(y)\,dy$. С помощью этого уравнения Моран нашел распределение величины Z_t, которое равно

$$\begin{aligned} &p_0 \text{ при } Z_t=0,\\ &g(x+M) \text{ при } Z_t=x,\ 0<x<K-M,\\ &p_1 \text{ при } Z_t=K-M. \end{aligned} \tag{14.100}$$

3. Плотины с конечной и бесконечной емкостью водохранилища

Если емкость водохранилища бесконечна, т. е. $K \to \infty$, то рассматриваемая задача упрощается, и уравнения для стационарного состояния принимают вид

$$P_0 = P_0(p_0 + \ldots + p_M) + P_1(p_0 + \ldots + p_{M-1}) + \\ + \ldots + P_M p_0, \qquad (14.101)$$

$$P_i = P_0 p_{M+1} + P_1 p_{M+i-1} + \ldots + p_{M+i} p_0, \quad i \geqslant 1.$$

Распределение $\{R_i\}$ при стационарном состоянии описывается системой уравнений

$$R_0 = p_0(R_0 + \ldots + R_M),$$

$$R_i = p_i(R_0 + \ldots + R_M) + p_{i-1} R_{M+1} + \ldots + p_0 R_{M+i}, \qquad (14.102)$$

$$\sum_{i=0}^{\infty} R_i = 1. \qquad i \geqslant 1,$$

Если отождествить длину очереди с количеством воды, находящейся в водохранилище, пуассоновский поток требований — с потоком воды, поступающим в водохранилище, а обслуживание группами, содержащими M требований (или всю очередь, если последняя меньше M), — с вытекающим потоком воды, то последняя система уравнений будет описывать длину очереди в моменты времени, непосредственно предшествующие началу обслуживания. Это приводит к определенному соотношению между плотинами и системами массового обслуживания.

Имеем следующее выражение:

$$P_0 = R_0 + \ldots + R_M, \quad P_i = R_{M+i}, \quad i \geqslant 1, \qquad (14.103)$$

следовательно, приведенные выше две системы уравнений являются эквивалентными.

Замечание. Представляют интерес два случая. Один из них имеет место, когда количество накопленной воды очень велико, и, следовательно, вероятность того, что величина Z_t становится равной нулю, очень мала. В другом случае мала вероятность того, что водохранилище всегда будет заполнено. Таким образом, возможны два вида решений.

Стационарное ненулевое решение существует, когда для первого случая справедливо соотношение $\sum_{n=0}^{\infty} n p_n > M$, а для второго случая — соотношение $\sum_{n=0}^{\infty} n p_n < M$.

Кроме того, для случая непрерывного распределения вероятностей при $K \to \infty$ имеем

$$g(x) = f(x) \int_0^M g(y)\,dy + \int_M^{M+x} f(M+x-y)\,g(y)\,dy, \quad (14.104)$$

при этом выполняется соотношение

$$\sum_{i=0}^{\infty} P_i = 1.$$

При $M = 1$ имеем

$$\begin{aligned} P_0 &= P_0(p_0 + p_1) + P_1 p_0, \\ P_1 &= P_0 p_2 + P_1 p_1 + P_2 p_0, \\ &\ldots\ldots\ldots\ldots\ldots \end{aligned} \quad (14.105)$$

Эта система уравнений нам уже знакома. Введем производящую функцию

$$P(z) = \sum_{i=0}^{\infty} P_i z^i, \quad p(z) = \sum_{i=0}^{\infty} p_i z^i,$$

и обычным способом найдем

$$P(z) = p_0 P_0 \frac{z-1}{z-p(z)}. \quad (14.106)$$

Математическое ожидание, т. е. среднее количество воды, находящейся в водохранилище перед поступлением нового притока, равно

$$\sum_{i=0}^{\infty} iP_i = \frac{p''(1)}{2[1-p'(1)]}. \quad (14.107)$$

Моменты высших порядков находятся аналогичным образом.

Заметим, что если $\{P_i\}$ является решением в случае бесконечного запаса воды при дискретном распределении вероятностей, то в случае конечного запаса воды имеем

$$P'_i = \frac{P_i}{P_0 + \ldots + P_{K-1}}. \quad (14.108)$$

Можно определить $\{P_i\}$ путем разложения в ряд функции $\frac{z-1}{z-p(z)}$ для достаточно малых значений z и приравнивания коэффициентов (при условии, что $p_0 > 0$). Кроме того, можно использовать функцию распределения

$$G(x) = \int_0^x G(x+M-t)\,dF(t). \quad (14.109)$$

В качестве примера можно решить это уравнение, когда

$$F(x) = \sum_i c_i e^{-\lambda_i x}, \quad (14.110)$$

где c_i — комплексные постоянные величины и $\operatorname{Re}(\lambda_i) > 0$.

Посредством подстановки находим условия, которым должны удовлетворять c_i и λ_i, в особенности, коль скоро известно распределение (общего типа) $G(x)$. В качестве упражнения определите $F(x)$, если $G(x) = 1 - e^{-\mu x}$ $(\mu > 0)$.

Гани и Прабху [279] рассматривали также входящие потоки, приводящие к экспоненциальным распределениям для стационарного состояния водохранилища, и получили приближенные решения для случая конечного запаса воды.

4. Непрерывное время и непрерывное распределение вероятностей

Чтобы рассмотреть задачу для непрерывного времени с непрерывным распределением вероятностей и неограниченной емкостью, разделим промежуток времени $(t, t+1)$ на отрезки длительностью $\frac{1}{n}$ каждый (в конце каждого отрезка расходуется единица количества воды). Распределение входящего потока в промежутке времени $(t, t+1)$ есть n-кратная свертка распределения $\{p_i\}$, поэтому его производящая функция равна $\{p(z)\}^n$. Заметим, что мы принимаем $M=1$ и используем ранее рассмотренные уравнения. При увеличении n функция $\{p_i\}$ изменяется таким образом, что $\{p(z)\}^n$ приближается к производящей функции для непрерывного распределения.

В качестве примера рассмотрим геометрический закон распределения

$$p_i = (1-r)r^i, \quad i = 0, 1, \ldots, \quad 0 < r < 1. \qquad (14.111)$$

Тогда

$$p(z) = (1-r)(1-rz)^{-1}. \qquad (14.112)$$

Если ξ — случайная величина, распределенная по этому закону, то ее характеристическая функция имеет вид

$$(1-r)(1-re^{i\theta})^{-1}, \qquad (14.113)$$

и случайная величина $X = \frac{\xi}{n}$ имеет математическое ожидание $\frac{r}{n}(1-r)$. Пусть при $n \to \infty$ и $r \to 1$ математическое ожидание стремится к $\mu_1 = \sum_{n=0}^{\infty} np_n$. Запишем $r = \frac{\mu_1 n}{(1-\mu_1 n)}$.

Характеристическая функция случайной величины X равна

$$\lim_{n \to \infty} \frac{1}{\mu_1 n}\left(1 - \frac{\mu_1 n}{1+\mu_1 n} e^{\frac{i\theta}{n}}\right)^{-1} = (1-\mu_1 i\theta)^{-1}. \qquad (4.114)$$

Таким образом, случайная величина X имеет асимптотическое распределение $\frac{e^{-\frac{x}{\mu_1}}dx}{\mu_1}$, которое мы будем рассматривать как распределение притока, поступающего в промежутке времени $(t, t+1)$, т. е. за время единичного промежутка. Тогда распределение входящего потока за любой промежуток времени T имеет вид

$$\frac{1}{\mu_1 \Gamma(T)}\left(\frac{x}{\mu_1}\right)^{T-1} e^{-\frac{x}{\mu_1}} dx. \tag{14.115}$$

Для случая непрерывного распределения вероятностей нужно вычислить вероятность того, что не происходит запасания воды. Начнем с того, что рассмотрим приведенное выше геометрическое распределение как n-кратную свертку распределения $\{p_n\}$, следовательно, распределение самой функции $\{p_n\}$ есть n-кратная свертка. Ее производящая функция «корня» из $\{p_n\}$ имеет вид

$$p(z) = (1-r)^{\frac{1}{n}} (1-rz)^{-\frac{1}{n}}. \tag{14.116}$$

Она может использоваться для вычисления $P(z)$, а следовательно, и $\{P_i\}$.

В данном случае $p_0 = (1-r)^{\frac{1}{n}}$, $\mu_i = \frac{r}{n(1-r)}$, и вероятность того, что в водохранилище нет воды, для случая дискретного распределения равна

$$\frac{1-p'(1)}{p_0} = \frac{1-\frac{r}{n(1-r)}}{(1-r)^{\frac{1}{n}}}. \tag{14.117}$$

При $n \to \infty$ эта вероятность стремится к $1-\mu_1$ (предполагаем, что $\mu_1 < 1$), т. е. к вероятности того, что для случая непрерывного распределения вероятностей в водохранилище нет воды. Можно строго показать, что данная вероятность определяется с помощью $H(z)$ — функции распределения величины Z_t. В свою очередь, можно показать, что при $Z_t > 0$ функция $H(t)$ определяется с помощью интегрального уравнения

$$H(z) - \frac{1}{2}[H(0+0) - H(0-0)] = H(z) - \frac{1}{2}(1-\mu_1) =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{(1-\mu_1)(1-e^{-i\theta z})}{i\theta + \log(1-\mu_1 i\theta)} d\theta. \tag{14.118}$$

Метод нахождения этого распределения будет теперь развиваться в различных направлениях.

Даунтон исследовал случай непрерывного распределения вероятностей, рассматривая групповое поступление требований в случае дискретной модели. В случае непрерывного распределения объем группы и промежутки времени между группами стремятся к нулю. Поступление требований группами, содержащими n требований, происходит в соответствии с функцией распределения

$$A(t) = 1 - e^{-nt}, \quad 0 \leqslant t < \infty. \tag{14.119}$$

Пусть $B_n(t)$ — функция распределения времени обслуживания группы, содержащей n требований, а $C_n(t)$ — соответствующая функция распределения времени ожидания. Требования обслуживаются в порядке поступления. Функция распределения времени обслуживания равна $B(t)$, а функция распределения времени ожидания равна $C(t) \equiv P(w \leqslant t)$ при $0 \leqslant t < \infty$. Выражение для загрузки имеет вид

$$\rho = \frac{\int\limits_0^\infty t dB(t)}{\int\limits_0^\infty t dA(t)}. \tag{14.120}$$

Таким образом,

$$\int\limits_0^\infty t B_n(t) = \frac{\rho}{n}.$$

Если

$$\beta_n(s) \equiv \int\limits_0^\infty e^{-st} dB_n(t),$$

то, воспользовавшись уравнением (9.9), получим преобразование функции распределения времени ожидания

$$\gamma_n(s) = \int\limits_0^\infty e^{-st} dC_n(t) = (1 - \rho)\left\{1 - \frac{n[1 - \beta_n(s)]}{s}\right\}^{-1}. \tag{14.121}$$

Требуется определить

$$\psi(s) = \lim_{n \to \infty} \gamma_n(s), \tag{14.122}$$

при условии, что приток за конечный отрезок времени принимает конечное значение.

В случае дискретного распределения вероятность поступления ровно r групп за время t определяется законом Пуассона

$$(nt)^r \frac{e^{-nt}}{r!},$$

а распределение общего числа требований, входящих в состав r групп, есть n-кратная свертка функции $B_n(t)$. Пусть $\Phi_n(s, t)$ — преобразование Лапласа—Стилтьеса вероятности $F_n(x, t)$ того, что в момент времени t число поступивших требований меньше x. На основании изложенного имеем

$$\Phi_n(s, t) = \frac{\sum_{r=0}^{\infty} (nt)^r e^{-nt} [\beta_n(s)]^r}{r!} = \exp\{-nt[1-\beta_n(s)]\}. \quad (14.123).$$

Предположим, что для входящего потока, непрерывного во времени, $F_n(x, t) \to F(x, t)$ при $t \to \infty$. Соответственно имеем $\Phi_n(s, t) \to \Phi(s, t)$. Условие непрерывности эквивалентно равенству

$$\lim_{n\to\infty} n[1-\beta_n(s)] = -\frac{\log \Phi(s, t)}{t}. \quad (14.124)$$

Левая часть этого выражения не содержит t; следовательно, и правая часть не содержит t, откуда вытекает существование решения для стационарного случая при непрерывном входящем потоке, коль скоро он имеет аддитивное распределение. В правой части последнего выражения можно произвести подстановку и определить $\psi(s)$.

В примере, который приводит Моран,

$$F(x, t) = \frac{1}{\Gamma(t)\rho^t} \int_0^x e^{-\frac{y}{\rho}} y^{t-1} dy. \quad (14.125)$$

В этом случае

$$\Phi(s, t) = (1+\rho s)^{-t} \quad (14.126)$$

и

$$\psi(s) = \frac{(1-\rho)s}{s - \log(1+\rho s)}. \quad (14.127)$$

Используя данное преобразование, определяем $H(z)$.

Введем обозначение

$$M(s) = \frac{1-\rho}{\rho}\psi(s). \quad (14.128)$$

Здесь член $1-\rho$ представляет значение вероятности равенства $Z=0$. Следовательно, $M(s)$ есть условное преобразование Лапласа $E(e^{-sZ} | Z > 0)$. Кендалл получил частное решение для нахождения обратного преобразования функции $M(s)$.

Описанный метод перехода от дискретного распределения вероятностей к непрерывному можно использовать также для определения функции распределения $H_k(z)$ для плотин с конечной емкостью водохранилища.

14.10. Больницы и медицинское обслуживание

Флегль, сотрудник Университета Джонса Гопкинса, построил подробную вероятностную модель, описывающую процесс оказания медицинской помощи [250]. Рассматривается определенная совокупность людей, которые могут заболевать случайным образом и обращаться за помощью к частнопрактикующим врачам, в поликлиники и больницы либо непосредственно, либо по этапам, например, больной—частнопрактикующий врач—клиника—больница—больной. После лечения пациент возвращается в совокупность. Не исключается также возможность, что пациент может покинуть совокупность и больше никогда в нее не возвратиться. Совокупность пополняется младенцами, поступающими из родильного дома; они рассматриваются как возможные пациенты, которым также может потребоваться медицинская помощь. Замечено, что потребность в обслуживании и характер требуемого обслуживания изменяются с возрастом. Таким образом, сам объект, обслуживаемый рассматриваемой системой, изменяется. Точка зрения персонала медицинского учебного заведения относительно цели этого процесса отличается от точки зрения пациента по этому вопросу, а у лиц, ответственных за работу учебного заведения, существуют свои взгляды на этот счет.

Важной проблемой является сокращение времени лечения. Для правильного управления этим видом обслуживания и оптимального размещения пациентов для более полного использования системы было проведено исследование интенсивности поступления пациентов на лечение и длительности обслуживания и проанализированы колебания общего числа пациентов, обслуженных в течение суток. Наблюдались (почти ежедневно) такие периоды, когда система была свободна, затем следовали периоды более интенсивного обращения за медицинской помощью. Для большинства больниц почти всегда наблюдается заметное колебание числа пациентов в зависимости от времени года.

Путем наблюдений было установлено, что поступление пациентов имеет пуассоновское распределение. Тяжело больные пациенты нуждаются в более длительном обслуживании; обычно проходит значительное время, пока такой больной начнет вставать с больничной койки (оставаясь в больнице), становится более независимым и самостоятельным и требует к себе значительно меньшего внимания. В случае наличия тяжело больных, которые первыми занимают обслуживающие устройства и раньше других получают обслуживание, рассматривается система с бесконечным числом каналов, если число таких больных не превышает емкости системы. Для группы тех пациентов, которые нуждаются в неотложной медицинской помощи и мало влияют на общую емкость, также применима система с бесконечным числом каналов. Практика показывает, что поток больных является пуассоновским, а время обслуживания имеет экспоненциальное распределение, за исключением случаев эпидемических заболеваний. Характер заболевания, а следовательно, и интенсивность поступления пациентов взаимно связаны.

Существует приоритет, прерывающий обслуживание, когда тяжело больной пациент поступает на лечение, а выздоравливающих пациентов просят покинуть больницу. Время лечения тяжело больных пациентов лучше всего описывается с помощью логарифмически нормального закона распределения (так как в данном случае дисперсия велика по сравнению с математическим ожиданием), который приближается к экспоненциальному распределению. Дисперсию можно уменьшить, объединяя усилия обслуживающего персонала по оказанию медицинской помощи. Это даст возможность принять большее число пациентов и лучше использовать имеющиеся средства обслуживания.

Анализ различных видов медицинской помощи, в которой нуждаются пациенты, показывает, что комбинация параллельных и последовательных каналов, действующих совместно, обеспечивает более эффективное обслуживание, чем ряд последовательных каналов, выполняющих те же функции. Так, два параллельных канала, каждый из которых может производить любое из двух видов требуемого обслуживания, могут действовать последовательно с двумя другими каналами, каждый из которых производит один из двух других видов обслуживания. Кроме того, в зависимости от потребности в обслуживании пациенты могут размещаться изолированно в различных обслуживающих устройствах.

Рассматривается также система, в которую пациенты прибывают с опережением или запаздыванием относительно назначенного времени приема. Когда становится возможным появление отклонения любой длительности, имеем пуассоновский процесс (он описывает предельный случай, когда происходит полная разладка расписания с запланированными моментами прибытия). Отклонения от графика последовательно связаны друг с другом.

Важным показателем эффективности для этого процесса является минимальная общая стоимость лечения. Сюда же включаются и производственные потери вследствие болезни. Всеобъемлющей аналитической модели, описывающей весь процесс, не существует. Процесс разбивается на отдельные части, которые затем исследуются аналитически.

Изучая статистические материалы, связанные с проектированием лечебных учреждений, Бейли [24] рассмотрел проблему максимального использования ограниченных ресурсов: искусства врачей, лекарств, помещений, денежных средств и т. д. Можно произвести оценку совокупности с некоторой ошибкой, т. е. определить число пациентов, обращающихся за медицинской помощью в данную группу больниц, производя выборку из списков адресов пациентов, поступивших в эту группу больниц в заданный период времени. Аналогичный анализ выполняется для соседних больниц, которые могут находиться во взаимодействии с этой группой больниц. Если в эту группу поступило a пациентов, а b пациентов поступило в соседние больницы, то доля пациентов, поступивших в эту группу больниц, равна $\frac{a}{a+b}$; после умножения этой величины на N — статистическую оценку объема совокупности — получим число пациентов, поступающих в данную группу больниц, определенное с некоторой погрешностью. Просуммировав по всем округам, получим общий эффективный объем совокупности, определенный также с некоторой погрешностью. Просуммировав по всем округам выражение $\frac{abN^2}{(a+b)^3}$, получим дисперсию этого числа. Допустим, что a и b — случайные величины, имеющие пуассоновское распределение. Если a и b являются оценками, основанными на постоянном выборочном весе λ, то приведенное выше выражение для дисперсии перед суммированием делится на λ.

Если характер обслуживания и длительность лечения известны, то время, затраченное на ожидание различного числа свободных коек, можно определить методами теории массового обслуживания. В больнице должно быть такое число коек, которое соответствует числу пациентов, обратившихся за помощью, и, кроме того, одна-две запасные койки на случай колебаний допускаемой длительности времени обслуживания. Влияние изменения интенсивности поступления пациентов (которая случайна) на среднее время ожидания не всегда можно объяснить. Напомним, что в исключительных случаях пациенты должны допускаться без ожидания либо за счет выделения запасных коек, либо за счет увеличения времени ожидания тех пациентов, которые не нуждаются в неотложной медицинской помощи.

Аналогичная ситуация имеет место и в случае амбулаторных больных. Можно установить такой порядок приема больных, при котором для пациента определяется допустимое среднее время ожидания, находится соотношение между ценностью времени доктора и ценностью времени пациента и допускается определенный риск иметь перерыв в работе. Назначенное время приема может колебаться, но в начале процесса не следует собирать большой очереди. Например, эксперимент, проведенный с помощью системы, в которой время приема больных было назначено заранее (средняя продолжительность консультации равнялась 5 минутам), показал, что для того чтобы процесс протекал устойчиво, консультация должна начинаться в тот момент, когда прибывает третий пациент, а промежутки времени между приемом пациентов должны быть постоянными, чуть большими 5 минут.

В другой статье [19] Бейли приводит вычисления, которые позволяют сделать вывод относительно путей уменьшения перегрузок в работе больниц. Перегрузки приводят к удлинению времени обслуживания и, таким образом, уменьшают емкость и пропускную способность лечебного учреждения, когда число имеющихся обслуживающих устройств ограничено.

В еще одной статье [22] Бейли рассматривает среднее время, в течение которого пациент должен ожидать, пока освободится одна из c больничных коек, если пациенты образуют пуассоновский поток с параметром λ, а время пребывания в больнице имеет экспоненциальное распределение с параметром μ [см. (4.130)]. Бейли показал, что длительное ожидание в очереди в значительной степени связано с постепенным накоплением невыполненного ранее обслуживания. Если бы с помощью экстренно принятых мер удалось устранить перегрузку, то время ожидания стало бы более коротким за счет лишь незначительного увеличения числа существующих коек. При этом максимальное число коек должно быть больше средней потребности в обслуживании. Можно использовать также формулу, позволяющую определить сокращение средней длительности лечения с целью достижения приемлемого времени ожидания. Некоторые работы Бейли, посвященные групповому обслуживанию, применяются для исследования оптимальных методов составления расписания приема амбулаторных больных.

14.11. Проектирование помещений для кафе

Детальное исследование важных вопросов, связанных с проектированием помещений для кафе, выполнил Чартренд [113]. Он провел всесторонний анализ причин задержек при обслуживании посетителей кафе. Чартренд рассмотрел такие вопросы, как сокращение времени ожидания и времени обслуживания; возможность обслуживания большего числа посетителей у буфетной стойки; раздельное размещение различных блюд и быстрая подача их посетителям, при этом обслуживание каждого клиента начинается раньше, чем в том случае, когда все блюда сосредоточены в одном месте. Чартренд показал, что задержки вызываются следующими причинами: 1) изменение числа постоянных клиентов, ожидающих в очереди; 2) медленное обслуживание вследствие того, что официанты плохо обучены, или вследствие того, что кафе неполностью укомплектовано обслуживающим персоналом; 3) непродуманное размещение горячих закусок и кофе, а также неудачное расположение кассовых аппаратов; 4) разнообразное меню.

Была поставлена задача установить равновесие между временем простоя обслуживающего устройства и временем ожидания посетителей. Для нахождения решения были рассмотрены: распределение входящего потока, число имеющихся буфетных стоек, виды обслуживания, а также различные дисциплины очереди. Было предложено параллельное размещение различных блюд (например, дессерта), так как при последовательном размещении блюд увеличивается среднее время обслуживания для всех посетителей. В случае параллельного размещения блюд клиент сразу выбирает то, что ему нужно, и становится в соответствующую очередь; при этом клиенты, прибывшие раньше и также производящие выбор закусок, не задерживают его, как это имеет место в случае последовательного размещения блюд. Но даже и при последовательном размещении можно избежать перегрузок, например, путем соответствующей группировки различных блюд.

Важным показателем является время продвижения в очереди одного клиента, когда его не подталкивают клиенты, стоящие сзади, и не задерживают клиенты, стоящие впереди; это время должно сокращаться. Нужно сокращать также и время, которое клиент тратит на обслуживание [ожидая, пока продвинутся клиенты, стоящие впереди (также находящиеся на обслуживании)]. Это можно сделать, например (в процессе решения этой задачи), путем сокращения числа посетителей, стоящих впереди. Обычно среднее время обслуживания клиента, который регулярно пользуется услугами кафе, меньше, чем время обслуживания клиента, который посещает кафе от случая к случаю.

Кроме тех вопросов, о которых уже упоминалось ранее, было рассмотрено также сокращение времени продвижения в очереди и времени обслуживания за счет того, что по окончании обслуживания посетители должны сразу покидать кафе. Можно добиться сокращения времени обслуживания также с помощью следующих мер: устранение многократных остановок при выборе блюд; обеспечение возможности свободного выхода из очереди; исключение необходимости обращаться к обслуживающему персоналу, например, при отпуске некоторых продуктов в расфасованном виде; выполнение минимального числа движений в процессе обслуживания. Те блюда, выбор которых производится в порядке самообслуживания, должны размещаться параллельно движению очереди. Было рассмотрено несколько схем процесса обслуживания и произведена их оценка.

14.12. Запасные забои угольной шахты

Тафт и Бутройд [839] рассмотрели следующую задачу, которая относится к типу задач об обслуживании станков. Для добычи необходимого количества угля нужно иметь m угольных забоев. Кроме этого числа забоев, нужно иметь также определенное число s запасных угольных забоев, чтобы во время подготовительных работ производить замену разрабатываемых забоев. Как только число разрабатываемых забоев становится меньше $m+s$, бригада начинает зачистку забоя. Работы ведут c бригад; следовательно, если требуется подготовить больше, чем c забоев, остальные забои должны ожидать. Прекращение работы в забое — пуассоновский процесс с параметром λ, а время зачистки забоя имеет экспоненциальное распределение с параметром μ.

Предположим, что разработка забоя и его зачистка производятся без задержки. Допустим, что все забои одинаковы (хотя на практике этого может и не быть). В действительности число работающих бригад при необходимости можно увеличить. При рассмотрении распределений случайных величин принимаются некоторые полезные упрощения. Если $P_n(t)$ — вероятность того, что в процессе подготовки и в процессе ожидания находится n забоев, то при $s \geqslant c$ имеем

$$P_n'(t) = \begin{cases} -(m\lambda + n\mu) P_n(t) + m\lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \\ \qquad\qquad 0 \leqslant n < c, \\ -(m\lambda + c\mu) P_n(t) + m\lambda P_{n-1}(t) + c\mu P_{n+1}(t), \\ \qquad\qquad c \leqslant n \leqslant s, \\ -[(m+s-n)\lambda + c\mu] P_n(t) + (m+s-n+ \\ +1)\lambda P_{n-1}(t) + c\mu P_{n+1}(t), \quad s < n \leqslant m+s; \end{cases} \tag{14.129}$$

при $s < c$

$$P_n'(t) = \begin{cases} -(m\lambda + n\mu) P_n(t) + m\lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \\ \qquad\qquad 0 \leqslant n \leqslant s, \\ -[(m+s-n)\lambda + n\mu] P_n(t) + \\ +(m+s-n+1)\lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \\ \qquad\qquad s < n < c, \\ -[(m+s-n)\lambda + c\mu] P_n(t) + \\ +(m+s-n+1)\lambda P_{n-1}(t) + c\mu P_{n+1}(t), \\ \qquad\qquad c \leqslant n \leqslant m+s. \end{cases} \tag{14.130}$$

и $P_{-1}(t) = P_{m+s+1}(t) = 0$.

Решение уравнений для стационарного состояния при $s \geqslant c$ имеет вид

$$p_n = \begin{cases} \dfrac{m^n \rho^n p_0}{n!}, & n \leqslant c, \\ \dfrac{m^n \rho^n p_0}{c!\, c^{n-c}}, & c \leqslant n \leqslant s+1, \\ \dfrac{m!\, m^s \rho^n p_0}{(m+s-n)!\, c!\, c^{n-c}}, & s \leqslant n \leqslant m+s. \end{cases} \tag{14.131}$$

При $s < c$ имеем

$$p_n = \begin{cases} \dfrac{m^n \rho^n p_0}{n!}, & n \leqslant s+1, \\ \dfrac{m!\, m^n \rho^n p_0}{n!\,(m+s-n)!} & s \leqslant n \leqslant c, \\ \dfrac{m! m^s \rho^n p_0}{c!\, c^{n-c}\,(m+s-n)!}, & c \leqslant n \leqslant m+s. \end{cases} \tag{14.132}$$

Из соотношения

$$\sum_{n=0}^{m+s} p_n = 1$$

находим p_0. Среднее число запасных забоев, сохраняемых одновременно, равно

$$\sum_{n=0}^{s} (s-n) p_n. \tag{14.133}$$

Средняя доля потерянного рабочего времени (когда $n>s$) равна

$$\frac{1}{m}\sum_{n=s+1}^{m+s}(n-s)p_n. \qquad (14.134)$$

Упражнение. Проверьте приведенные решения уравнений. Убедитесь в справедливости двух последних выражений.

Можно определить оптимальные значения s и c с тем, чтобы поддерживать состояние равновесия между стоимостью запасных забоев и стоимостью потерянного рабочего времени в действующих забоях. Можно использовать также графический анализ для получения решений, которые обычно трудно найти другими способами.

14.13. Шумы в полупроводниках как задача массового обслуживания

Этот вопрос представляет интерес, главным образом, для научных работников, занимающихся приложениями, и инженеров. Белл [38] исследовал шумы в полупроводниках как задачу теории массового обслуживания. Понятие об автокорреляционной функции уже давалось в задаче, помещенной в конце гл. 4. Энергетический спектр есть косинус-преобразование Фурье для автокорреляционной функции.

Очень важной проблемой при изучении шумов в полупроводниках является расчет энергетического спектра, интенсивность которого неограниченно возрастает при уменьшении частоты. Однако с помощью времени релаксации не всегда можно легко определить энергетический спектр. В некоторых случаях решение получить проще, если определение скорости, с которой носители зарядов покидают зону проводимости, рассматривается как задача массового обслуживания и не связывается со временем релаксации. Рассматриваемое явление связано с наличием тока; это означает, что носители зарядов перемещаются из центров, в которых они возникли, а время рекомбинации зависит от времени, необходимого для отыскания какого-нибудь другого свободного центра, и не зависит от вероятности попадания носителей зарядов обратно в те центры, где они возникли.

Задача определения шумов в полупроводниках представлена Беллом как задача массового обслуживания при следующих условиях:

1. Носители зарядов образуют очередь в зоне проводимости.

2. Элементы, составляющие очередь, присоединяются к ней в случайные моменты времени в результате возбуждения из исходного уровня; обычно это возбуждение является случайным процессом.

3. «Выводы обслуживающего устройства» — это свободные уровни, на которые могут возвращаться носители зарядов (или

возможные рекомбинационные центры, если не накладывается ограничений на способ возвращения носителей зарядов из зоны проводимости к уровням).

4. Длительность времени, в течение которого заняты выводы обслуживающего устройства, равна промежутку между моментом времени, когда уровень занимается носителем заряда, и моментом времени, в который он снова становится свободным, когда в результате возбуждения носителя заряда он снова возвращается с этого уровня в зону проводимости. Предполагается, что длительность этого промежутка времени описывается статистически с помощью экспоненциального распределения со значением математического ожидания, которое определяется энергией возбуждения через термодинамическое соотношение между «свободным» и «связанным» временем. (Это соотношение дает только значение пропорциональности, действительное время зависит также от «частотного скачка» релаксационного процесса.)

Рассматриваемая модель предполагает, что любая малая область полупроводника содержит: 1) некоторое число свободных мест, образованных теми возбуждениями носителей зарядов, которые происходят в этой области; 2) некоторое число носителей зарядов, определяемое числом возбуждений, происходящих в этой и смежных областях. Хотя оба эти числа имеют одинаковые средние значения, они не могут быть в точности одинаковыми в любой момент времени. В данном случае имеем ситуацию, напоминающую ту, которая рассматривалась в случае работы телефонных систем. Хотя общее число выводов коммутатора соответствует среднему числу поступающих вызовов, все же возможны локальные отклонения, которые приводят к полной занятости, и отдельные абоненты вынуждены ожидать в течение времени, длительность которого не имеет явной связи с длительностью процесса обслуживания.

Автокорреляционная функция имеет вид

$$\psi(\tau) = \int P(\geqslant t) P(\geqslant t + \tau)\, dt.$$

Здесь вероятность $P(\geqslant t)$ находится из уравнения (11.76) (так как число выводов велико и возбуждения возникают случайным образом), в котором функция Шлефли заменена функцией Ганкеля; между этими двумя функциями существует определенное соотношение. Функция Ганкеля табулирована. Несмотря на упрощение вероятности $P(\geqslant t)$, Белл не смог получить выражение для энергетического спектра из-за аналитических трудностей. Эти трудности заключаются в том, что формулы теории массового обслуживания не применимы при интенсивностях поступления требований, больших или равных интенсивности обслуживания; если выполняется подобное неравенство, поступление требований в систему образует расходящийся процесс; в случае выполнения равенства число требований, находящихся в очереди, становится неопределенным, одним из возможных значений является бесконечность. Аналогично ведет себя и время ожидания. Одной из трудностей аналитического представления энергетического спектра шумов в полупроводниках и является то, что в данном случае мы сталкиваемся именно с таким неопределенным случаем. То обстоятельство, что существование низкочастотного предела не было подтверждено экспериментально, т. е. интеграл энергетического спектра расходится, означает, что распределение времени существования также не имеет сходимости, что может иметь место только в такой системе массового обслуживания, в которой интенсивность поступления требований равна интенсивности обслуживания. По этому поводу см. также [749].

14.14. Другие приложения

Эдай [193] использовал различные положения теории массового обслуживания для анализа работы в установившемся состоянии пунктов сбора платы за проезд транспорта по шоссе. Он разработал оптимальную схему назначения различного числа сборщиков в различное время суток. Использование его метода позволило уменьшить расходы, связанные с организацией этого процесса, и обеспечило лучшее обслуживание.

С помощью методов теории массового обслуживания Брайхем [84] определил для заводов фирмы «Боинг» число рабочих, которое должно быть занято на выдаче инструментов. В кладовых завода хранится большое число различных инструментов, которые необходимы рабочим, занятым в мастерских и на сборочных линиях. Постановка задачи была вызвана жалобами мастеров, у которых рабочие подолгу простаивали в очереди за инструментом. Напрашивался вывод, что компания должна нанимать новых людей. Однако администрация, испытывая финансовые затруднения, была вынуждена сокращать накладные расходы и уменьшать число рабочих. Оптимальное число рабочих было определено с помощью математического расчета.

Сеспаньяк [760] рассмотрел вероятность ожидания $P(>0)$ в многоканальной системе с пуассоновским входящим потоком интенсивности a требований в единицу времени и экспоненциальным распределением времени обслуживания при обслуживании требований в порядке поступления, а также определил оптимальное число c контролеров, необходимое для проверки производственного процесса. Время обслуживания включает в себя также и время перехода на пункт обслуживания. Если $r=\frac{r_a}{r_c}$ — отношение стоимости единицы времени ожидания проверки к стоимости простоя контролера в течение одной временно́й единицы, то нормированная стоимость выражается как

$$(c-a)-\frac{raP(>0)}{c-a}, \quad c>a.$$

Получены графики, изображающие нормированную стоимость как функцию от a для различных значений c и различных значений r. Затем для фиксированных значений c и различных значений r строится семейство кривых, которые пересекаются с соответствующими кривыми для следующего значения c и того же самого значения r. Точки пересечения соединяются прямой линией. Другая линия проводится через точки пересечения этих кривых с теми, которые получены для малых значений c (по существу, это метод Брайхема). Зона, заключенная между этими двумя линиями, позволяет определить такое число контролеров, при котором обеспечивается минимальная нормированная стоимость, соответствующая данной интенсивности поступления требований.

Шиллер и Лейвин [754] применили моделирование с помощью метода Монте-Карло с целью определения протяженности новой площадки для погрузки и разгрузки автомобилей, которая оборудовалась перед объединенным складом. При этом должен был обслуживаться такой поток грузов, который раньше поступал к трем складам. Все исходные данные были разделены на 14 категорий в зависимости от грузоподъемности автомобиля, характера перевозимого груза, расположения склада и т. д. Грузовые автомобили были разделены на два класса, каждый из которых имел различное экспоненциальное распределение времени обслуживания. Движение автомобилей моделировалось для погрузочных площадок, способных принять одновременно 12, 15, 18 и 21 автомобилей. Было вычислено время ожидания и число ожидающих автомобилей. После рассмотрения процесса в течение нескольких суток и правильного усреднения оказалось возможным описать этот процесс с помощью теории массового обслуживания для площадок, способных принять различное число автомобилей.

Делькур [174] рассмотрел задачу, в которой требовалось определить, какое количество нефти должна подавать каждая из нескольких насосных станций, принимая во внимание вероятность допустимой длительности ожидания поступившей заявки. Поскольку общая потребность в нефти возрастает с каждым годом, то каждый насос должен будет работать в течение более длительного времени; это означает, что и время ожидания для поступающих заявок будет более длительным. Однако насосы не работают непрерывно. Таким образом, в пределах заданного плана обеспечения заявок важно установить продолжительность времени, в течение которого должны работать насосы, допуская возможность некоторой задержки в удовлетворении заявок.

Было найдено распределение входящего потока и распределение времени обслуживания для каждого насоса, проведена их статистическая проверка и приняты допущения, упрощающие вычисления. Затем с помощью теории массового обслуживания (рассматривалась система с пуассоновским входящим потоком и экспоненциальным временем обслуживания в стационарном состоянии) было определено среднее число ожидающих заявок, среднее время ожидания в очереди, среднее время пребывания в системе и т. д. Это позволило также определить, какое количество нефти можно перекачать дополнительно.

Возможны и другие приложения теории массового обслуживания (многие из перечисляемых здесь приложений теории вполне реальны). С помощью теории массового обслуживания можно исследовать прибытие пароходов в порт и их разгрузку; доставку грузов автомобилями; проход судов через узкие каналы, например через Суэцкий канал; высадку пассажиров в иностранных портах и проведение таможенного досмотра; ожидание в очереди к различным обслуживающим устройствам, например в парикмахерских, в магазинах самообслуживания и универмагах, в театрах, на стоянках такси, на автобусных остановках и т. д.; прохождение молекул газа через отверстие; действие катализаторов в химических реакциях; фильтрацию; почтовые операции; производственные процессы; технический осмотр автомобилей; работу конвейера; протекание нервных реакций; телеграфную связь и доставку почтовых отправлений, обладающих приоритетом; работу железнодорожных сортировочных станций.

В Соединенных Штатах теория массового обслуживания может найти важное применение в связи с тем, что вследствие нехватки судей растет число случаев, когда откладывается судебное разбирательсто и вынесение приговоров. Если бы удалось провести психометрические измерения процесса мышления и процессов возникновения импульсов и целенаправленных возбуждений, поступающих в нервную систему и головной мозг, то можно было бы исследовать процессы высшей нервной деятельности человека, выраженные через взаимное влияние различных факторов, через заявки на выполнение действий (умственных, физиологических или физических), через ошибки в назначении приоритетов, через задержки в обслуживании заявок с высоким приоритетом и т. д.

Задачи

1. Предположим, что автомобили прибывают к участку с однорядным движением, как принято в статье Таннера, имея экспоненциальное распределение промежутков времени между машинами с математическими ожиданиями, равными соответственно 0,2 и 0,5. Используя для моделирования этой задачи метод Монте-Карло, найдите среднее время ожидания для автомобилей, приближающихся с обеих сторон. Время, необходимое для проезда любого автомобиля через этот участок, равно 0,1 мин.

2. Воспользовавшись статьей Брайхема [84], вычерчите некоторые кривые, которые применяются при таком анализе.

3. Допустим, что

$$p_0 = \frac{\dfrac{(\lambda T)^c}{c!}}{\sum\limits_{k=0}^{c} \dfrac{(\lambda T)^k}{k!}}.$$

Вычертите ряд кривых для p_0 как функции от c для значений λT, лежащих в пределах от 0,01 до 9,0 [182]. Доу в письме редактору журнала *Operations Research* сообщает следующие две задачи:

а) За три года, в течение которых были собраны эти данные, с самолетов, базировавшихся на одном из авианосцев в Южной Америке, было снято 253 распределителя зажигания для проведения полной разборки и ремонта. Анализ записей в книге ремонта показал, что средняя продолжительность ремонта составила 37,6 дней. Обслуживающий персонал, несущий ответственность за сборочные работы и знакомый с требованиями безопасности, считает, что гарантированный срок хранения запасных распределителей зажигания составляет два года. Каким должно быть начальное число запасных деталей? Было показано, что $\lambda T=8{,}7$, $p_0=0{,}0059$; с помощью графиков было найдено $c=16$ запасным деталям.

б) После окончания обслуживания посетителей бара официантка должна сдавать деньги в кассу. В среднем за час она подает 16 чашек кофе, а затем в течение 45 минут ожидает, пока посетители закончат еду. Если каждая чашка кофе стоит 60 сентаво и официантка будет стремиться к тому, чтобы обслужить всех посетителей (допускается возможность потери лишь одного клиента из тысячи), то какую сумму денег она должна внести в кассу? В данном случае $\lambda T=12$, $p_0=0{,}001$, следовательно, $c=24$ и сумма наличных денег составит 144 песо. Один песо равен ста сентаво.

С помощью приведенной выше формулы покажите, что при пуассоновском входящем потоке с параметром $\lambda=2{,}24$ вызовов в час и средней продолжительностью разговора $T=0{,}00625$ часа при $p_0=0{,}001$ число каналов должно равняться двум.

4. Рассмотрим c-канальную систему с потерями, в которую поступает пуассоновский поток с параметром λ, а время обслуживания имеет экспоненциальное распределение с параметром μ. Предположим, что если занято $b<c$ каналов при фиксированном значении b и поступает новое требование, то все каналы оказываются занятыми или запертыми вследствие неисправности приборов. Однако каналы становятся свободными в моменты окончания обслуживания требований. Если все каналы заняты или заперты, то требования теряются. Если начальное число требований, имеющихся в момент времени $t=0$, равно i, то можно записать систему уравнений [795]

$$P'_{in}(t) = -(\lambda + n\mu)P_{in}(t) + \lambda P_{i,n-1}(t) + (n+1)\mu P_{i,n+1}(t), \quad (0 \leqslant n \leqslant b),$$

$$P'_{in}(t) = -n\lambda P_{in}(t) + (n+1)\mu P_{i,n+1}(t), \quad b-1 \leqslant n \leqslant c-1,$$

$$P'_{ic}(t) = -\mu c P_{ic}(t) + \lambda P_{ib}(t).$$

Покажите, что решение этих уравнений для стационарного состояния имеет вид

$$p_n = p_0 \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n, \quad n = 1, \ldots, b,$$

$$p_n = \frac{\lambda}{n\mu} p_b, \quad n = k+1, \ldots, c,$$

и из соотношения $\sum\limits_{n=0}^{c} p_n = 1$ определите p_0. Покажите, что при $b = c-1$ полученное решение превращается в формулу Эрланга. Найдите вероятность полной занятости $\sum\limits_{n=b+1}^{c} p_n$, т. е. вероятность того, что требование теряется. Вычислите интенсивность поступления требований, допускаемых к обслуживанию:

$$\lambda_0 = \sum_{n=0}^{b} \lambda p_n.$$

Заметим, что в этом случае вероятность полной занятости равна $\frac{\lambda - \lambda_0}{\lambda}$. Общая загрузка равна $\frac{\lambda}{\mu}$, а загрузка требованиями, допускаемыми к обслуживанию,— $\frac{\lambda_0}{\mu}$. Среднее число каналов, занятых реальными требованиями (в отличие от «ложных» требований, которые приводят к запиранию системы после того, как оказываются занятыми b каналов), равно

$$\sum_{n=0}^{b} np_n + cp_c. \tag{1}$$

Загрузка за счет «ложных» требований составляет

$$\sum_{n=b+1}^{c-1} np_n; \tag{2}$$

при $b = c - 1$ она равна нулю. Вычислите общее среднее число занятых каналов $\sum_{n=0}^{c} np_n$ и покажите, что оно равно сумме выражений (1) и (2).

5. В c-канальную систему с ожиданием поступает пуассоновский входящий поток с параметром λ, время обслуживания имеет экспоненциальное распределение с параметром μ. Предположим, что имеется q мест для ожидания. Если все эти места заняты, то вызов теряется. В этом случае нарушение работы состоит в том, что, когда в системе уже находится $c<b$ вызовов, вновь поступающий вызов приводит к запиранию всех остальных мест для ожидания. Таким образом, для состояний $b+1, \ldots, c+q$ единственно возможным переходом является окончание обслуживания.

Уравнения для стационарных вероятностей имеют вид [795]

$$\lambda p_{n-1} - (\lambda + n\mu) p_n + (n+1)\mu p_{n+1} = 0, \quad 0 \leqslant n \leqslant c-1,$$

$$\lambda p_{n-1} - (\lambda + c\mu) p_n + c\mu p_{n+1} = 0, \quad c \leqslant n \leqslant b,$$

$$-c\mu p_n + c\mu p_{n+1} = 0, \quad b+1 \leqslant n \leqslant c+q-1,$$

$$\lambda p_b - c\mu p_{c+q} = 0.$$

В данном случае нормирующим условием является

$$\sum_{n=0}^{c+q} p_n = 1.$$

Найдите решение этой системы уравнений и покажите, что при $b=c+q-1$ оно приводится к формуле Эрланга для смешанной системы, в которой может ожидать ограниченное число требований. Если $q=1$, то $b=c$. Покажите, что при бесконечном значении q и конечном значении b имеем $p_0=0$ и стационарное решение не существует. Если параметры $b=c+q-1$ и q равны бесконечности, то получим предельное распределение Эрланга для системы с потерями при $\rho = \frac{\lambda}{c\mu} < 1$. Найдите интенсивность поступления требований, допускаемых на обслуживание:

$$\lambda_0 = \lambda \sum_{n=0}^{b} p_n.$$

Покажите, что $\frac{\lambda_0}{\mu}$ — среднее число требований, допускаемых на обслуживание. Вычислите вероятность полной занятости

$$\sum_{n=b+1}^{c+q} p_n$$

и загрузку за счет потока ложных требований

$$c \sum_{n=b+1}^{c+q-1} p_n.$$

Среднее время ожидания для всего потока равно $\sum_{n=c}^{c+q} (n-c) p_n$. Покажите, что это выражение можно разбить на время ожидания для реальных требований и время ожидания для ложных требований. Вычислите загрузку, соответствующую поступлению ложных требований:

$$\sum_{n=b+1}^{c+q-1} n p_n.$$

Заметим, что вероятность того, что время ожидания n-го требования больше t, равна

$$W_n(\geqslant t) = \int_t^{\infty} \frac{(cnx)^{n-c}}{(n-c)!} e^{-\mu c x} c\mu dx, \quad n \geqslant c.$$

Отсюда найдем распределение времени ожидания

$$P(>t) = \sum_{n=c}^{b-1} W_n(\geqslant t) p_n + W_{c+q-1}(>t) p_b.$$

Вычислите это выражение при $b=c+q-1$, что справедливо для смешанной системы Эрланга. Если $q=\infty$, то получим экспоненциальное распределение, соответствующее классической системе с ожиданием. Покажите, что вероятность ожидания $1-P(=0)$ равна

$$p_c \frac{1-p^{b-c+1}}{1-\rho}.$$

Покажите, что среднее время ожидания равно

$$\frac{\rho[1-P(=0)] - p_{c+q}[q\rho + (b-c+1-q)]}{\lambda(1-\rho)}.$$

Покажите, что для нахождения среднего числа действительно ожидающих требований

$$\sum_{n=c}^{b} (n-c) p_n + q p_{q+c},$$

нужно умножить λ на среднее время ожидания.

6. В одноканальную систему поступает пуассоновский поток; время обслуживания в приборе, управляющем входящим потоком, имеет экспоненциальное распределение с параметром η; время обслуживания в канале имеет экспоненциальное распределение с параметром μ; требования обслуживаются в порядке

поступления. Покажите, что распределение времени ожидания в очереди (исключая время обслуживания) имеет вид

$$P(>t)=\sum_{i=1}^{\infty} P(i-1,1)\int_{t}^{\infty} g_0(i, y)\,dy+\sum_{i=1}^{\infty} Q(i-1, 0)\int_{t}^{\infty} g_1(i, y)\,dy.$$

Преобразование Лапласа функций g_0 и g_1, соответственно, имеет вид

$$g_1^*(i, s)=\sigma^i, \quad g_0^*(i, s)=\frac{\mu}{\mu+s}\sigma^{i+1}$$

и

$$\sigma=\frac{\mu}{\mu+s}\frac{\eta}{\eta+s}.$$

Здесь $g_1(i, t)$ — плотность вероятности того, что требование, поступившее в систему, когда в очереди находится $i-1$ требований, будет ожидать в течение времени t, и при этом оказывается, что управляющий прибор занят. Аналогично, $g_0(i, t)$ — та же плотность вероятности при свободном управляющем приборе. (Указание: общая длительность времени, затраченного после ожидания в очереди, равна сумме двух случайных величин, имеющих экспоненциальное распределение с параметрами η и μ.)

7. Покажите, что в аналогичном случае при бесконечном числе каналов имеем

$$P(>t)=e^{-(\eta-\lambda)t}-\left(1-\frac{\lambda}{\eta}\right)e^{-\eta t}.$$

Среднее время ожидания равно $\dfrac{\frac{\lambda}{\eta}}{\eta-\lambda}$ и $P(=0)=\dfrac{\lambda}{\eta}$.

8. Допустим, что в системе с N возможными позициями, установленными равномерно по окружности, в j-ю позицию поступает пуассоновский поток с параметром λ_j и обслуживается одним оператором, который движется по кругу все время в одном и том же направлении, последовательно обходя все позиции. Время обслуживания имеет экспоненциальное распределение с параметром μ. Если оператор занят, то требование теряется. Когда система свободна, оператор остается в позиции, которая обслуживалась последней. Когда поступает новое требование, оператор немедленно перемещается в эту позицию, чтобы произвести обслуживание поступившего требования. Пусть $P_j(t)$ — вероятность того, что в момент времени t оператор находится в j-й позиции и он свободен, а $Q_j(t)$ — вероятность того, что оператор находится в j-й позиции и занят обслуживанием требования. Пусть также

$$Q(t)=\sum_{j=1}^{N} Q_j(t), \quad P(t)=\sum_{j=1}^{N} P_j(t).$$

Заметим, что $P(t)+Q(t)=1$. Пусть $\lambda=\sum_{j=1}^{N}\lambda_j$. Вероятность того, что за время dt оператор, который находится в i-й позиции, будет вынужден переместиться в j-ю позицию и будет производить обслуживание требования, поступившего в j-ю позицию, равна $\lambda_j dt$. Если в j-й позиции нет требований, подлежащих обслуживанию, то при $i=j$ вероятность этого события равна $\lambda_j dt$, а при $i\neq j$ она равна нулю. Если оператор занят в i-й позиции, и в j-й позиции нет требований, ожидающих обслуживания, то эти вероятности равны μdt. Если в j-ю позицию поступает одно требование, то оно будет потеряно, так как оператор занят; значит, он останется в i-й позиции. Следовательно, вероятность того, что оператор переместится в эту позицию, равна $1-\mu dt$ при $i=j$ и равна нулю при $i \neq j$. Уравнения перехода имеют вид

$$P_j'(t) = \lambda_j Q(t) - \mu P_j(t),$$

$$Q_j'(t) = \mu P_j(t) - \lambda Q_j(t).$$

Решите эти уравнения для стационарного случая. Решите задачу для нестационарного случая при помощи преобразований Лапласа. Найдите решение для стационарного состояния, беря обратное преобразование Лапласа и переходя к пределу, а также путем вычисления значений $\lim\limits_{s \to 0} sP_j^*(s)$ и $\lim\limits_{s \to 0} sQ_j^*(s)$, когда рассматриваются преобразования Лапласа функций $P_j(t)$ и $Q_j(t)$ с параметром s (неопубликованная работа Сиски).

9. Допустим, что требования поступают на обслуживание из конечной совокупности объема N (такая система рассматривается в этой главе при решении задачи Энгсета—О'Делла) и интенсивность поступления требований пропорциональна числу требований, еще не находящихся в системе. Если A — число требований, находящихся в системе, то вероятность того, что в стационарном состоянии занято x каналов, полученная О'Деллом, аналогична формуле Эрланга для случая конечной совокупности:

$$\frac{\binom{N}{x}\left[\frac{A}{N-A}\right]^x}{1+N\left[\frac{A}{N-A}\right]+\binom{N}{2}\left[\frac{A}{N-A}\right]^2+\ldots+\binom{N}{x}\left[\frac{A}{N-A}\right]^x}.$$

Доля теряемых требований равна

$$\frac{A+(A-1)N\left[\frac{A}{N-A}\right]+(A-2)\binom{N}{2}\left[\frac{A}{N-A}\right]^2+\ldots+(A-x)\binom{N}{x}\left[\frac{A}{N-A}\right]^x}{1+N\left[\frac{A}{N-A}\right]+\binom{N}{2}\left[\frac{A}{N-A}\right]^2+\ldots+\binom{N}{x}\left[\frac{A}{N-A}\right]^x}.$$

Для системы с c параллельными каналами, в которой ожидание невозможно, время обслуживания имеет экспоненциальное распределение и число источников N конечно; выражение для вероятности потери требования, полученное Энгсетом, имеет вид

$$p_N = \frac{\binom{N-1}{c}\left\{\frac{\rho}{[1-\rho(1-p_N)]}\right\}^c}{\sum\limits_{k=0}^{c}\binom{N-1}{c}\left\{\frac{\rho}{[1-\rho(1-p_N)]}\right\}^k},$$

где $\rho = \frac{\lambda}{\mu}$; λdt — вероятность поступления в систему для каждого требования из конечной совокупности;

μ — интенсивность обслуживания.

Вероятность того, что все c каналов заняты, равна

$$\frac{\binom{N}{c}\left\{\frac{\rho}{1-\rho(1-p_N)}\right\}^c}{\sum\limits_{k=0}^{c}\binom{N}{k}\left\{\frac{\rho}{1-\rho(1-p_N)}\right\}^k}.$$

Определите, какая связь существует между этими двумя формулами? Предположим, что интенсивность поступления требований в систему равна $(N-i)a$. Выразите a через ρ и p_N.

10. В своей статье [139] Коэн детально рассмотрел телефонную систему с c каналами, имеющими одинаковое экспоненциальное время обслуживания с параметром $\frac{1}{h}$. Входящий поток — пуассоновский с параметром νs. Рассматривается вызов, который застал все каналы занятыми. Вызов можно повторить; время до повторного поступления вызова имеет экспоненциальное распределение с параметром νs, где ν — постоянная величина. Если поток, поступающий в группу каналов, обладает интенсивностью $a=sh$, то средний период повторения вызова равен $\frac{h}{\mu a}$. Кроме того, допустимое время ожидания вызова, поступившего во время полной занятости, имеет экспоненциальное распределение с параметром μs, где μ — постоянная величина. После того как допустимое время ожидания израсходовано, вызов теряется. Среднее время ожидания равно $\frac{h}{\mu a}$. Величины ν и μ являются показателями, описывающими характер абонента. Допустим, что как только число вызовов, поступивших во время полной занятости, становится равным m и их время ожидания не истекло, вновь поступающий вызов теряется и не может быть возобновлен. После определения $P(x, y, i)$ (вероятности того, что в момент времени t занято x каналов и имеется y возобновленных вызовов при начальных условиях $x=x_0$, $y=y_0$, $t=t_0$) можно найти также число занятых каналов, вероятность того, что все каналы заняты, и относительное число потерянных вызовов.

Покажите, что эта система описывается приведенными ниже уравнениями. Укажите, какие известные вам задачи массового обслуживания можно решить путем использования подобной модели.

$$\begin{aligned}
P(x, y, t+\Delta t) &= s\Delta t P(x-1, y, t) + \\
&+ \frac{(y+1) P(x-1, y+1, t)\nu a\Delta t}{h} + \frac{(y+1) P(x, y+1, t)\mu a\Delta t}{h} + \\
&+ \frac{(x+1) P(x+1, y, t)\Delta t}{h} + \left[1 - s\Delta t - y(\mu a + \nu a)\frac{\Delta t}{h} - \right. \\
&\left. - x\frac{\Delta t}{h}\right] P(x, y, t), \quad 0 < x < n, \ 0 \leqslant y < m,
\end{aligned}$$

$$\begin{aligned}
P(x, y, t+\Delta t) &= s\Delta t P(n-1, y, t) + \frac{(y+1) P(n-1, y+1, t)\nu a\Delta t}{h} + \\
&+ (y+1) P(n, y+1, t)\frac{\mu a\Delta t}{h} + s\Delta t P(n, y-1, t) + \\
&+ \left(1 - s\Delta t - \frac{y\mu a\Delta t}{h} - \frac{n\Delta t}{h}\right) P(n, y, t), \quad x = n, \quad 0 < y < m,
\end{aligned}$$

$$\begin{aligned}
P(x, m, t+\Delta t) &= s\Delta t P(x-1, m, t) + \frac{(x+1) P(x+1, m, t)\Delta t}{h} + \\
&+ \left[1 - s\Delta t - m(\mu a + \nu a)\frac{\Delta t}{h} - \frac{x\Delta t}{h}\right] P(x, m, t), \ 0 < x < n, \ y = m,
\end{aligned}$$

$$\begin{aligned}
P(n, m, t+\Delta t) &= s\Delta t P(n-1, m, t) + s\Delta t P(n, m-1, t) + \\
&+ \left(1 - ma\mu\frac{\Delta t}{h} - \frac{n\Delta t}{h}\right) P(n, m, t), \quad x = n, \ y = m.
\end{aligned}$$

Заметим, что $\sum\limits_{x=0}^{n}\sum\limits_{y=0}^{m} P(x, y, t) = 1$.

Эти уравнения обычным способом приводятся к системе дифференциально-разностных уравнений. Коэн получил решение этих уравнений для стационарного случая. Он показал, что стационарное состояние существует при $m \to \infty$. Были рассмотрены некоторые частные случаи для различных значений параметров μ, ν и m. Для систем с потерями и систем с ожиданием получены известные формулы Эрланга, Молина и Пальма. В частном случае, когда $m \to \infty$, $\mu \to 0$, $\nu \to 0$, $\lim\left(\frac{\nu}{\mu}\right) = \beta$, среднее время ожидания $\frac{h}{\mu a}$ и среднее время возобновления $\frac{h}{\nu a}$ неограниченно возрастают. Это означает, что вызов, поступивший во время полной занятости, всегда останется в системе и возобновляется лишь через очень продолжительный промежуток времени. Таким образом, распределение входящего потока, состоящего из первоначально и повторно поступающих вызовов, является пуассоновским с математическим ожиданием $(a+\varepsilon)h$, где ε должно быть определено. Покажите, что распределение числа занятых каналов является эрланговским, т. е.

$$f(x) = \frac{\frac{(a+\varepsilon)^x}{x!}}{N_n(a+\varepsilon)},$$

где

$$N_n(y) = 1 + y + \frac{1}{2!}y^2 + \ldots + \frac{1}{n!}y^n.$$

Покажите, что интенсивность потока принятых требований составляет

$$a_h = \sum_{x=0}^{n} x f(x) = (a+\varepsilon)\left[1 - E_n(a+\varepsilon)\right],$$

где

$$E_n(y) = f(n).$$

Если бы каждый возобновляемый вызов рассматривался как новый, то число вызовов, потерянных в единицу времени, было бы равно $\frac{(a+\varepsilon)E_n(a+\varepsilon)}{h}$. Однако вызов, поступивший во время полной занятости, возобновляется в среднем β раз, и поэтому число вызовов, потерянных в единицу времени, будет равно $\frac{a+\varepsilon}{h(1+\beta)} E_n(a+\varepsilon)$. Покажите, что интенсивность потока потерянных вызовов равна

$$\frac{a+\varepsilon}{1+\beta} E_n(a+\varepsilon).$$

Вычислите суммарную интенсивность потока a, равную сумме интенсивностей потоков потерянных и принятых вызовов, и покажите, что выражение для ε имеет вид

$$\varepsilon = \frac{\beta}{1+\beta}(a+\varepsilon)E_n(a+\varepsilon).$$

Глава 15

ЭЛЕМЕНТЫ ТЕОРИИ ВОССТАНОВЛЕНИЯ

15.1. Введение

В теории восстановления рассматриваются, главным образом, свойства процесса замены элементов, которые выходят из строя и заменяются новыми. Такими элементами могут быть, например, электрические лампочки, авиационные моторы и др. Когда элемент выходит из строя, он заменяется новым; когда выходит из строя и этот элемент, то его снова заменяют и т. д. Обычно представляет интерес среднее число восстановлений за промежуток времени и его дисперсия.

В данной главе будет дан вывод этих величин. Важность их можно легко показать, однако для вывода формул потребуются сложные выкладки. Полностью мы их не приводим, однако излагаем некоторые полезные методы, которые помогут читателю самому понять опущенные подробности доказательств, а также освоиться с упражнениями.

Процесс восстановления есть последовательность взаимно независимых неотрицательных одинаково распределенных случайных величин $\{X_i\}$. Здесь X_i обозначает продолжительность существования элемента, вводимого при i-й замене. Так, промежутки времени между требованиями, поступающими в систему массового обслуживания в соответствии с пуассоновским распределением с параметром λ, образуют процесс восстановления с общим законом распределения $F(t) \equiv 1 - e^{-\lambda t}$. Поток с ограниченным последействием также является процессом восстановления. Большинство глубоких результатов теории восстановления связано с асимптотическим поведением функций, так как обычно представляет интерес поведение функции восстановления (определяемой ниже) для продолжительных промежутков времени.

Запишем n-ю частичную сумму в виде $S_n = X_1 + \ldots + X_n$, тогда $S_n (n = 1, 2, \ldots)$ будет моментом времени, в который происходит n-е восстановление. Иногда добавляется дополнительная неотрицательная случайная величина X_0, не зависящая от $\{X_i\}$. В этом случае получаем *общий процесс восстановления* $\{X_0, X_1, \ldots\}$, где X_0 — остаточная продолжительность существования элемента исходной совокупности, имеющегося в начальный момент времени.

Опустим условие неотрицательности $\{X_i\}$, получим *обобщенный процесс восстановления.* Обозначив через Z_1 первую неотрицательную частичную сумму, а через $Z_1+\ldots+Z_k$ первую неотрицательную частичную сумму S_n, бóльшую $Z_1+\ldots+Z_{k-1}$, Блекуэлл получил процесс восстановления $\{Z_n\}$. Полученные таким способом Z_n являются неотрицательными одинаково распределенными случайными величинами.

Смит [770] обозначил через N_t наибольшее значение случайной величины n, для которой $S_n \leqslant t$. Он показал, что процесс восстановления есть частный случай процесса случайного блуждания при наличии поглощающих барьеров, когда отбор величин $\{X_i\}$ продолжается до тех пор, пока S_n в первый раз не превзойдет t, а N_t+1 есть такой объем выборки, при котором происходит остановка. Теория случайного блуждания приводит к таким результатам, как существование конечных моментов всех порядков случайной величины N_t, но другие методы дают более сильные результаты. В случае непрерывного процесса рассматривается функция восстановления $U(t) \equiv E(N_t)$. Начнем изучение теории восстановления с дискретного процесса.

Многие положения, которые приводятся в этой главе, взяты из превосходных работ Смита, посвященных этому вопросу. Как в случае дискретного процесса, так и в случае непрерывного процесса, рассмотрение начнем с вывода выражений для среднего числа элементов, заменяемых в данный момент времени. Полученные формулы являются основными в теории восстановления и находят широкое применение.

15.2. Дискретный процесс восстановления

Рассмотрим задачу восстановления элементов, например, электрических лампочек, выходящих из строя в процессе эксплуатации. Первая замена неисправных элементов начальной совокупности, существовавшей в момент времени $t=0$, происходит в момент времени $t=X_1$. Эти восстановленные элементы затем заменяются в момент времени $t=X_1+X_2$ и т. д.

Вначале рассмотрим случай, когда время разделено на дискретные промежутки. Наши рассуждения лишь незначительно отличаются от изложения этого вопроса в статье Феллера [236]. Пусть a_n — вероятность того, что новый элемент, установленный в данный момент времени, выйдет из строя ровно через n временных единиц. Можно принять допущение, что продолжительность существования элемента положительна, и поэтому $a_0=0$. Предполагается, что замена элементов производится без перерыва, и, как только элемент выходит из строя, он заменяется новым; при этом общее число элементов не уменьшается. Допускается, что исходная конечная совокупность элементов, равная N, не обязательно должна состоять только из новых элементов. Так, если первоначально имеется c_k элементов, продолжительность существования которых равна k временным единицам, то $N = \sum_{k=0}^{\infty} c_k$. Заметим, что продолжительность существования k может принимать любое значение (см. пример на стр. 440).

В данном случае для любых промежутков времени восстановление заключается в замене элементов, которые сами заменили другие элементы. Пусть $g_{i+1}(n)$ — среднее число элементов, которые в момент времени n были заменены $(i+1)$-й раз. Имеем следующее соотношение между числом элементов, заменяемых $(i+1)$-й раз, и числом элементов, заменяемых i-й раз:

$$g_{i+1}(n) = \sum_{k=0}^{n} g_i(n-k)\, a_k. \tag{15.1}$$

Чтобы показать, что это соотношение справедливо, заметим, что в момент времени n заменяются $(i+1)$-й раз те элементы, которые до момента времени n были заменены i раз и к моменту времени n вышли из строя. Из числа элементов, установленных i-й раз в момент времени $n-k$ $(0 \leqslant k \leqslant n)$, к моменту времени n выйдет из строя $g_i(h-k)\,a_k$ элементов. Таким образом, число элементов, заменяемых $(i+1)$-й раз, определяется суммированием по k. Это дает приведенное выше соотношение.

Если обозначить среднее число восстановлений, произведенных в момент времени n, через

$$u_n \equiv \sum_{i=1}^{\infty} g_i(n), \tag{15.2}$$

то, просуммировав равенство (15.1) по i, получим среднее число элементов, замененных в момент времени n:

$$u_n - g_1(n) = \sum_{k=0}^{n} u_{n-k} a_k. \tag{15.3}$$

Здесь $g_1(n)$ — среднее число элементов исходной совокупности (которая равна сумме чисел c_k), замененных в момент времени n. В данном случае можно допустить, что $u_0 = 0$. Значения a_k были определены только для новых элементов. В исходной совокупности могут находиться также элементы, бывшие в употреблении. Вероятность того, что в момент времени n такой элемент (с продолжительностью существования, равной k временным единицам, которая уже истекла) выходит из строя, зависит от r_k, вероятности того, что продолжительность существования элемента больше k временных единиц, т. е. $r_k = a_{k+1} + a_{k+2} + \ldots = 1 - \sum_{n=0}^{k} a_n$, если $\{a_n\}$ — такая функция распределения, что $\sum a_n = 1$. Таким образом, вероятность того, что в момент времени n из строя выйдет элемент, прослуживший уже k временных единиц, равна

$$\frac{a_{n+k}}{r_k}. \tag{15.4}$$

Так как имеется c_k элементов, продолжительность существования которых уже превысила k, то величина

$$\frac{c_k a_{n+k}}{r_k} \tag{15.5}$$

есть среднее число таких элементов, нуждающихся в замене в момент времени n. Если просуммировать по всем возможным значениям k, то получим среднее число элементов исходной совокупности, замененных в момент времени n. Таким образом,

$$g_1(n) = \sum_{k=0}^{\infty} \frac{c_k a_{n+k}}{r_k}, \tag{15.6}$$

и окончательно получим искомое выражение для среднего числа восстановлений в момент времени n в виде

$$u_n = \sum_{k=0}^{n} u_{n-k} a_k + b_n, \quad n \geqslant 1, \tag{15.7}$$

где $b_n = g_1(n)$.

Пример. Предположим, что при замене элементов в дискретные моменты времени вероятность того, что продолжительность существования элемента равна k временны́м единицам, имеет биномиальное распределение $\left[\frac{m!}{(m-k)!\,k!}\right] p^k q^{m-k}$, где m равно пяти временны́м единицам, а вероятность того, что продолжительность существования элемента не меньше одной временно́й единицы, равна 0,75. Допустим также, что исходная совокупность состоит из 100 элементов, из которых первоначально 10 элементов были новыми, 15 проработали в течение одной временно́й единицы, 25 — в течение трех временны́х единиц, а 50 — в течение четырех временны́х единиц. Заметим, что ни один элемент не может существовать более пяти временны́х единиц. С помощью биномиального распределения с точностью до двух десятичных знаков находим значения a_n, которые будем считать точными. Так, $a_0 = 0$; $a_1 = 0{,}01$; $a_2 = 0{,}09$; $a_3 = 0{,}26$; $a_4 = 0{,}40$; $a_5 = 0{,}24$. Ясно, что при $n > 5$ имеем $a_n = 0$. Тогда $r_0 = 1$; $r_1 = 0{,}99$; $r_2 = 0{,}90$; $r_3 = 0{,}64$; $r_4 = 0{,}24$ и $r_5 = 0$.

Используя имеющуюся информацию об исходной совокупности, запишем $c_0 = 10$, так как первоначально 10 элементов имели нулевой срок службы. Аналогично, $c_1 = 15$, $c_2 = 0$, $c_3 = 25$, $c_4 = 50$. Напомним, что максимальная продолжительность существования была принята равной пяти временны́м единицам. Следовательно, нельзя принимать во внимание элементы, продолжительность существования которых равна пяти единицам, так как они уже использованы.

Кроме того, имеем

$$g_1(1) = 10 \cdot 0{,}01 + \frac{15 \cdot 0{,}09}{0{,}99} + \frac{25 \cdot 0{,}40}{0{,}64} + \frac{50 \cdot 0{,}24}{0{,}24} = 67{,}09.$$

Аналогично:

$$g_1(2) = 14{,}21,$$

$$g_1(3) = 8{,}66,$$

$$g_1(4) = 7{,}63,$$

$$g_1(5) = 2{,}4.$$

При $n > 5$ имеем $g_1(n) = 0$. Окончательно находим $u_0 = 0$; $u_1 = 67{,}09$; $u_2 = 14{,}88$; $u_3 = 14{,}85$; $u_4 = 26{,}57$; $u_5 = 34{,}72$ и $u_6 = 23{,}16$. По истечении шести временных единиц суммарное среднее число восстановлений оказалось равным 181,27.

Заметим, что если нужно вычислить среднее число восстановлений для каждого момента времени, то можно записать выражения для производящих функций $a(s)$ и $b(s)$, которые дают возможность определить $u(s)$, откуда можно получить все значения u_n. Если корни знаменателя $u(s)$ известны, то $u(s)$ можно разложить на простые дроби (каждую из которых можно разложить в ряд); затем производится группировка коэффициентов, содержащих s_n. Этот коэффициент равен u_n. Общее число восстановлений равно сумме значений u_n, полученных при $s \to 1$ в $u(s)$.

Первое выражение в правой части формулы (15.7) есть свертка последовательностей $\{a_n\}$ и $\{u_n\}$. Введем производящие функции: $a(s) = \sum\limits_{n=0}^{\infty} a_n s^n$, $b(s) = \sum\limits_{n=0}^{\infty} b_n s^n$ и $u(s) = \sum\limits_{n=0}^{\infty} u_n s^n$, $u_0 = 0$, $b_0 = 0$. Применив производящую функцию к уравнениям (15.7) и вспомнив, что производящая функция свертки есть произведение производящих функций двух составляющих, получим

$$u(s) = u(s)\,a(s) + b(s).$$

Следовательно,

$$u(s) = \frac{b(s)}{1 - a(s)} \tag{15.8}$$

и

$$u_n \equiv \frac{1}{n!} \frac{d^n u(s)}{ds^n}\bigg|_{s=0}. \tag{15.9}$$

Заметим следующее.

1. Ряды, определяющие $a(s)$ и $b(s)$, сходятся при $|s| < 1$, так как их коэффициенты ограничены.

2. Если имеется фиксированная верхняя граница долговечности, то производящая функция $a(s)$ становится полиномом относительно s и ее существование гарантируется. Этот случай имеет наибольшее практическое применение.

3. Общее число восстановлений для всех моментов времени выражается через $\sum_{n=0}^{\infty} u_n$. Этот ряд сходится, если $\sum a_k < 1$, что можно видеть из (15.8), разложив $\frac{1}{[1-a(s)]}$ в ряд по степеням $a(s)$ для таких значений s, при которых $|a(s)| < 1$, т. е. для $|s| \leqslant 1$. Сходимость обеспечивается, так как $\sum_{n=0}^{\infty} b_n$ есть число элементов, замененных один раз. Оно равно числу элементов исходной совокупности N. Так как $b_0 = 0$, то

$$\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} \frac{c_k a_{n+k}}{r_k} = \sum_{k=0}^{\infty} c_k = N.$$

4. Если $a(1) = \sum a_k = 1$, то при $n \to \infty$ имеем

$$u_n \to \frac{b(1)}{\sum_{n=0}^{\infty} n a_n} = \frac{N}{\sum_{n=0}^{\infty} n a_n} \tag{15.10}$$

(за исключением периодического процесса, при котором все a_k равны нулю, кроме, возможно, тех, у которых индексы являются целыми числами, кратными данному фиксированному индексу, большему единицы).

По существу, эта теорема утверждает, что общее число восстановлений в любой момент времени стремится к числу элементов исходной совокупности, деленному на среднюю продолжительность существования элемента. Этот вывод согласуется с интуитивными соображениями. В случае периодического процесса в пределе и при $a(1) = 1$ $\{u_n\}$ становится периодической последовательностью: $u_{np+m} \to \frac{p(b_m + b_{p+m} + b_{p+2m} + \ldots)}{\sum_{n=0}^{\infty} n a_n}$ при $n \to \infty$, где p — период.

5. Обозначим через $c_k(n)$ число элементов, продолжительность существования которых в момент времени n составляет k временных единиц. Это число равно общему числу элементов, которые были установлены в момент времени $k-n$ и к моменту времени n не вышли из строя. Таким образом,

$$c_k(n) = u_{k-n} r_k. \tag{15.11}$$

Если $k > n$, то в нулевой момент времени эти элементы могут иметь продолжительность существования, равную $k-n$ временным единицам, и не выйти из строя к моменту времени n, т. е.

$$c_k(n) = \frac{c_k(0)\, r_k}{r_{k-n}}. \tag{15.12}$$

Используя (15.10) и (15.11), находим, что распределение предельной продолжительности существования равно

$$\frac{Nr_k}{\sum\limits_{n=0}^{\infty} na_n} \tag{15.13}$$

и не зависит от распределения исходной продолжительности существования.

15.3. Непрерывный процесс восстановления

Обозначим для непрерывного процесса восстановления через $f(x)dx$ вероятность выхода элемента из строя в промежутке времени dx; $f(x)=0$ для $x<0$. Получим следующее соотношение между числом элементов, заменяемых $(i+1)$-й раз, и числом элементов, заменяемых i-й раз в момент времени t:

$$g_{i+1}(t)=\int\limits_0^t g_i(t-x)f(x)\,dx, \tag{15.14}$$

так как из $g_i(t-x)$ элементов, замененных i-й раз в момент времени $t-x$, к концу промежутка времени x выйдет из строя $g_i(t-x)f(x)dx$ элементов.

Если введем функцию

$$u(t)\equiv\sum_{i=1}^{\infty} g_i(t), \tag{15.15}$$

то, как и ранее, получим среднее число восстановлений в промежутке $(t,\ t+dt)$:

$$u(t)=g_1(t)+\int\limits_0^t u(t-x)f(x)\,dx. \tag{15.16}$$

Если допустить, что исходная совокупность состоит из N элементов, то вероятность того, что в промежутке времени $(t;\ t+dt)$ произойдет первая замена элементов этой совокупности, равна

$$g_1(t)=Nf(t)\,dt. \tag{15.17}$$

Если рассмотреть совокупность $N=1$ и оценить число восстановлений в промежутке времени $(t,\ t+dt)$, получим

$$u(t)=f(t)\,dt+\int\limits_0^t u(t-x)f(x)\,dx. \tag{15.18}$$

Это выражение показывает среднее число восстановлений в промежутке времени $(t,\ t+dt)$. Суммарное среднее число восстановлений за время t описывается функцией восстановления, которая определяется из интегрального уравнения теории восстановления [эту функцию получим, проинтегрировав (15.18) от нуля до t]:

$$U(t)=F(t)+\int_0^t dy\int_0^y u(y-x)f(x)\,dx=$$

$$=F(t)+\int_0^t dx\int_0^t u(y-x)f(x)\,dy=$$

$$=F(t)+\int_0^t U(t-x)\,dF(x). \tag{15.19}$$

Заметим, что $U(t)$ — первый момент случайной величины N_t, т. е. $U_t=E(N_t)$, что будет показано далее. В данном случае

$$F(t)=\int_0^t f(x)\,dx. \tag{15.20}$$

Значительная часть теории восстановления связана с уравнением (15.19). Для существования $U(t)$ не требуется, чтобы функция $F(t)$ была абсолютно непрерывной, следовательно, существование функции $f(x)$ не обязательно. Позже мы дадим вывод $U(t)$, при котором не обязательно предполагается существование $f(x)$. Поскольку функция $U(t)$ известна, можно определить все характеристики процесса восстановления.

Решение приведенного выше уравнения можно найти в виде обратного преобразования Лапласа. Если функция $F(x)$ абсолютно непрерывна, т. е. существует функция $f(x)=F'(x)$, то к (15.16) можно применить преобразование Лапласа. Получим

$$u^*(s)=\frac{f^*(s)}{1-f^*(s)}.$$

Здесь звездочка является общепринятым обозначением преобразования Лапласа.

15.4. Некоторые свойства непрерывного процесса восстановления

Свойства дискретного процесса восстановления можно получить по аналогии со свойствами непрерывного процесса, которые здесь рассматриваются.

Используя функцию $f(t)=\mu e^{-\mu t}$, находим, что $U(t)=\mu t$, или $\frac{U(t)}{t}=\mu$ для любых t. Одна из задач состоит в том, чтобы при данном распределении вероятностей выхода из строя показать, что общее число восстановлений есть сумма неполных гамма-функций. Решая уравнение для $U(t)$, используя свойство преобразования Лапласа $\lim_{t\to\infty} u(t) = \lim_{s\to 0} su(s)$ и интегрируя этот результат от нуля до t, получаем, что $\frac{U(t)}{t} \approx h$, где h — обратная величина к математическому ожиданию продолжительности существования элемента.

Заметим, что мы используем возможность интегрирования асимптотического ряда и то обстоятельство, что конечные (или сходящиеся) асимптотические ряды можно перемножать. Приведенные выше примеры асимптотического поведения функции $U(t)$ при больших t подтверждаются важным положением, известным под названием элементарной теоремы восстановления.

Теорема 15.1. При $t \to \infty$

$$\frac{U(t)}{t} \to \mu_1^{-1}, \tag{15.21}$$

где μ_1 — первый момент распределения $dF(x)$.

Таким образом, общее число восстановлений для большого промежутка времени приближенно равно интенсивности появления неисправностей, умноженной на время. Доказательство этой теоремы будет дано в конце главы.

Что касается плотности восстановления, то необходимым условием того, чтобы функция $u(t) = U'(t)$ при $t \to \infty$ стремилась к пределу, является соотношение $f(x) \to 0$ при $x \to \infty$. *Теорема о плотностях восстановления* утверждает, что если при $x \to \infty$ функция $f(x) \to 0$ и если для некоторых значений $p > 1$ величина $|f(x)|^p$ интегрируема, то при $t \to \infty$ имеем $u(t) \to \mu_1^{-1}$. Для этого не требуется, чтобы выполнялось условие $\mu_1 < \infty$.

Теорема 15.2 (теорема Блекуэлла). Если для непрерывного процесса восстановления $\mu_1 < \infty$, а $\alpha > 0$ — постоянная, то при $t \to \infty$

$$U(t+\alpha) - U(t) \to \alpha\mu_1^{-1}. \tag{15.22}$$

Это частный случай *ключевой теоремы восстановления*, доказанной Смитом.

Теорема 15.3. При $t \to \infty$

$$\int_0^t Q(t-x)\,dU(x) \to \mu_1^{-1} \int_0^\infty Q(y)\,dy, \tag{15.23}$$

где $Q(t)$ — любая неотрицательная невозрастающая функция, заданная при положительных t и принадлежащая классу $L_1(0, \infty)$ функций, интегрируемых в смысле Лебега; при этом значение μ_1 может быть бесконечным.

Если $Q(t)=\frac{1}{a}$ для $0<t\leqslant a$ и $Q(t)=0$ для остальных t, то получим теорему Блекуэлла. Если для $dF(x)$ существует конечный второй момент μ_2, то, подставив

$$Q(t)\equiv 1-\mu_1^{-1}\int_0^t[1-F(x)]\,dx,$$

при $t\to\infty$ получим

$$U(t)-\frac{1}{\mu_1}\to\frac{\mu_2}{2\mu_1^2}-1. \tag{15.24}$$

15.5. Второй момент функции восстановления

При выводе второго момента найдем функцию восстановления другим способом. Второй момент $U_2(t)\equiv E(N_t^2)$ определяется, как обычно. Если написать

$$z_r=\begin{cases}1 & \text{для } S_r\leqslant t,\\ 0 & \text{для } S_r>t,\end{cases}$$

то $N_t=\sum_{r=1}^{\infty}z_r$ и $E(z_r)=P(S_r\leqslant t)\equiv F_r(t)$. Кроме того,

$$U(t)=E(N_t)=\sum_{r=1}^{\infty}E(z_r)=\sum_{r=1}^{\infty}F_r(t). \tag{15.25}$$

В данном случае характеристическая функция, т. е. преобразование Фурье—Стилтьеса суммы независимых в совокупности случайных величин, равна произведению их характеристических функций, которые в данном случае одинаковы. Обозначим это преобразование кружочком. Имеем

$$F_r^{\circ}(s)=[F^{\circ}(s)]^r,$$

$$U^{\circ}(s)=\sum_{y=1}^{\infty}[F^{\circ}(s)]^r=\frac{F^{\circ}(s)}{1-F^{\circ}(s)}, \tag{15.26}$$

где $F(x)$ — функция распределения.

Воспользовавшись теоремой о свертке, получим

$$U(t)=F(t)+\int_0^t U(t-x)\,dF(x).$$

Вывод этой промежуточной формулы оставим в качестве упражнения.

Теперь на основании определения z_r имеем $z_r = z_p z_q$, если $r = \max(p, q)$. Ясно, что $z_r^2 = z_r$. Используя эти соотношения, получаем

$$N_t^2 = \left(\sum_{r=1}^{\infty} z_r\right)^2 = z_1^2 + 2z_1z_2 + 2z_1z_3 + \ldots + 2z_2z_1 + $$
$$+ z_2^2 + 2z_2z_3 + \ldots + 2z_3z_1 + 2z_3z_2 + z_3^2 + \ldots =$$
$$= \sum_{r=1}^{\infty} z_r + 2\sum_{r=1}^{\infty} (r-1) z_r = \sum_{r=1}^{\infty} (2r-1) z_r.$$

Найдем выражение для второго момента $U_2(t) = E(N_t^2)$:

$$U_2(t) = U(t) + 2\int_0^t U(t-x)\, dU(x). \tag{15.27}$$

Вывод этой формулы также оставим в качестве упражнений. Моменты высших порядков находятся аналогичным образом.

При $\mu_2 < \infty$ для больших t с помощью ключевой теоремы восстановления Смит получил соотношение

$$\sigma^2(N_t) \approx \frac{\mu_2 - \mu_1^2}{\mu_1^3}\, t. \tag{15.28}$$

Если определить коэффициент вариации C как $\left(\frac{\sigma^2}{\mu_1^2}\right)^{\frac{1}{2}}$, где $\sigma^2 = \mu_2 - \mu_1^2$, а σ — среднее квадратическое отклонение, то формулу для дисперсии числа восстановлений за время t можно записать в виде

$$\sigma^2(N_t) \approx \frac{C^2 t}{\mu_1}. \tag{15.29}$$

Эвристическое доказательство справедливости этого выражения, при котором рассматривается последовательность $\{X_i\}$ и принимается допущение, что $f(x)$ существует, производится следующим образом (см. в гл. 3 о сумме случайного числа случайных величин):

$$E(X_1 + \ldots + X_{N_t}) = E(N_t)\mu_1. \tag{15.30}$$

Таким образом, среднее значение времени, при котором происходит N_t-е восстановление элемента, равно среднему числу восстановлений, умноженному на среднюю продолжительность существования элемента. Кроме того,

$$E\left[(X_1 + \ldots + X_{N_t}) - N_t\mu_1\right]^2 = E(N_t)\sigma^2. \tag{15.31}$$

Заметим, что это выражение не является дисперсией для S_{N_t}, а является математическим ожиданием квадрата разности между фактическим временем и произведением N_t на среднее значение μ_1. Выражение (15.31) следует непосредственно из формулы для дисперсии распределения суммы. Через продолжительный промежуток времени t, $X_1 + \ldots + X_{N_t} \approx t$, где N — число событий, появившихся за промежуток времени t, откуда

$$E(N_t) \approx \frac{t}{\mu_1}. \tag{15.32}$$

После подстановки этого выражения в формулу (15.31) последняя принимает вид

$$E(t - N_t\mu_1)^2 \approx \frac{\sigma^2 t}{\mu_1}. \tag{15.33}$$

Но

$$\sigma^2(N_t) = E[N_t - E(N_t)]^2 \approx E\left(N_t - \frac{t}{\mu_1}\right)^2, \tag{15.34}$$

поэтому

$$\sigma^2(N_t) \approx \frac{\sigma^2}{\mu_1^2}\,\frac{t}{\mu_1} = \frac{C^2 t}{\mu_1}. \tag{15.35}$$

Если и $\mu_3 < \infty$, то при $t \to \infty$

$$\sigma^2(N_t) \approx \frac{\mu_2 - \mu_1^2}{\mu_1^3}\,t + \frac{5\mu_2^2}{4\mu_1^4} - \frac{2\mu_3}{3\mu_1^3} - \frac{\mu_2}{2\mu_1^2} + o(1), \tag{15.36}$$

что является еще одним показателем рассеяния.

15.6. Остаточное время существования

Ясно, что для общего процесса восстановления с функцией распределения

$$G(x) = P(X_0 \leqslant x) \neq F(x),$$

описывающей остаточное время существования элемента, при помощи выкладок, аналогичных приведенным выше, получается формула

$$U(t) = G(t) + \int_0^t U(t - x)\,dF(x). \tag{15.37}$$

Используя тильду для обозначения преобразования Лапласа—Стилтьеса,

$$\widetilde{U}(t) = \int_0^t e^{-st}\,dU(t),$$

получим

$$\widetilde{U}(s) = \frac{\widetilde{G}(s)}{1 - \widetilde{F}(s)}. \tag{15.38}$$

Аналогично можно найти плотность восстановления.

Для этого случая Феллер рассмотрел обратное преобразование функции $u^*(s)$ и представил его в виде

$$u(t) = \sum_{k=0}^{\infty} C_k \exp(s_k t). \tag{15.39}$$

Здесь s_k — корни уравнения $f^*(s) = 1$, где $f^*(s)$ — преобразование Лапласа функции $f(t)$. Этот ряд абсолютно сходится при $t \geqslant 0$. Необходимым и достаточным условием для подобного представления является представление

$$u^*(s) = \sum_{k=0}^{\infty} \frac{C_k}{s - s_k} \tag{15.40}$$

и сходимость ряда $\sum_{k=0}^{\infty} |C_k|$. В этом случае имеем

$$C_k = \frac{-g^{*\prime}(s_k)}{f^*(s_k)}, \tag{15.41}$$

где $g(t) = G'(t)$.

Следует различать полученное распределение и известное распределение остаточного времени существования, имеющее вид

$$\mu \int_0^t [1 - F(x)]\, dx;$$

это распределение времени, которое в произвольный момент остается до следующего восстановления (т. е. вероятность того, что оставшееся время не больше t), или, что то же самое, оно является распределением времени, прошедшего до фиксированного момента времени с момента последнего восстановления. Данное распределение нам встречалось в формуле (9.13) (там мы рассматривали случай $\geqslant t$). В качестве упражнения читателю предлагается найти математическое ожидание и преобразование Лапласа этого распределения.

15.7. Замечания

1. На основании соотношения для процессов восстановления доказывается центральная предельная теорема

$$P(S_n \leqslant t) = P(N_t \geqslant n),$$

если $\mu_2 < \infty$ и $\sigma^2 = \mu^2 - \mu_1^2$, то при $t \to \infty$

$$P\left(N_t \geqslant \frac{t}{\mu_1} - \frac{\alpha\sigma}{\mu_1}\sqrt{\frac{t}{\mu_1}}\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy. \tag{15.42}$$

2. Случайный процесс, в который может быть вложен процесс восстановления, называется регенерирующим процессом. Это понятие уже использовалось в связи с вложенными цепями Маркова (точки регенерации), рассмотренными выше. Промежутки времени между этими точками взаимно независимы и имеют одинаковое распределение.

3. Предположим, что существует n независимых процессов восстановления, каждый из которых имеет свое собственное распределение. Теорема о суперпозиции утверждает, что когда математические ожидания каждого отдельного процесса стремятся к бесконечности таким образом, что при $n \to \infty$ сумма их обратных значений $\frac{1}{m}$ остается постоянной, то результирующий процесс приближается к процессу Пуассона с математическим ожиданием m интервала между событиями. Более подробно этот вопрос рассмотрен в работах Пальма [634] и Кокса и Смита [159].

4. Из регенерирующего процесса можно получить процесс накопления. Например, в теории массового обслуживания общее время занятости после момента времени $t=0$ или число клиентов, в момент прибытия которых очередь превышает N человек, образуют процесс накопления.

5. Дадим доказательство элементарной теоремы восстановления. Пусть $S_n = X_1 + \ldots + X_n$, тогда события $n = N_t + 1$ и X_{n+1}, $X_{n+2}, \ldots$ взаимно независимы. По определению имеем

$$E(N_t + 1) = 1 + U(t) < \infty.$$

Пусть $E(X_1) = \mu_1$. Среднее значение времени, когда производится $(N_1 + 1)$-е восстановление, приближенно равно общему среднему числу восстановлений $[U(t) + 1]$, умноженному на среднюю продолжительность существования элемента μ_1. Таким образом,

$$E(S_{N_t+1}) = E(N_t + 1)E(X_1) = [1 + U(t)]\,\mu_1.$$

Если записать $S_{N+1} = t + \xi_t$, где ξ_t — время, прошедшее после момента времени t (остаточное время существования элемента, используемого в момент времени t), то из предыдущего выражения находим

$$E(S_{N_t+1}) = t + E(\xi_t) = [1 + U(t)]\,\mu_1.$$

Воспользовавшись тем обстоятельством, что ξ_t — положительная величина, из последнего равенства получим

$$\lim_{t\to\infty} \inf \frac{U(t)}{t} = \frac{1}{\mu_1},$$

поскольку тогда $E(\xi_t) \to 0$.

Рассмотрим, наконец, еще один процесс восстановления

$$Y_i = \begin{cases} X_i & \text{при } X_i \leqslant \Delta, \\ \Delta & \text{при } X_i > \Delta, \end{cases}$$

где Δ — некоторая постоянная. Новая частичная сумма не превосходит соответствующей предыдущей частичной суммы; следовательно, новое значение N_t не меньше предыдущего значения N_t. Таким образом, это неравенство справедливо как для новой, так и для прежней функции восстановления. Обозначим их соответственно через $U^y(t)$ и $U^x(t)$. Следовательно,

$$\limsup_{t\to\infty} \frac{U^x(t)}{t} \leqslant \limsup_{t\to\infty} \frac{U^y(t)}{t} \leqslant \frac{1}{\mu_1^y},$$

где μ_1^y — математическое ожидание новой функции распределения $F^y(x)$. Выбирая Δ достаточно большим, можно сделать величину μ_1^y как угодно близкой к $\mu_1^x \equiv \mu$; тем самым теорема доказана.

Задачи

1. Найдите решение уравнения восстановления для непрерывного процесса с помощью преобразований Лапласа, предположив существование функции $f(x) = \frac{dF(x)}{dx}$.

2. Допустим, что требуется оценить восстановления совокупности элементов, которые выходят из строя вследствие n взаимно независимых видов неисправностей, имеющих распределение $\mu_j e^{-\mu_j x}$, и предположим, что доля используемых элементов, которые выходят из строя вследствие j-го вида неисправности, равна a_j, а всего элементов N. Общее распределение вероятностей появления неисправностей имеет вид

$$\sum_{j=1}^{n} a_j \mu_j e^{-\mu_j x}, \quad \sum_{j=1}^{n} a_j = 1.$$

Использовав функцию

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad a > 0 \text{ и } b > 0,$$

в качестве аппроксимации для распределения общего числа неисправностей, определите a и b, приравняв первый и второй моменты. Используйте полученную функцию распределения для вычисления с помощью преобразований Лапласа общего числа элементов, вышедших из строя к моменту t. Покажите, что полученный результат есть бесконечная сумма неполных гамма-функций. Оперируя с $u(t)$, покажите, что в этом случае справедлива элементарная теорема восстановления. Покажите, кроме того, что остается в силе теорема Блекуэлла и что формула для $U(t)$, содержащая два момента μ_1 и μ_2, также справедлива.

3. Можно ли вывести теорему Блекуэлла из элементарной теоремы восстановления, пользуясь разложением в асимптотический ряд?

4. Дайте приближенный вывод уравнения восстановления и уравнения для второго момента, используя функции $F_r(t)$.

5. Найдите разложение функции $u(t)$ в экспоненциальный ряд по степеням s_k, если при $x \geqslant 0$ функция $f(x)$ равна

1) $\mu e^{-\mu x}$,

2) $\frac{x^2 e^{-x}}{2}$.

6. Пусть $p_n(t) = F_n(t) - F_{n+1}(t)$ — вероятность того, что $N_t = n$ для больших t. Тогда для $\operatorname{Re}(a) > 0$

$$p_n(t) = \frac{1}{2\pi i} \int\limits_{a-\infty i}^{a+\infty i} \frac{e^{st}}{s} [1 - \widetilde{F}(s)] [\widetilde{F}(s)]^n ds.$$

Используя метод быстрейшего спуска, получим

$$p_n(t) \approx \frac{e^{\alpha t} [1 - \widetilde{F}(\alpha)] [\widetilde{F}(\alpha)]^n}{\alpha [2\pi n K''(\alpha)]^{\frac{1}{2}}}.$$

В данном случае α — единственный корень уравнения

$$K'(s) = -\frac{t}{n}$$

и

$$K(s) = \log \widetilde{F}(s).$$

Вычислите выражение для $p_n(t)$ при

$$f(t) = \frac{t^{n-1} e^{-t}}{\Gamma(n)}, \quad n > 0,$$

и при

$$f(t) = \begin{cases} 0, & t < \tau \\ e^{-t+\tau}, & t \geqslant \tau. \end{cases}$$

Здесь $F(s)$ — преобразование Лапласа—Стилтьеса функции распределения $F(x)$, которой соответствует плотность $f(x)$.

Метод быстрейшего спуска дает возможность получить асимптотическое представление

$$\int\limits_A^B f(x) e^{zg(x)} dx \approx \frac{f(\alpha) \sqrt{2\pi}\, e^{zg(\alpha)}}{\sqrt{-z g''(\alpha)}}$$

при $-\frac{\pi}{2} + \varepsilon \leqslant \arg z \leqslant \frac{\pi}{2} + \varepsilon$,

где $f(x)$ и $g(x)$ — аналитические функции в связной области комплексной плоскости, содержащей A и B, а α — единственная точка, где функция $g(z)$ имеет наибольшее значение. Для больших z значение этого интеграла изменится незначительно, если A и B заменить соответственно на $-\infty$ и ∞. Заметим, что посредством замены переменной интегрирования, можно сделать пределы интеграла, рассматриваемого в этой задаче, действительными числами.

ОБЩИЕ ЗАМЕЧАНИЯ. БИБЛИОГРАФИЯ

Некоторые общие замечания

В последние несколько лет происходило очень быстрое развитие теории массового обслуживания. Многие задачи, которые казались сложными, были поставлены и решены; так, например, были исследованы уравнения процесса рождения и гибели. Естественно попытаться сделать обзор теории массового обслуживания в более широком плане и построить единую теорию.

Обычно отдается предпочтение решениям, полученным с учетом временной зависимости, особенно когда стационарное состояние не существует. Наиболее важные случаи, в которых наблюдается такая зависимость, требуют дальнейшего рассмотрения (см. работу Вантюра, цитированную в гл. 9). До сих пор в большинстве допущений, принятых при рассмотрении задач массового обслуживания, предполагалась независимость распределения входящего потока и распределения времени обслуживания. Очевидно, что для многих задач это не соответствует действительному положению вещей и вызывается лишь сложностью анализа. Это положение встречается и во многих других областях, в которых наблюдается и необходимо присутствует зависимость.

Как видно из изложенного ранее, в теории массового обслуживания, если трактовать ее достаточно широко, должны рассматриваться разнообразные свойства входящих потоков. В теории должно рассматриваться большинство явлений, о которых упоминалось в гл. I, например неприсоединение к очереди, уход из очереди, свойства группового поступления и группового обслуживания и т. д. В теории массового обслуживания должен рассматриваться гораздо больший круг вопросов, чем тот, который представлен в этой книге. (Работы Б. В. Гнеденко [971—990] являются примером более разностороннего исследования.) В теории массового обслуживания должны изучаться все существенные дисциплины очереди, при этом нужно определить, какие дисциплины можно получить за счет варьирования параметров.

Дальнейшего исследования требуют также системы с параллельными и последовательными каналами. Если часть этих объединенных каналов специализирована, число их является переменным и характер очереди перед каналами изменяется, причем некоторые из каналов могут иметь ограничения в отношении числа ожидающих клиентов, то получим очень важную задачу, в которой содержатся различные обобщения систем массового обслуживания. В ходе исследований нужно решить вопрос о зависимости распределения выходящего потока от произвольного распределения входящего потока и произвольного распределения времени обслуживания. Последние работы по динамическим приоритетам указывают еще одно направление, в котором нужно вести работу — необходимо более детальное исследование тех явлений массового обслуживания, которые могут найти широкое применение. Соглашение между клиентами и переход клиентов из очереди в очередь являются примерами таких важных задач массового обслуживания, которые трудно формализировать.

Конечно, существует немало и простых задач, к решению которых необходимо приступать, как, например, произвольный входящий поток, произвольное распределение времени обслуживания, одноканальная система массового обслуживания с обслуживанием требований в порядке поступления; нестационарный случай. Предметом исследования могут быть также распределение числа требований, находящихся в очереди, распределение длительности периода занятости, а также распределение времени ожидания (исследование которого предприняли Кифер и Вольфовиц). Как видно из литературы, в этом направлении наблюдается большой прогресс.

Нередко рассмотрение таких вопросов приводит к постановке интересных математических задач. Решение их зависит, в частности, от того, насколько быстро развиваются исследования вероятностных процессов.

Моделирование процессов массового обслуживания для нахождения численных решений применяется при решении срочных и очень важных задач. Однако нередко эта операция оказывается неэффективной. Проведение моделирования целесообразно, когда трудно получить аналитическое решение и имеется вычислительная машина. Наличие вычислительной машины, если ею пользоваться разумно, может обеспечить постановку эксперимента при различных гипотезах и улучшить результаты аналитических исследований.

Чтобы получить более полное использование результатов теории массового обслуживания, в нее должна систематически вводиться оптимизация. В тексте мы приводили некоторые простые примеры оптимизации.

До сих пор еще не существует геометрического языка для описания систем массового обслуживания; теория графов не дает понятий, необходимых для полного представления. Чтобы установить связь понятий, полагаются на интуицию и прежнее знакомство с теорией массового обслуживания. Для решения этой важной задачи необходимо принять срочные меры, с тем чтобы улучшить связи между понятиями и направить творческое воображение на создание новых типов систем массового обслуживания как для теоретических, так и для прикладных целей.

Библиография

Приведенная библиография по теории массового обслуживания является сравнительно полной. Здесь представлена также литература по таким вопросам, как движение различных видов транспорта, телефонное дело, управление плотиной, управление запасами, обслуживание станков, теория восстановления и теория счетчиков, так как они тесно связаны с теорией массового обслуживания. По названиям этих работ исследователь всегда может найти статью, основные положения которой применимы к исследуемому им вопросу теории массового обслуживания. Хотя теория управления запасами и теория расписаний охватывают большое число исследователей и в некоторых аспектах связаны с теорией массового обслуживания, число работ по этим вопросам слишком велико, чтобы привести их полностью. В этом перечне приведены только наиболее важные из них.

При подготовке этого списка использовались библиографии Дойга и Лангера.

1. Adams, W. F.: Road Traffic Considered as a Random Series, *J. Inst. Civil Engrs. (London)*, vol. 4, pp. 121—130, 1936.

2. Adler, H. A., and K. W. Miller: A New Approach to Probability Problems in Electrical Engineering, *Trans. Am. Inst. Elec. Engrs.*, vol. 65, pp. 630—632, 1946.

3. Adler, R. B., and S. J. Fricker: The Flow of Scheduled Air Traffic, parts I and II, *Mass. Inst. Technol., Research Lab. Electronics, Tech. Repts.* 189, 199, 1951.

4. Adler, R. B., and S. J. Fricker: Notes on the Flow of Scheduled Air Traffic, *J. Roy. Aeronaut. Soc.*, vol. 58, pp. 475—484, 1954.

5. Adler, R. B., and S. J. Fricker: Notes on the Flow of Scheduled Air Traffic, *IRE Trans. on Aeronaut. Navigational Electronics*, vol. AN-2, p. 22, 1955.

6. Agard, J.: Problèmes d'attente dans une comagnie aérienne, *Rev. franç. recherche opérationelle*, vol. 4, no. 4, pp. 229—243, 1960.

7. Akaike, Hirotugu: On Ergodic Property of a Tandem Type Queueing Process, *Ann. Inst. Statist. Math. (Tokyo)*, vol. 9, pp. 13—22, 1957.

8. Albert, G. E., and L. Nelson: Contributions to the Statistical Theory of Counter Data, *Ann. Math. Statist.*, vol. 24, pp. 9—22, 1953.

9. Ancker, C. J., and A. V. Gafarian. Queueing with Multiple Poisson Inputs and Exponential Service Times, System Development Corporation, *Tech. Mem.* 503, Santa Monica, Calif., 1960.

10. Anis, A. A.: Statistical Aspects of Storage Problems, Thesis, London University, 1952.

11. Anson, C. J.: Determining the Number of Machines to be Attended by an Operator in Order to Minimize the Total Cost per Unit of Production, *Time and Motion Study J.*, vol. 6, pp. 13—27, April, 1957.

12. Arfwedson, G.; Research in Collective Risk Theory, part I, *Skand. Aktuar Tidskr.*, vol. 37, pp. 191—223, 1954.

13. Arley, N.: «On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation,» G. E. C. Cads Forlag, Copenhagen, 1943.

14. Arrow, K. J., S. Karlin, and H. Scarf: «Studies in the Mathematical Theory of Inventory and Production», Stanford University Press, Stanford, Calif., 1958.

15. Ashcroft, H.: The Productivity of Several Machines Under the Care of One Operator, *J. Roy. Statist. Soc., Ser. B*, vol. 12, pp. 145—151, 1950.

16. Atkinson, J.: «Telephony», vol. II, Sir Isaac Pitman & Sonc Ltd., London, 1950.

17. J. H. B.: A Partial Call Queueing Scheme for Sleeve-control Exchanges, *P. O. Elec. Engrs J.*, vol. 46, part I, April, 1953.

18. Bailey, N. T. J., and J. D. Welch: Appointment Systems in Hospital Out-patient Departments, *Lancet*, vol. 262, pp. 1105—1108, 1952.

19. Bailey, N. T. J.: Study of Queues and Appointment Systems in Outpatient Departments with Special Reference to Waiting Times, *J. Roy. Statist. Soc., Ser. B*, vol. 14, pp. 185—199, 1952.

20. Bailey, N. T. J.: A Continuous Time Treatment of a Simple Queue, Using Generating Functions, *J. Roy. Statist. Soc., Ser. B*, vol. 16, pp. 288—291, 1954.

21. Bailey, N. T. J.: On Queueing Processes with Bulk Service, *J. Roy. Statist. Soc., Ser. B*, pp. 80—87, 1954.

22. Bailey, N. T. J.: Queueing for Medical Care, *Appl. Statist.*, vol. 3, pp. 137—145, 1954.

23. Bailey, N. T. J.: A Note on Equalizing the Mean Waiting Time of Successive Customers in a Finite Queue, *J. Roy. Statist. Soc., Ser. B*, vol. 17, no. 1, pp. 262—263, 1955.

24. Bailey, N. T. J.: Statistics in Hospital Planning and Design, *Appl. Statist.*, vol. 5, no. 3, November, 1956.

25. Bailey, N. T. J.: «The Mathematical Theory of Epidemics», Hafner Publishing Company, New York, 1957.

26. Bailey, N. T. J.: Some Further Results in the Non-equilibrium Theory of a Simple Queue, *J. Roy. Statist. Soc., Ser. B*, vol. 19, no. 2, 1957.

27. Bähler, W. Th., J. W. Cohen, and M. M. Jung: Calculation of the Number of First Group-selectors for Telephone Systems Provided with First and Second Concentration Stages, Taking into Account the Internal Blocking, *Commun. News*, vol. 14, pp. 51—58, 1954.

28. Barrer, D. Y.: Queueing with Impatient Customers and Indifferent Clerks, *Operations Research*, vol. 5, no. 5, 1957.

29. Barrer, D. Y.: Queueing with Impatient Customers and Ordered Service, *Operations Research*, vol. 5, pp. 650—656, 1957.

30. Barry, J. Y.: A Priority Queueing Problem, *Operations Research*, vol. 4, pp. 385—386, 1956.

31. Bartlett, M. S.: «An Introduction to Stochastic Processes», Cambridge University Press, London, 1955.

32. Bathgate, R. J. P.: Automatic Telephone Traffic with Reference to Production Processes, *Trans. S. African Inst. Elec. Engrs.*, vol. 42, p. 384, 1951.

33. Bauer, F. L., and H. Störmer: Calculation of Waiting Time in Switching Systems with Small Input Line Groups, *Arch. Electrotech.*, vol. 9, pp. 69—73, 1955.

34. Bech, N. I.: Calculation of the Probabilities of Possible States of Occupancy in a Simple Group with Simple Ordered Searching, *Teleteknik*, vol. 5, pp. 342—349, 1954 (Danish; English summary).

35. Bech, N. I.: A Method of Computing the Loss in Alternative Trunking and Grading Systems, Copenhagen Telephone Company, Copenhagen (translation of article from *Teleteknik*, vol. 5, pp. 435—448, 1954).

36. Bech, N. I.: Statistical Equilibrium Equations for Overflow Systems, *Teleteknik* (English ed.), vol. 1, pp. 66—71, 1957.

37. Bekessy, A.: On the Probability Distribution of Pulse Counts Made by a Faulty Counter, *Magyar Tudományos Akad. Alkalm. Mat. Int. Közleményei*, vol. 3, pp. 171—181, 1955 (Hungarian; Russian and German summaries).

38. Bell, D. A.: Semiconductor Noise as a Queueing Problem, *Proc. Phys. Soc. (London)*, vol. 72, pp. 27—32, 1958.

39. Bell, G. E.: Theoretical Studies into Traffic Congestion, part II, *Brit. Ministry of Civil Aviation, Operation Research Sect., ORS/MCA Rept.* 3, June, 1948.

40. Bell, G. E.: Operational Research into Air Traffic Control, *J. Roy. Aeronaut. Soc.*, vol. 53, pp. 965—978, 1949.

41. Bell, G. E., D. Barnett, M. A. Young, and D. C. T. Bennett: Air Traffic Control, *J. Inst. Navigation*, pp. 211—250, November, 1950.

42. Bell, G. E.: Queueing Problems in Civil Aviation, *Operational Research Quart.*, vol. 3, pp. 9—11, 1952.

43. Bell, G. E., D. V. Lindley, K. D. Tocher, and J. D. Welch: Marshalling and Queueing, *Operational Research Quart.*, vol. 3, pp. 4—13, 1952.

44. Beneš, V. E.: A Continuous Time Treatment of the Waiting Time in a Queueing System Having Poisson Arrivals, a General Distribution of Service-time and a Single Service Unit (Preliminary Report), *Ann. Math. Statist.*, vol. 27, p. 872, 1956.

45. Beneš, V. E.: On Queues with Poisson Arrivals, *Ann. Math. Statist.*, vol. 28, pp. 670—677, 1957.

46. Beneš, V. E.: The Joint Distribution of a Set of Sufficient Statistics for the Parameters of a Single Telephone Exchange Model, *Ann. Math. Statist.*, vol. 28, p. 525, 1957.

47. Beneš, V. E.: A Note on Fluctuations of Telephone Traffic, *Ann. Math. Statist.*, vol. 28, 1957.

48. Beneš, V. E.: A Sufficient Set of Statistics for a Simple Telephone Exchange Model, *Bell System Tech. J.*, vol. 36, pp. 939—964, 1957,

49. Beneš, V. E.: Fluctuations of Telephone Traffic, *Bell System Tech. J.*, vol. 36, pp. 965—973, 1957.

50. Beneš, V. E.: Generalization of Palm's Loss Formula for Telephone Traffic, presented at Annual Convention of Institute of Mathematics, 1958.

51. Beneš, V. E.: The General Queue with One Server, presented at summer meeting of American Mathematics Society, 1958.

52. Beneš, V. E.: On Trunks with Negative Exponential Holding Times Serving a Renewal Process, *Bell System Tech. J.*, vol. 38, pp. 211—258, 1959.

53. Beneš, V. E.: General Stochastic Processes in Traffic Systems with One Server, *Bell System Tech. J.*, vol. 39, no. 1, pp. 127—160, January, 1960.

54. Beneš, V. E.: Combinatory Methods and Stochastic Kolmogorov Equations in the Theory of Queues with One Server, *Trans. Am. Math. Soc.*, vol. 94, pp. 282—294, 1960.

55. Beneš, V. E.: Transition Probabilities for Telephone Traffic, *Bell System Tech. J.*, vol. 39, no. 5, pp. 1279—1320, 1960.

*56. Beneš, V. E.: "General Stochastic Processes in the Theory of Queues", Reading, Mass., 1963.

57. Ben-Israel, A., and P. Naor: A Problem of Delayed Service, part I. *J. Roy. Statist. Soc.*, vol. 22., no. 2, pp. 245—269, 1960.

58. Ben-Israel, A., and P. Naor: A Problem of Delayed Service, part II, *J. Roy. Statist. Soc.*, vol. 22, no. 2, pp. 270—276, 1960.

59. Benson, F., and D. R. Cox: The Productivity ot Machines Requiring Attention at Random Intervals, *J. Roy. Statist. Soc., Ser. B*, vol. 13, pp. 65—82, 1951.

60. Benson, F.: Further Notes on the Productivity of Machines Requiring Attention at Random Intervals, *J. Roy. Statist. Soc., Ser. B*, vol. 14, pp. 200—210, 1952.

61. Benson, F., J. Miller, and M. W. H. Townsend: Machine Interference — A Solution of Some Workload Problems Arising When an Operator Has Charge of More than One Machine, *J. Textile Inst.*, vol. 44, pp. T619—T644, 1953.

62. Berkeley, G. S.: "Traffic and Trunking Principles in Automatic Telephony", Ernest Benn, Ltd., London, 1934 (reprinted, 1949).

63. Berkeley, G. S.: Traffic and Delay Formulae, *P. O. Elec. Engrs. J.*, vol. 29, pp. 188—195, 1936.

64. Berkowitz, S. M.: Analysis of a Fixed-block Terminal Area, Air Traffic Control System, *Franklin Inst. Final Rept.* F-2164-2, 1951.

65. Bernstein, P.: How Many Automatics Should a Man Run, *Factory Management and Maintenance*, vol. 99, pp. 85—164, 1941.

66. Berry, D. S., and F. H. Green: Techniques for Measuring Overall Speed in Urban Areas, *Highway Research Board, Proc.*, vol. 29, p. 311, 1949.

67. Berry, D. S., and D. M. Belmont: Distribution of Vehicle Speed and Travel Times, *Proc. Second Berkeley Symposium on Math. Statist. and Probability*, pp. 589—602, University of California Press, Berkeley, Calif., 1951.

68. Berry, D. S.: Field Measurement of Delay at Signalized Intersections, *Highway Research Board, Proc.*, vol. 35, pp. 505—522, 1956.

69. Beukelman, B. J., and J. W Cohen: Call Congestion of Transposed Multiples, *Philips Telecommun. Rev.*, vol. 17, no. 4, pp. 145—154, April, 1957.

70. Bharucha-Reid, A. T.: "Elements of the Theory of Markov Processes and Their Applications", McGrow-Hill Book Company, Inc., New York, 1960.

71. Bickelhaupt, C. O.: Mathematical Study of Toll Traffic, *Telephony*, vol. 77, no. 9, p. 12, 1919.

72. Bishop, D. J.: The Renewal of Aircraft, *Brit. Ministry of Aircraft Production, Aeronaut. Research Comm. Rept. and Mem.*, 1907 (6342), 1942.

73. Blackwell, D: A Renewal Theorem, *Duke Math. J.*, vol. 15, pp. 145—150, 1948.

74. Bloemena, A. R.: On Queueing Processes with a Certain Type of Bulk Service, *Math. Centrum, Statist. Afdel. Rept., Rept* SP-67, Amsterdam (mimeo. copy).

75. Blumstein, A.: The Operations Capacity of a Runway Used for Landings and Take-offs, Cornell Aeronautical Laboratory, Inc., Buffalo, N. Y., Feb. 11, 1960, prepared for Second International Conference on Operations Research, Aix-en-Provence, France, September, 1960.

76. Bodino, G. A., and F. Brambilla.: "Teoria delle code," Instituto Editoriale Cisalpino, Milano-Varese, 1959.

77. Boiteux, M.: Coûts et tarifs en face d'une demande aléatoire, *Econométrie*, pp. 223—227; Discussion (French), pp. 227—230 (*Coll. intern. centre natl. recherche sci. (Paris)*, no. 40, 1953.

78. Boldyref, A.: Determination of the Maximal Steady State Flow of Traffic through a Railroad Network, *J. Operations Research Soc. Am.*, vol. 3, pp. 443—465, 1955.

---

* Звездочкой отмечена библиография, добавленная переводчиком. — *Прим. ред.*

79. B o r e l, E.: Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'une infinité de coefficients, application au problème de l'attente à un guichet, *Compt. rend.*, vol. 214, p. 452, 1942.
80. B o w e n, E. G.: Operational Research into the Air Traffic Problem, *J. Inst. Navigation*, vol. 1, pp. 338—341, 1948.
81. B o w e n, E. G., and T. P e a r c e y.: Delays in Air Traffic Flow, *J. Roy. Aeronaut. Soc.*, vol. 52, pp. 251—258, 1948.
82. B r a d l e y, D. P.: Overflow Trunking of Switching and Discriminating Selector Repeaters, *Telecommun. J. Australia*, vol. 9, p. 254, 1953.
83. B r e t s c h n e i d e r, G.: The Calculation of Groups of Lines for Overflow Traffic in Automatic Telephone Networks, *NTZ — Nachrtech. Z.*, vol. 9, pp. 533—540, 1956.
84. B r i g h a m, G.: On a Congestion Problem in an Aircraft Factory, *J. Operations Research. Soc. Am.*, vol. 3, pp. 412—428, 1955.
85. B r i s b y, M. D. J., and R. T. E d d i s o n: Train Arrivals — Handling Costs, and the Holding and Storage of Raw Materials, *J. Iron Steel Inst. (London)*, vol. 172, pt. 2, pp. 171—183, October, 1952.
86. British Ministry of Civil Aviation, Operational Research Section: Notes and Data on Traffic Flow, 1950.
87. B r o a d h u r s t, S. W., and A. T. H a r m s t o n: An Electronic Traffic Analyser, *P. O. Elec. Engrs. J.*, vol. 42, pp. 181—187, 1949.
88. B r o a d h u r s t, S. W., and A. T. H a r m s t o n: Studies of Telephone Traffic with the Aid of a Machine, *Proc. Inst. Elec. Engrs. (London)*, pt. I, vol. 100, pp. 259—274, 1953.
89. B r o c k m e y e r, E., H. L. H a l s t r o m, and A. J e n s e n: "The Life and Works of A. K. Erlang" (all of Erlang's papers translated into English), Copenhagen Telephone Company, Copenhagen, 1948.
90. B r o c k m e y e r, E.: The Use of Probability Calculations in Telephone Engineering on the Basis of the Research of Erlang and Moe, *Teleteknik*, vol. 3, p. 95, 1952.
91. B r o c k m e y e r, E.: The Application of the Theory of Probability to Telephony, in J. Rybner (ed.), "Larebog in Telefonteknik," chap. 9, pp. 144—187, Copenhagen, 1949 (Danish).
92. B r o c k m e y e r, E.: "The Simple Overflow Problem in the Theory of Telephone Traffic", Copenhagen Telephone Company, Copenhagen (Danish); *Teleteknik*, vol. 5, pp. 361—375, 1955 (English synopsis).
93. B r o c k m e y e r, E.: A Survey of Traffic Measuring Methods in the Copenhagen Telephone Exchange., *Teleteknik* (English ed.), vol. 1, pp. 92—105, 1957.
94. B r o t m a n, L., and J. M i n k e r: Computer Simulation of a Complex Communications System, *Operations Research*, vol. 5, pp. 138—139, 1957.
95. B r o w n, A. W.: A Note on the Use of a Pearson Type III Function in Renewal Theory, *Ann. Math. Statist.*, vol. 11, pp. 448—453, 1940.
96. B r u c e, J. A, and J. B. R u d d e n: Denver's Traffic Control System Electronically Moves Auto Flow, *Western City*, November, 1953.
97. B r u n n s c h w e i l e r, D.: Machine Interference, *J. Textile Inst.*, vol. 45, no. 12, pp. T886—T895, 1954.
98. B u c h, K. R.: On a Special Use of the Erlang Methods in Industry, *Teleteknik* (English ed.) vol. 1, pp. 76—80, 1957.
99. B u r k e, P. J.: The Output of a Queueing System, *Operations Research*, vol. 4, pp. 699—704, 1956.
100. B u r k e, P. J.: Equilibrium Delay Distribution for One Channel with Constant Holding Time, Poisson Input and Random Service. *Bell System Tech. J.*, pp. 1021—1031, July, 1959.
101. B u r r o w s, C.: Some Numerical Results for Waiting Times in the Queue $E_k/M/1$, *Biometrika*, vol. 47, pts. 1 and 2, p. 202, 1960.
102. B u x t o n, H. A.: Estimating Trunking Requirements, *Telephony*, vol. 75, no. 2, 1918.
103. C a m p, G. D.: Some Approximate Solutions of the Single-step Queueing Equation, *J. Operations Research Soc. Am.*, vol. 3, p. 360, 1955.

104. Camp, G. D.: Bounding the Solutions of Practical Queueing Problems by Analitic Methods, in "Operations Research for Management", vol. II, pp. 307—339, Johns Hopkins Press, Baltimore, 1956.

105. Campbell, G. A.: Probability Curves Showing Poisson's Exponential Summation, *Bell System Tech. J.*, vol. 2, pp. 95—113, January, 1923.

106. Campbell, L. L.: Standard Deviation of Dead Time Correction in Counters, *Can. J. Phys.*, vol. 34, pp. 929—937, 1956.

107. Capello, F., and A. Sanneris: Studio ecomomico delle accessibilita in una centrale telefonica automatica, *Alta frequenza*, vol. 25, nos. 3—4, pp. 305—318, 1956.

108. Carlsson S. G., and A. Elldin: Solving Equations of State in Telephone Traffic Theory with Digital Computers, *Ericsson Tech.*, no. 2, 1958.

109. Champernowne, D. G.: An Elementary Method of the Solution of the Queueing Problem with a Single Server and Constant Parameter, *J. Roy. Statist. Soc., Ser. B*, vol. 18 no. 1, pp. 125—128, 1956.

110. Chandler, R. E. R. Herman, and E. W. Montroll: Traffic Dynamics: Studies in Car Following, *Operations Research*, pp. 165—184, March-April, 1958.

111. Chantal, R. de: Sur des probabilités relatives au trafic aérien dans les aéroports, *Travaux*, vol. 37, pp. 397—403, 1953.

112. Chantal, R. de: The Scientific Basis for Decision (Examples), *Travail et méthodes*, pp. 3—7, December, 1954; Review, *Operational Research Quart.*, vol. 6, p. 87, 1955.

113. Chartrand, M.: The Design of Cafeteria Counters, Norton Company Rept., Worcester, Mass., 1957.

114. Chauveau, J., and E. Vaulot: Extension de la formule d'Erlang au cas où le trafic est fonction du nombre d'abonnés occupés, *Ann. télécommun.*, nos. 8—9, 1949.

115. Chernoff, H.: The Distributions ot Shadows with Applications to Traffic and Counter Problems, *Ann. Math. Statist.*, vol. 27, p. 217, 1956.

116. Cherry, E. C.: The Conceivable Future of Telecommunications, *Trans. S. African Inst. Elect. Engrs.*, vol. 49, pt. 8, pp. 271—286, August, 1958.

117. Chew, W. G. N.: The Automatic Traffic Recorder, *P. O. Elec. Engrs. J.*, vol. 28, p. 1, 1935.

118. Chovet, A.: Determination of the Allowable Load for Telephone Operators, *Ann. télécommun.*, vol. 9, pp. 35—43, 1954.

119. Christensen, P. V.: Die Wählerzahl in Automatischen Fernsprechämtern, *Elektrotech. Z.*, vol. 34, p. 1314, 1913.

120. Churchman, C. W., R. L. Ackoff, and E. N. Arnoff: „Introduction to Operations Research", John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, pp. 389—416, 1957.

121. Clarke, A. B.: On the Solution of the "Telephone Problem", *Univ. Michigan Eng. Research Inst. Rept. R*-32, March, 1952.

122. Clarke, A. B.: The Time Dependent Waiting Line Problem, *Univ. Michigan Rept.* M720—1R39, 1953.

123. Clarke, A. B.: On Time-dependent Waiting Line Processes, *Ann. Math. Statist.*, vol. 24, pp. 491—492, 1953 (abstract).

124. Clarke, A. B.: A Waiting Time Process of Markov Type, *Ann. Math. Statist.*, vol. 27, pp. 452—459, 1956.

125. Clarke, A. B.: Maximum Likelihood Estimates in a Simple Queue, *Ann. Math. Statist.*, vol. 28, pp. 1036—1040, 1957.

126. Clayton, A. J. H.: Road Traffic Calculations, *J. Inst. Civil Engrs. (London)*, vol. 16, 1941.

127. Clifford, E. J.: The Measurement of Traffic Flow, *Traffic Eng.*, vol. 26, pp. 243—246, 1956.

128. Clos, C.: An Aspect of the Dialing Behavior of Subscribers and Its Effect on the Trunk Plant, *Bell System Tech. J*, vol. 27, pp. 424—445, 1948.

129. Clos, C., and R. I. Wilkinson: Dialing Habits of Telephone Customers, *Bell System Tech. J.*, vol. 31, pp. 32—67, 1952.

130. Clos, C.: A Study of Non-blocking Switching Networks, *Bell System Tech. J.*, vol. 32, pp. 406—424, 1953.

131. Clos, C.: Automatic Alternate Routing of Telephone Traffic, *Bell Labs. Record*, vol. 32, pp. 51—57, February, 1954.

132. Cobham, A.: Priority Assignment in Waiting Line Problems, *J. Operations Research Soc. Am.*, vol. 2, pp. 70—76, 1954; A Correction, *ibid.*, vol. 3, p. 547, 1955.

133. Cohen, J. W., and P. Harkema: A Study of the Delay Encountered in Telegraph Timerelay Switching, *Commun. News*, vol. 15, p. 47, 1954.

134. Cohen, J. W.: Berechnung der Verkehrsgrössen in Wartezeitsystem aus den Verkehrsgrössen eines Verlustsystems, *NTZ — Nachrtech. Z.*, vol. 8, p. 139, 1955.

135. Cohen, J. W.: Das Warteproblem für das Volkommene Bündel mit einer Endlichen Quellenzahl, *NTZ — Nachrtech. Z.*, vol. 8, pp. 641—645, 1955.

136. Cohen, J. W.: Some Examples in the Use of Implication in Switching Algebra, *Commun. News*, vol. 16, no. 1, pp. 1—10, October, 1955.

137. Cohen, J. W.: On the Queueing Process of Lanes, *Philips Tech. Rept.*, 1956

138. Cohen, J. W.: Certain Delay Problems for a Full Availability Trunk Group Loaded by Two Sources, *Commun. News*, vol. 16, pp. 105—113, 1956.

139. Cohen J. W.: Basic Problems of Telephone Traffic Theory and the Influence of Repeated Calls, *Philips Telecommun. Rev.*, vol. 17—18, 1956—1957.

140. Cohen, J. W.: The Full Availability Group of Trunks with an Arbitrary Distribution of the Interarrival Times and a Negative Exponential Holding Time Distribution, *Natuurk. Tydschr.* (*Ghent*), vol. 26, no. 4, pp. 169—181, 1957.

141. Cohen, J. W.: A Survey of Queueing Problems Occurring in Telephone and Telegraph Traffic Theory, *Internatl. Conf. Operations Research*, Oxford, 1957.

142. Cohen, J. W.: The Generalized Engset Formulae, *Philips Telecommun. Rev.*, vol. 18, no. 4, pp. 158—170, November, 1957.

143. Cohen, J. W.: On the Fundamental Problem of Telephone Traffic Theory and the Influence of Repeated Calls, *Philips Telecommun. Rev.*, vol. 18, pp. 49—100, 1957.

144. Cohen, J. W., and B. J. Beukelman: Call Congestion of Transposed Multiples, *Philips Telecommun. Rev.*, vol. 17, no. 4, pp. 145—154, 1957.

145. Conolly, B. W.: A Difference Equation Technique Applied to the Simple Queue, *J. Roy. Statist. Soc., Ser. B*, vol. 20, pp. 165—167, 1958.

146. Conolly, B. W.: A Difference Equation Technique Applied to the Simple Queue with Arbitrary Arrival Interval Distribution, *J. Roy. Statist. Soc., Ser. B*, vol. 20, pp. 167—175, 1958.

147. Conolly, B. W.: The Busy Period in Relation to the Single-server Queueing System with General Independent Arrivals and Erlangian Service-time, *J. Roy. Statist. Soc., Ser. B*, vol. 22, no. 1, pp. 89—96, 1960.

148. Conolly, B. W.: The Busy Period in Relation to the Queueing Process, GI/M/1, *Biometrika*, vol. 46, pp. 246—251, 1959.

149. Conolly, B. W. Queueing at a Single Serving Point with Group Arrival, *J. Roy. Statist. Soc.*, vol. 22, no. 2, pp. 285—298, 1960.

150. Conrad, R.: Performance of Telephone Operators Relative to Traffic Level, *Nature*, no. 4548, pp. 1480, December, 1956.

151. Conrad, R., and B. A. Hille: Telephone Operators' Adaptation to Traffic Variations, *J. Inst. Elec. Engrs.* (*London*), vol. 4., pp. 10—14, 1958.

152. Cox, D. R.: A Table for Predicting the Production from a Group of Machines under the Care of One Operator, *J. Roy. Statist. Soc., Ser. B*, vol. 16, pp. 285—287, 1954.

153. Cox, D. R.: A Use of Complex Probabilities in the Theory of Stochastic Process, *Proc. Cambridge Phil. Soc.*, vol. 51, pp. 313—319, 1955.

154. Cox, D. R.: Some Statistical Methods Connected with Series of Events, *J. Roy. Statist. Soc., Ser. B*, vol. 17, pp. 129—164, 1955.

155. Cox, D. R.: The Analysis of Non-Markovian Stochastic Processes by the Inclusion of Supplementary Variables, *Proc. Cambridge Phil. Soc.*, vol. 51, pt. 3, pp. 433—441, July, 1955 (see also *Math. Rev.*, p. 277, 1956).

156. C o x, D. R.: Prévision de la production d'un groupe de machinès surveillées par un exécutant, *Cahiers bur. temps elémentaires, sér. 5*, no. 502—503, p. 18, 1955 (French).

157. C o x D. R.: The Statistical Analysis of Congestion, *J. Roy. Statist. Soc., Ser. A*, vol. 118, pp. 324—335, 1955.

158. C o x, D. R., and W. L. S m i t h: The Superposition of Several Strictly Periodic Sequences of Events, *Biometrica*, vol. 40, pp. 1—11, 1953.

159. C o x, D. R., and W. L. S m i t h: On the Superposition of Renewal Processes, *Biometrica*, vol. 41, 1954.

*160. C o x, D. R., and W. L. S m i t h: "Queues", N. Y., John Wiley & Sons, 1961.

161. C o x, R. E.: Traffic Flow in an Exponential Delay System with Priority Categories. *Proc. Inst. Elec. Engrs. (London), Ser. B*, vol. 102, pp. 815—818, 1955.

162. C r a m e r, H.: "The Elements of Probability Theory and Some of Its Applications", 1st Am. ed., John Wiley & Sons, Inc., New York, 1955.

163. C r a n e, R. T., F. B. B r o w n, and R. O. B l a n c h a r d: An Analysis of a Railroad Classification Yard, *J. Operations Research Soc. Am.*, vol. 3, pp. 262—271; also "Operations Research by Management", chap. 7, Johns Hopkins Press, Baltimore, 1956.

164. C r o m m e l i n, C. D.: Delay Probability Formulae When the Holding Times are Constant, *P. O. Elec. Engrs. J.*, vol. 25, pp. 41—50, 1932.

165. C r o m m e l i n, C. D.: Delay Probability Formulae, *P. O. Elec. Engrs. J.*, vol. 26, pp. 266—274, 1934.

166. D a n t z i g, G. B.: A Comment on Edie's "Traffic Delays at Toll Booths", *J. Operations Research Soc. Am.*, vol. 2, pp. 339—341, 1954.

167. D a r u, E.: Nombres de machines automatiques à confier à un ouvrier en vue d'atteindre le prix de revient industriel minimum, *Rev. statist. appl.*, vol. 3, pp. 103—108, 1955 (French).

168. D a r u, E.: Mèthodes de resolution pour quelques problèmes de files d'attente comportant des "serveurs" d'efficacités différentes, *Rev. franç. recherche opérationelle*, vol. 2, no. 8, pp. 137—152, 1958.

169. D a v i s, H.: The Build-up Time of Waiting Lines, *Naval Research Logistics Quart.*, vol. 7, no. 2, p. 185, 1960.

170. D e B a u n, R. M., and S. K a t z: An Approximation to Distributions of Summed Waiting Times, *Operations Research*, vol. 7, no. 6, November—December, 1959.

171. D e b r y, R., and J. P. D r e z e: Files d'attente à plusieurs priorités absolues, *Cahiers centre d'études recherche opérationnelle*, vol. 2, no. 3, pp. 201—222, 1960.

172. D e s c a m p s, R.: Calcul direct des probabilités d'attente dans une file, *Rev. franç. recherche opérationnelle*, vol. 3, no. 2, pp. 88—100, 1959.

173. D e G o u r r i e r e c, B.: Economic Study of the Sub-division of the Lines at the Output of a Stage of Selection in Automatic Telephony, *Ann. télécommun.*, vol. 9, pp. 335—344, 1954 (French).

174. D e l c o u r t, J.: Équipement d'une station de pompage, *Rev. statist. appl.*, vol. 1, no. 2, pp. 77—85, 1959.

175. D e u t s c h, K. H.: Messung und Berechnung von Wartenzeiten in Unvolkommenen Bündeln bei Fernsprechwählanlagen, *Funk u. Ton*, no. 8, pp. 415—421, 469—478, 1950 (German).

176. D h o n d t, A.: Sur le comportement transitoire du processus d'attente simple, *Cahiers centre d'etudes recherche opérationnelle*, vol. 2, no. 3, 1960.

177. D h o n d t, A.: Le Problème transitoire de la file finie avec source illimitée, *Cahiers centre d'etudes recherche opérationnelle*, vol. 2, no. 4, pp. 309—318, 1960.

178. D i e t r i c h, H.: The Problem of the Optimum Placing of the Town Telephone Exchange, *Preglad Telekomunikacyjni*, no. 12, pp. 371—377, December, 1952 (Polish).

179. D o b b e n De B r u y n, M. V a n.: The Condition for Equilibrium in Automatic Telephone Traffic, *Ingenieur (Utrecht)*, vol. 59, pp. E1-12, 1947 (Dutch).

180. D o b b i e, J. M.: Comments on Expectancy of Multiple Vehicular Breakdowns in a Tunnel, *Operations Research*, vol. 4, pp. 609—613, 740—742, 1956.

181. D o b e r m a n n, H.: Formeln für den Mittleren Besetzteinfluss von Vielfach-Schaltungen, *FTZ — Fernmeldetech. Z.*, vol. 7, pp. 23—24, 1954.

182. D o e h, G: A Graphic Tool for the No-queue Model, *Operations Research*, vol. 8, no. 1, pp. 143—145, January—February, 1960.

183. D o i g, A.: A Bibliography on the Theory of Queues, *Biometrika*, vol. 44, pts. 3 and 4, pp. 490—514, December, 1957.

184. D o m b, C.: The Problem of Random Intervals on a Line, *Proc. Cambridge Phil. Soc.*, vol. 43, pp. 329—341, 1947.

185. D o m b, C.: Some Probability Distributions Connected with Recording Apparatus, *Proc. Cambridge Phil. Soc.*, vol. 44, pp. 335—341, 1948.

186. D o m b, C. Some Probability Distributions Connected with Recording Apparatus, part II, *Proc. Cambridge Phil. Soc.*, vol. 46, pp. 429—435, 1950.

187. D o o b, J. L.: "Stochastic Processes", John Wiley & Sons, Inc., New York, 1953.

Русский перевод: Д у б Дж. Л. Вероятностные процессы. Изд-во иностранной литературы, 1956.

188. D o w n t o n, F.: Waiting Times in Bulk Service Queues, *J. Roy. Statist. Soc., Ser. B*, vol. 17, pp. 256—261, 1955.

189. D o w n t o n, F.: On Limiting Distributions Arising in Bulk Service Queues, *J. Roy. Statist. Soc., Ser. B*, vol. 18, pp. 265—274, 1956.

190. D o w n t o n, F.: A Note on Moran's Theory of Dams, *Quart. J. Math. Oxford, Ser. 2*, vol. 8, pp. 282—286, 1957.

191. D r e s s i n, S. A.: Priority Assignment — Single Channel, *U. S. Marine Corps* (10 pages).

192. D u n n, P. F., C. F. F l a g l e, and P. A. H i c k s: The Queuiac — An Electromechanical Analog for the Simulation of Waiting Line Problems, *Operations Research*, vol. 4, pp. 648—662, 1956.

193. E d i e, L. C.: Traffic Delays at Toll Booths, *J. Operations Research Soc. Am.*, vol. 2, pp. 107—138, 1954.

194. E d i e, L. C.: Operations Research Techniques: Waiting Problems, *Operations Research*, section of *Proc. Soc. for Advancement of Management Conf.*, January, 1954.

195. E d i e, L. C.: Expectancy of Multiple Vehicular Breakdowns in a Tunnel, *J. Operations Research Soc. Am.*, vol. 3, pp. 513—522, 1955.

196. E d i e, L. C.: A Reply to Comments by James M. Dobbie, *Operations Research*, vol. 4, pp. 614—619, 1956.

197. E d i e, L. C.: Operations Research in a Public Corporation, *Operations Research*, vol. 5, pp. 111—122, 1957.

198. E d i e, L. C.: Optimization of Traffic Delays at Toll Booths, in "Operations Research for Management", vol. 2, chap. 3, Johns Hopkins Press, Baltimore, 1956.

199. E d i e, L. C.: Car Following and Steady-state Theory for Non-congested Traffic, *Operations Research*, vol. 9, no. 1, p. 66, January-February, 1961.

200. E g e r t, P.: The Mathematical Principles of Traffic Statistics, *Tech. u. Volkswirt. Berichte des Wirtschafts und Verkehrsministerium Nordrhein—Westfalen*, 34 pages, 1954 (German).

201. E k b e r g, S.: The Telephone Traffic Machine, *Tele.* (English ed.), vol. 3, pp. 150—155, 1953.

202. E k b e r g, S.: Telephone Traffic Research with the Aid of the Swedish Traffic Analyser, with two Appendices. Appendix 1, The Call Generating Equipment and Its Statistical Accuracy; Appendix 2, Exact Computation of a Grade of Service for Some Types of Interconnecting Arrangements, *Tele.* (English ed.), no. 1, pp. 1—29, 1955.

203. E k e l ö f, S.: Calculation of Delays in an Automatic Telephone System, *Ericsson. Rev.*, nos. 7—8, 1930.

204. E k e l ö f, S.: On the Hunting of Line Finders, *P. O. Elec. Engrs. J.* vol. 28, p. 52, April, 1935.

205. E l l d i n, A.: Traffic Measurements with Lamp Panel, *Ericsson Tech.* vol. 10, pp. 107—187, 1954.

206. Elldin, A., C. Jacobaeus, and L. von Sydow: Traffic Measurements with Lamp Panels, *Ericsson Tech.*, vol. 10, pp. 107—187, 1954.
207. Elldin, A.: Automatic Telephone Exchanges with Cross-bar Switches (Switches Calculations; General Survey), Switch Calculation Book B11265, Telefonaktienbolaget L. M. Ericsson, Stockholm, November, 1955.
208. Elldin, A.: On the Congestion in Gradings with Random Hunting, *Ericsson Tech. Ser. II*, vol. 2, pp. 33—37, 1955.
209. Elldin, A., and G. Lind: Statistical Methods for Supervision of Telephone Exchanges and Networks, *Ericsson Tech.*, no. 1, 1956.
210. Elldin, A.: Applications of Equations of State in the Theory of Telephone Traffic, Thesis, University of Stockholm, 1957.
211. Elldin, A.: Further Studies on Gradings with Random Hunting, *Ericsson Tech.*, vol. 13, pp. 175—257, 1957.
212. Elldin, A., and I. Tange: Long-time Observations of Telephone Traffic, *Tele.* (English ed.), no. 1, 1960.
213. Elliman, E. A., and R. W. Fraser: An Artificial Traffic Machine for Automatic Telephone Studies, *Elec. Commun.*, October, 1928.
214. Ellis, H. F.: *Punch*, May 30, 1951.
215. Ellis, H. F.: Written in a Queue, *Punch*, vol. 232, pp. 407—408, 1957.
216. Engset, T.: Die Wahrscheinlichkeits Rechnung zur Bestimmung der Wählerzahl in Automatischen Fernsprechämtern, *Electrotech. Z.*, vol. 31, p. 304, 1918.
217. Engset, T.: Emploi du calcul des probabilités pour la détermination du nombre de sélecteurs dans les Bureaux Téléphoniques Centreaux, *Rev. gén. elect.*, vol. 9, pp. 138—140, 1921.
218. Erdélyi, A. (ed.): "Tables of Integral Transforms", vol. 1, McGraw-Hill Book Company, Inc., New York, 1954.
219. Erlang, A. K.: Probability and Telephone Calls, *Nyt Tidsskr. Mat.*, *Ser. B*, vol. 20, pp. 33—39, 1909.
220. Erlang, A. K.: Solution of Some Probability Problems of Significance for Automatic Telephone Exchanges, *Elektroteknikeren*, vol. 13, pp. 5—13, 1917.
221. Erlang, A. K.: Lösung einiger Probleme der Wahrscheinlichkeitrechnung von Bedeutung für die selbsttätigen Fernsprechämter, *Electrotech. Z.*, vol. 39, p. 504, 1918 (German).
222. Erlang, A. K.: Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges, *P. O. Elec. Engrs. J.*, vol. 10, p. 189, 1918.
223. Erlang, A. K.: Telefon-Ventetider. Et Stykke Sandsynlighedsregning, *Mat. Tidsskr.*, *Ser. B*, vol. 13, pp. 25—42, 1920.
224. Erlang, A. K.: Solution de quelques problèmes de la théorie des probabilités presentant de l'importance pour les Bureaux Téléphoniques Automatiques, *Ann P. T. T.*, vol. 11, p. 800, 1922 (French).
225. Erlang, A. K.: Sandsynlighedsregningens Anvendelse i Telefondrift, *Elektroteknikeren*, vol. 19, pp. 99—110, 1923.
226. Erlang, A. K.: Application du calcul des probabilités en téléphonie, *Ann. P. T. T.*, vol. 14, p. 617, 1925 (French).
227. Erlang, A. K.: Problem 15 in *Matematisk Tidsskrift*, vol. B, p. 36, 1929. (This is Erlang's version of the problem of the extinction of surnames.)
228. Everett, J. L.: State Probabilities in Congestion Problems Characterized by Constant Holding Times, *J. Operations Research Soc. Am.*, vol. 1, pp. 279—285, 1953.
229. Everett, J. L.: Seaport Operations as a Stochastic Process, *J. Operations Research Soc. Am.*, vol. 1, p. 76, 1953.
230. Fabens, A. J.: The Solution of Queueing and Inventory Models, *Stanford Univ.*, *Applied Math. and Statist. Labs. Tech. Rept. 20*, Dec. 7, 1959.
231. Fagen, R. E., and J. Riordan: Queueing Systems for Single and Multiple Operations, *J. Soc. Ind. Appl. Math.*, vol. 3, pp. 73—79, 1955.
232. Farlie, D. J.: Canteen Queues, *Incorp. Statistician*, vol. 7, pp. 73—84, 1956.

233. F e l d e r, H. H., and E. N. L i t t l e: Intertoll Trunk Net Loss Maintenance under Operator Distance and Direct Distance Dialling, *Bell System Tech. J.*, vol. 35, pp. 955—972, 1956.

234. F e l l e r, W.: On the Theory of Stochastic Processes, with Particular Reference to Applications, First Berkeley Symposium on Mathematical Statistics and Probability, pp. 403—432, University of California Press, Berkeley, Calif., 1949.

235. F e l l e r, W.: Fluctuation Theory of Recurrent Events, *Trans. Am. Math. Soc.*, vol. 67, pp. 98—119, 1949.

236. F e l l e r, W.: "An Introduction to Probability Theory and Its Applications", John Wiley & Sons, Inc., New York, 1950 (first edition); 1957 (second edition).

Русский перевод: Ф е л л е р В., Введение в теорию вероятностей и ее приложения. 1-е изд. Изд-во иностранной литературы, 1952. 2-е изд. "Мир", 1964.

237. F e l l e r, W.: Diffusion Processes in One Dimension, *Trans. Am. Math. Soc.*, vol. 77, p. 1—31, 1954.

238. F e l l e r, W.: On Boundaries and Lateral Conditions for Kolmogorov Differential Equations, *Ann. Math.*, vol. 65, no. 2, pp. 527—570, 1957.

239. F e l l e r. W.: The Birth and Death Processes as Diffusion Processes, Princeton University, Princeton, N. J., 1958.

240. F e t t e r, R. B.: The Assignment of Operators to Service Automatic Machines, *J. Ind. Eng.*, vol. 6, pp. 22—30, 1955.

241. F i e l d, J. W.: Machine Utilization and Economical Assignment, *Factory Management and Maintenance*, vol. 104, pp. 288—296, 1946.

242. F i n c h, P. D.: The Effect of the Size of the Waiting Room on a Simple Queue, *J. Roy. Statist. Soc., Ser. B*, vol. 1, pp. 182—186, 1958.

243. F i n c h, P. D.: Cyclic Queues with Feedback, *J. Roy. Statist. Soc., Ser. B*, vol. 21, no. 1, pp. 153—157, 1959.

244. F i n c h, P. D.: Balking in the Queueing System $GI/M/1$, *Acta Math. Acad. Sci. Hung.*, vol. 10, no. 1/2, pp. 241—247, 1959.

245. F i n c h, P. D.: The Output Process of the Queueing System $M/G/1$, *J. Roy. Statist. Soc., Ser. B*, vol. 21, no. 2, pp. 375—380, 1959.

246. F i n c h, P. D.: On the Transient Behavior of a Simple Queue, *J. Roy. Statist. Soc.*, vol. 22, no. 2, pp. 277—283, 1960.

247. F i n c h, P. D.: Mixed Customer Impatience in a Simple Queue, paper — London School of Economics, Research Techniques Unit (mimeo. paper).

248. F i n c h, P. D.: On Distributions of Queue Size in Queueing Problems, London School of Economics, Research Techniques Unit.

249. F l a g l e, C. D., I. W. G a b r i e l s o n, A. S o r i a n o, and M. M. T a y l o r: Analysis of Congestion in an Outpatient Clinic, Johns Hopkins University, Operation Research Division (study conducted).

250. F l a g l e, C. D.: The Problem of Organization for Hospital Impatient Care, paper presented at the Sixth Annual International Meeting of the Institute of Management Science, Paris, September, 1959; Reprint 55, Pergamon Press Ltd., London.

251. F l a g l e C. D.: Queuing Theory and Cost Concept Applied to a Problem in Inventory Control, in "Operations Research Management", vol. 11, pp.160—177, Johns Hopkins Press, Baltimore, 1956.

252. F o g e l, L. J.: An Analytic Approach toward Air Traffic Control, *IRE Trans. on Aeronaut. Navigational Electronic*, vol. AN-2, no 2, 1955.

253. F o r b e s, T. W.: Statistical Techniques in the Field on Traffic Engineering and Traffic Research, *Proc. Second Berkeley Symposium on Math. Statist. and Probability*, pp. 603—625, 1951.

254. F o r d, L. R., and D. R. F u l k e r s o n: A Simple Algorithm for Finding Maximal Network Flows and an Application to the Hitchcock Problem, The Rand Corporation, P-743 (alternate numbering RM1604), September, 1955.

255. F o r t e t, R.: Probabilitè de perte d'un appel téléphonique: Régime nonstationnaire, influence du temps d'orientaton et du groupement des lignes, *Actes colloq. du calcul des probabilités de Lyon*, June, 1948.

256. F o r t e t, R.: Sur la probabilité de perte d'un appel téléphonique, *Compt. rend.*, vol. 226, pp. 1502—1507, 1948.

257. F o r t e t, R.: Sur la probabilité de perte d'un appel téléphonique dans un groupe de x sélecteurs, commandés par un orienteur unique, *Compt. rend.*, vol. 226, p. 159, 1948.
258. F o r t e t, R.: Évaluation de la probabilité de perte d'un appel téléphonique, compte tenu du temps d'orientation et du groupement des lignes, *Ann. télécommun.*, vol. 5, 1950.
259. F o r t e t, R.: Calcul des probabilités, in „Application des théories mathématiques", part I, Centre national recherche scientifique, Paris, 1950.
260. F o r t e t, R., and A. B l a n c-L a p i e r r e: "Théorie des fonctions aléatoires", Masson et Cie, Paris, 1953.
261. F o r t e t, R.: Les fonctions aléatoires en téléphonie automatique. Probabilités de perte en sélection conjuguée, *Ann. télécommun.*, vol. 11, pp. 85—88, 1956.
262. F o r t e t, R.: Random Distributions with an Application to Telephone Engineering, Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 11, 1956.
263. F o r t e t, R., and B. C a n c e i l l: Probabilités de perte en sélection conjuguée, *Teleteknik* (English ed.), vol. 1, pp. 41—55, 1957.
264. F o s t e r, F. G.: On Stochastic Matrices Associated with Certain Queueing Processes, *Ann. Math. Statist.*, vol. 24, pp. 355—360, 1953.
265. F o s t e r, F. G.: A Unified Theory for Stock, Storage and Queue Control, *Operational Research Quart.*, vol. 10, no. 3, September, 1959.
266. F o u l k e s, J. D., W. P r a g e r, and W. H. W a r n e r: On Bus Schedules, *Management Sci.*, vol. 1, pp. 41—48, 1954.
267. F r a n k, P.: Taboo Generating Functions and Other Topics in Markov Chains, *Columbia Univ. Statist. Engr. Group, Tech. Rept. 2(N)*, Feb. 26, 1959.
268. F r a z e r, R. A., W. J. D u n c a n, and A. R. C o l l a r: "Elementary Matrices", Cambridge University Press, London, 1955.
269. F r i e n d, J. K.: Two Studies in Airport Congestion, *Operational Research Quart.*, vol. 9, pp. 234—253, 1958.
270. F r o s t, G. R., W. K e i s t e r, and A. E. R i t c h i e: A Throwdown Machine for Telephone Traffic Studies, *Bell System Tech. J.*, vol. 32, pp. 292—359, 1953.
271. F r y, T. C.: The Theory of Probability as Applied to Problems of Congestion, in „Probability and Its Engineering Uses", D. Van Nostrand Company, Inc., Princeton, N. J., 1928.
Русский перевод: Ф р а й Т. Теория вероятностей для инженеров. Гостехиздат, М.-Л., 1934. Гл. X. Приложения теории вероятностей к проблемам скученности, с. 254—308.
272. F u l k e r s o n, D. R., and G. B. D a n t z i g: Computation of Maximum Flows in Networks, *Naval Research Logistics Quart.*, vol. 2, pp. 277—283, 1955.
273. G a l l i h e r, H. P., and R. C. W h e e l e r: Nonstationary Queueing Probabilities for Landing Congestion of Aircraft, *Operations Research*, vol. 6, no. 2, pp. 264—275, 1958.
274. G a n d a i s, M., and A. S a n n e r i s: "Principi di traffico telefonico", Stet-Torino, 1956.
275. G a n d e r, R. S.: Operational Research on Queueing Problems, *Research*, vol. 9, pp. 295—301, 1956.
276. G a n i, J.: Some Problems in the Theory of Provisioning and of Dams, *Biometrika*, vol. 42, pp. 179—200, 1955.
277. G a n i, J., and P. A. P. M o r a n: The Solution of Dam Equations by Monte Carlo Methods, *Australian J. Appl. Sci.*, vol. 6, pp. 267—273, 1955.
278. G a n i, J.: Problems in the Probability Theory of Storage Systems, *J. Roy. Statist. Soc., Ser. B*, vol. 19, no. 2, 1957.
279. G a n i, J. and N. V. P r a b h u: Stationary Distributions of the Negative Exponential Type for the Infinite Dam, *J. Roy. Statist. Soc., Ser. B*, vol. 19, pp. 342—351, 1957.
280. G a r b e r, N. H.: A Class of Queueing Problems, Ph. D. Thesis, Massachusetts Institute of Technology, Electrical Engineering Department, 1955.
281. G a r w o o d, F.: Application of the Theory of Probability to the Operation of Vehicular-controlled Traffic Signals, *J. Roy. Statist. Soc.* (suppl.), vol. 7, pp. 65—77, 1940—1941.

282. Gautzsch, O.: Verlustzeitung und Entlohnung bei mehr Maschinenbedienung, *Maschinenbau*, vol. 15, pp.179—182, 1936.

283. Gaver, D. P.: The Influence of Servicing Times in Queueing Processes, *J. Operations Research Soc. Am.*, vol. 2, pp. 139—149, 1954.

284. Gaver, D. P.: Imbedded Markov Chain Analysis of a Waiting Line Process in Continuous Time, *Ann. Math. Statist.*, vol. 30, no. 3, pp. 698—720, September, 1959.

285. Gaver, D. P. Jr.: A Waiting Line with Interrupted Service, Including Priorities, *Westinghouse Research Labs., Sci. Paper*, 6-41210-2-P4, Pittsburgh, Pa., August, 1960.

286. Gazis, D. G., R. Hermann and R. B. Potts: *Operations Research*, pp. 499—505, July—August, 1959.

287. Gerlough, D. L., and A. Schuhl: Poisson and Traffic. part I, The Use of the Poisson Distribution in Highway Traffic (D.L.G.); The Calculation of Probabilities and the Distribution of Traffic on Two Lane Roads. Eno Foundation for Highway Control, Saugatuck, Conn. (Schul's paper is a revised translation of *Travaux*, vol. 39, p. 24, 1955).

288. Gerlough, D. L., and J. H. Mathewson: Approaches to Operational Problems in Street and Highway Traffic — A Review, *Operations Research*, vol. 4, pp. 32—41, 1956.

289. Gerlough, D. L.: Simulation of Freeway Traffic by Electronic Computer, *Highway Research Board, Proc.*, vol. 35, pp. 543—547, 1956.

290. Gerlough, D. L., W. W. Mosher, and C. T. Weingarten: The Optimization of Traffic Flow through Certain Signalized Intersections, *Operations Research*, vol. 5, p. 135, 1957.

291. Gilbert, E. N.: N-terminal Switchinq Circuits, *Bell System Tech. J.*, vol. 30, p. 668, 1951.

292. Gilbert, E. N.: Enumeration of Labelled Graphs, *Can. J. Math.*, vol. 8, no. 3, pp. 405—411, 1956.

293. Gilbert, E. N., and H. O. Pollak: Coincidences in Poisson Patterns, *Bell System Tech J.*, vol. 35, no. 4, pp. 1005—1033, July. 1957.

294. Giltay, J.: Over Stationnaire en Overgangsstatisticken bij Telefoonverkeer, *Ingenieur (Utrecht)*, vol. 61, p. E45, 1949 (Dutch).

295. Giltay, J.: Static and Transient Statistics in Telephone-traffic Problems, *Appl. Sci. Research, Ser. B*, vol. B1, pp. 413—419, 1950.

296. Giltay, J.: Over Rangeringen bij de Automatische Telefonie, *Ingeniuer (Utrecht)*, vol. 65, pp. 107—118, 1953 (Dutch).

297. Giltay, J.: Bijdrage tot de Stochastiek van het Telefoonverkehr, in het bijzonder van de Volkomen Bundel, Dissertation, *University of Delft*, 1953 (Dutch; English summary).

298. Girault, M.: Quelques exemples d'analyse opèrationnelle des files d'attente et du stockage-mèthodes statistiques et calculs probabilistes, *Rev. franç recherche opérationnelle*, vol. 1, p. 106, 1957 (French).

299. Girault, M.: "Initiation aux processus aléatoires", Dunod, Paris, 1959.

300. Glanville, W. H.: Road Safety and Traffic Research in Great Britain, *J. Operations Research Soc. Am.*, vol. 3, pp. 283—299, 1955; also "Operations Research for Management," vol. 11, chap. 4, Johns Hopkins Press, Baltimore, 1956.

301. Glazier, E. D.: The Variation of the Holding Times of Telephone Calls, *P. O. Elec. Engrs. J.*, vol. 30, pp. 46—47, 1937.

302. Gluss, B.: Four Streams of Traffic Converging on a Cross-road, *Ann. Math. Statist.*, vol. 27, pp. 215—216, 1956.

303. Godfrey, A. I.: Stockyards, D. I. C. Thesis, Imperial College, London, 1953.

304. Gomm, G.: Wahrscheinlichkeitsprobleme im Fernsprechverkehr, *Mitteilungsb. Math. Statist.*, vol. 4, pp. 183—204, 1952 (German).

305. Good, I. J.: The Number of Individuals in a Cascade Process, *Proc. Cambridge Phil. Soc.*, vol. 45, pp. 360—363, 1948.

306. Goode, H. H., C. H. Pollmar, P. W. Warren, and J. B. Wright: Computer Simulation of Traffic, *Highway Research Board, Proc.*, vol. 35, pp. 548—557, 1956.

307. G o o d e, H. H., and R. E. M a c h o l: "Systems Engineering", chap. 23 Mc Graw-Hill Book Company, Inc., New York, 1957.
308. G r e e n b e r g, H., and A. D a o u: The Control of Traffic Flow to Increase the Flow, *Operations Research*, vol. 8, no. 4, p. 524, 1960.
309. G r e e n f i e l d, M. N.: Priority Effect in Queueing Theory, Methods of Operations Research, Course 41,616; American University, Washington, D. C., 1958.
310. G r e e n s h i e l d s, B. D., and F. M. W e i d a: "Statistics with Applications to Highway Traffic Analyses", Eno Foundation for Traffic Control, Saugatuck, Conn., 1947 (238 pages).
311. G r e e n s h i e l d s, B. D., D. S c h a p i r o, and E. L. E r i c k s e n: Traffic Performance at Urban Street Intersections. *Yale Univ., Bur. Highway Traffic Tech. Rept. 1*, 1947.
312. G r i n s t e d, W. H.: A Study of Telephone Traffic Problems with the Aid of the Principle of Probability, *P. O. Elec. Engrs. J.*, vol. 8, p. 33, 1915.
313. G r i n s t e d, W. H.: The Theory of Probability in Telephone Traffic Problems, *P. O. Elec. Engrs. J.*, vol. 11. pp. 148—152, 1918.
314. G r o s s, M.: Leistungvorrechnung bei mehr Maschinenbedienung, *Textile-Praxis*, vol. 4, p. 113, 1949.
315. G r o s s, M.: High-level Railroad Cooperation plus Operations Research Methods Equal More Efficient Railroading, *Ry. Age*, vol. 134, pp. 71—76, 1953.
316. G u m b e l, H.: Waiting Lines with Heterogeneous Servers, *Operations Research*, vol. 8, no. 4, p. 504, 1960.
317. G u r k, H. M.: Single Server, Time-limited Queue, *Bull Am. Math. Soc.*, vol. 63, p. 400, 1957 (abstract only).
318. H a h n, K.: Normalivkatalog-tiberlagerung, *Textile-u-Fasertofftech* vol. 5, pp. 641—643, 1955.
319. H a i g h t, F. A.: Queueing with Balking, *Biometrika*, vol. 44, pts. 3 and 4, December, 1957.
320. H a i g h t, F. A.: Queueing with Reneging, *Metrika*, vol. 2, no. 3, pp. 186—197, 1959.
321. H a i g h t, F. A.: Overflow at a Traffic Light, *Biometrika*, vol. 46, pts. 3 and 4, December, 1959.
322. H a i g h t, F. A.: The Volume, Density Relation in the Theory of Road Traffic, *Operations Research*, vol. 8, no. 4, p. 572, 1960.
323. H a i g h t, F. A.: Queueing with Balking, part II, *Biometrika*, vol. 47, nos. 3 and 4, pp. 285—296, 1960.
324. H a l l e r, R. and B r o w n, Inc.: Queueing Theory Applied to Military Communications Systems, State College, Pa., 1956.
325. H a m m e r s l e y, J. M.: On Counters with Random Dead Time, *Proc. Cambridge Phil. Soc.*, vol. 49, pp. 623—637, 1953.
326. H a m m e r s l e y, J. M.: Storage Problems, *Math. Ann.*, vol. 128, pp.475—478, 1955.
327. H a r d y, G. H.: "Orders of Infinity, The 'Infinitärcalcül' of Paul du-Bois-Reymond", Cambridge University Press, London, 1954.
328. H a r r i s, L. R. F.: Time Sharing as a Basis for Electronic Telephone Switching, *Proc. Inst. Elec. Engrs. (London)*, Paper 1993R, vol. 103, pt. B, 1956.
329. H a r r i s, T. E.: Branching Processes, *Ann. Math. Statist.*, vol. 19, p. 474, 1948.
330. H a r r i s, T. E.: Some Mathematical Models for Branching Processes, *Proc. Second Berkeley Symposium on Math. Statist. and Probability*, 1951.
331. H a r r i s, T. E.: Stationary Single-server Queueing Processes with a Finite Number of Sources, *Operations Research*, vol. 7, no. 4, pp. 458—467, July, 1959.
332. H a r r i s o n, H. H.: "An Introduction to the Strowger System of Automatic Telephony", pp. 29—39 (remarks on traffic), Longmans, Green & Co., Ltd., London, 1924.
333. H a r r i s o n, S.: Queueing Theory, Booz Allen Applied Research, Bethesda, Md., paper submitted to *Scientific American*, 1960.
334. H a r t l e y, H. O. and E. R. F i t c h: A Chart for the Complete Beta Function and the Cumulative Binomial Distribution, *Biometrika*, vol. 38, 1951.

335. Haselton, M. L., and E. L. Schmidt: Automatic Inventory System for Air Travel Reservations, *Elec. Eng.*, pp. 641—646, July, 1954.

336. Hawkins, N. A.: Problems in Automatic Trunking-Last Contact Traffic, *P. O. Elec. Engrs. J.*, vol. 23, pp. 272—281, 1930—1931.

337. Hawkins, N. A.: Further Problems in Automatic Trunking—Some Interesting Solutions by Means of the Calculus, *P. O. Elec. Engrs. J.*, vol. 25, pp. 289—294, 1932.

338. Hayward. W. S.: The Reliability of Telephone Traffic Load Measurements by Switch Counts, *Bill System Tech. J.*, vol. 31, pp. 357—377, 1952.

339. Healy, T. L.: Queues with "Exponential-type" Service-time Distributions, *Operations Research*, vol. 8, no. 5, pp. 719—721, 1960.

340. Heathcote C. R.: The Time Dependent Problem for a Queue with Preemptive Priorities, *Operations Research*, vol. 7, no. 5, September-October, 1959.

341. Heathcote, C. R., and J. E. Moyal: The Random Walk (in Continuous Time) and Its Application to the Theory of Queues, *Biometrika*, vol. 46, pts. 3 and 4, December, 1959.

342. Heathcote, C. R.: A Single Queue with Several Preemptive Priority Classes, *Operations Research*, vol. 8, no. 5, pp. 630—638, 1960.

343. Hènon, R.: Combien de machines peut-on confier à un seul agent? *Rev. statist. appl*, vol. 3, pp. 73—81, 1955 (French).

344. Herman, R., E. W. Montroll, R. B. Potts, and R. W. Rothery: Traffic Dynamics: Analysis of Stability in Car Following, *Operations Research*, pp. 86—106, January-February, 1959.

345. Herne, H. J. L., and D. G. Nickols: *Proc. Radioactivity Conference*, Oxford, H. M. Stationery Office, London, 1951.

346. Herniter, J. D., and J. F. Magee: Customer Behavior as a Markov Process, *Operations Research*, vol. 9, no. 1, p. 105, 1961.

347. Herrey, E. M. J., and H. Herrey: Principles of Physics Applied to Traffic Movement and Road Conditions, *Am. J. Phys.*, vol. 13, pp. 1—14, 1945.

348. Hettwig, V. E., and K. Rohde: New Calculation Methods for the Design of Telephone Systems, *Siemens-Z.*, vol. 30, no. 1, pp. 1—10, January, 1956 (German).

349. Heyveart, A. C., and A. Hurt: Inventory Management of Slow Moving Parts, *Operations Research*, vol. 4, pp. 572—580, 1956.

350. Hildebrand, F. B.: "Introduction to Numerical Analysis", McGraw-Hill Book Company, Inc., New York, 1956.

351. Hille, E.: Functional Analysis and Semi-groups, *Am. Math. Soc. Colloq.*, Publ., vol. 31, 1948.

352. Hille, E.: On the Integration of Kolmogoroff's Differential Equations, *Proc. Natl. Acad. Sci. U. S.*, vol. 40, no. 1, pp. 20—25, January, 1954.

353. Hines, J. G.: The Anticipation of Demand and Economic Selection, Provision and Layont of Plant (Telephone Systems), *J. Inst. Elec. Engrs (London)*, vol. 67, pp. 594—618, 1929.

354. Holley, J. L.: Waiting Line Subject to Priorities, *J. Operations Research Soc. Am.*, vol. 2, pp. 341—343, 1954.

355. Holm, R.: Über die Benutzung der Wahrscheinlichkeitstheorie für Telefonverkehrsprobleme, *Arch. Elektrotech.*, vol. 8, p. 413, 1920.

356. Holm, R.: Calculation of Blocking Factors of Automatic Exchanges, *P. O. Elec. Engrs. J.*, vol. 15, pp. 22—38, 1922.

357. Holm, R.: The Validity of Erlang's Trunk Congestion Formula, *P. O. Elec. Engrs. J.*, vol. 17, pp. 318—321, 1925.

358. Homma, T.: On a Certain Queueing Process, *Rept. Statist. Application Research, Union Japanese Scientists and Engrs.*, vol. 4, no. 1, 1955.

359. Homma, T.: On the Many Server Queueing Process with a Particular Type of Queue Discipline, *Rept. Statist. Application Research, Union Japanese Scientists and Engrs.*, vol. 4, pp. 20—31, 1956.

360. Homma, T.: On the Theory of Queues with Some Types of Queue Discipline, *Yokohama Math. J.*, vol. 3, 1956.

361. Hunt, G. C.: Sequential Arrays of Waiting Lines, *Operations Research*, vol. 4, pp. 674—683, December, 1956.

362. Hurst, H. E.: Long-term Storage Capacity of Reservoirs, *Trans. Am. Soc. Civil. Engrs.*, vol. 116, p. 170, 1951.
363. Inglis, B. G.: Traffic Signal Systems, *J. Inst. Munic. Engrs.*, vol. 80, pp. 369—392, 1954.
364. Inose H.: An Artificial Call Generator Employing Random Noise Resistors, *J. Inst. Elec. Commun. Engrs. (Japan)*, vol. 36, pp. 177—181, 1951 (Japanese).
365. Irwin, J. O.: The Frequency Distribution of the Difference between Two Independent Variates Following the Same Poisson Distribution, *J. Roy. Statist. Soc.*, vol. 100, p. 415, 1937.
366. Jackson, J. R.: Networks of Waiting Lines, *Operations Research*, vol. 5, no. 4, August, 1957.
367. Jackson, J. R.: A Note on Transient Phenomena in Simple Queues, *Univ. Calif., Management Sci. Research Project Discussion Paper* 74, Nov. 16, 1959,
368. Jackson, J. R.: Some Problems in Queueing with Dynamic Priorities, *Univ. Calif., Management Sci. Research Project Paper* 62, Nov. 30, 1959.
369. Jackson, R. R. P.: Queueing Systems with Phase Type Service, *Operations Research Quart.*, vol. 5, pp. 109—120, 1954.
370. Jackson, R. R. P.: Queueing Processes with Phase Type Service, *J. Roy. Statist. Soc., Ser. B*, vol. 18, no. 1, pp. 129—132, 1956.
371. Jackson, R. R. P., and D. G. Nickols: Some Equilibrium Results for the Queueing Process *E/M/*1, *J. Roy. Statist. Soc., Ser. B*, vol. 18, pp. 275—279, 1956.
372. Jackson, R. R. P.: Notes on Queueing Processes and Stochastic Stock Models, London School of Economics, *Operations Research Seminar Paper*, June 27, 1957.
373. Jackobs, W. W.: The Caterer Problem, *Naval Research Logistics Quart.*, vol. 1, pp. 154—165, 1954.
374. Jacobaeus, C.: The Employment of Cross-bar Selectors in Bypass Circuit Systems, *Ericsson Rev.*, no. 2, p. 30, 1945.
375. Jacobaeus, C.: Influence of the Size of Selectors on the Number of Selectors in Telephone Plants, *Ericsson Rev.*, vol. 24, pp. 2—9, 1947.
376. Jacobaeus, C.: Blocking Computations in Link Systems, *Ericsson Rev.*, vol. 24, p. 86, 1947.
377. Jacobaeus, C.: A Study on Congestion in Link Systems, part I, Theoretical Analysis and Results; part II, Experimental Methods and Results. *Ericsson Tech.*, vol. 8, pp. 1—68, 1950.
378. Jacobaeus, C: Theoretical Considerations in the Design of Crossbar Exchanges, *Teleteknik*, vol. 2, p. 105, 1951.
379. Jacobaeus, C.: Investigations with an Electronic Telephone Traffic Machine, *Teleteknik*, vol. 5, no. 2, pp. 260—262, July, 1954 (Danish).
380. Jacobaeus, C.: Measurements with Lamp Panel — A New Method for Traffic Studies (International Teletraffic Congress, Copenhagen, 1955), *Teleteknik*, vol. 1, p. 91, 1957.
381. Jacobsen, C.: Investigations with an Electronic Telephone Traffic Machine, *Teleteknik*, vol. 5, pp. 260—262, 1954 (Danish).
382. Jaiswal, N. K.: Bulk-service Queueing Problem, *Operations Research*, vol. 8, no. 1, pp. 139—143, January—February, 1960.
383. Jaiswal, N. K.: Time Dependent Solution of the Bulk-Service Queueing Problem, *Operations Research*, vol. 8, no. 6, p. 773, 1960.
384. Jaiswal, N. K., and H. C. Jain: A Priority Problem in Queueing Theory, *Defence Sci. J. India* (to be published in *Operations Research*, 1961).
385. Jenkins, J. L.: A Survey of Some Application of Queueing, May 13, 1958 (mimeo. copy).
386. Jensen, A.: An Elucidation of A. K. Erlang's Statistical Works through the Theory of Stochastic Processes, in E. Brockmeyer et al., "The Life and Works of A. K. Erlang," pp. 23—100, Copenhagen Telephone Company, Copenhagen, 1948.
387. Jensen, A.: „Moe's Principle". Copenhagen Telephone Company, Copenhagen, 1950.

388. Jensen, A.: Distribution Patterns Composed of a Limited Number of Exponential Distributions, Second Scandinavian Mathematics Congress, Trondheim, 1949, pp. 210—215, Johan Grundt Tanums Forlag, Oslo, 1952.

389. Jensen, A.: A Basis for the Calculation of Congestion in Crossbar Systems, *Teleteknik*, vol. 3, p. 123, 1952.

390. Jensen, A.: Caclulation of Loss in Crossbar Automatic Exchanges, *Teleteknik*, vol. 4, pp. 176—200, 1952 (Danish).

391. Jensen, A.: Markoff Chains as an Aid in the Study of Markoff Processes, *Skand. Aktuar. Tidskr.*, vol. 36, pp. 87—91, 1953.

392. Jensen, A.: "A Distribution Model", Munksgaard, Copenhagen, 1954.

393. Jensen, A.: The Applicability of Decision Theory in the Planning and Operation of Telephone Plant (International Teletraffic Congress, Copenhagen, 1955), *Teleteknik*, vol. 1, pp. 126—129, 1957.

394. Jensen, E. L.: Elementary Queueing Theory, Elementaer Koteori, *Nord. Mat. Tidskr.*, 1958 (Danish).

395. Joffe, A., and P. E. Ney: A Convergence Problem in the Theory of Multiserver Queues, Cornell University, Department of Mathematics, Ithaca, N. Y., 1960.

396. Johannsen, F. W.: Waiting Times and Number of Calls, *P. O. Elec. Engrs. J.*, 1907; reprinted October, 1910, and January, 1911.

397. Johannsen, F. W.: Busy, *Ingeniorvidenskab. Skrifter.*, 1908; Revised, *ibid.*, Ser. A, no. 32, 1932, as Appendix I of the Development of Telephonic Communication in Copenhagen, 1881—1931. Translated by A. S. Jarvis from original Danish written in 1908, pp. 150—153.

398. Johnson, M. H.: Theoretical Studies into Traffic Congestion, part IV, *British Ministry of Civil Aviation, Operations Research Sect., ORS/MCA Rept.* 9, May, 1949.

399. Jones, W. D.: A Simple Way to Figure Machine Down Time, *Factory Management and Maintenance*, vol. 104, pp. 118—121, 1946.

400. Jones, W. D.: Machine Interference, *J. Ind. Eng.*, vol. 1, pp. 10, 11, 19, 1949.

401. Jones, W. D.: Determining Inefficiencies in Multiple-machine Assignments, *Georgia Institute of Technology, Experiment Station Research in Engineering*, 1953.

402. Jones, W. D.: Graphical Determination of Work Loads for Multiple Machine Assignments, *J. Ind. Eng.*, vol. 4, pp. 16, 19, 26, 1953.

403. Josephs, H. J., and R. A. Hastie: Operational Research in the Post Office (Part 2, Probability Models), *P. O. Elec. Engrs. J.*, vol. 50, pp. 81—85, July, 1957.

404. Jowett, G. H.: The Exponential Distribution and Its Applications, *Incorp. Statistician*, vol. 8, no. 2, pp. 89—96, January, 1958.

405. K. T. A. S.: The Development of Telephone Communication in Copenhagen, 1881—1931, *Ingeniorvidenskab. Skrifter, Ser. A*, no. 32, 1932.

406. Kac, M.: Random Walk in the Presence of Absorbing Barriers, *Ann. Math. Statist.*, vol. 16, 1945.

407. Kac, M.: Random Walk and the Theory of Browjian Motion, *Am. Math. Monthly*, vol. 54, pp. 369—391, 1947.

408. Karlin, S., and J. McGregor: Representation of a Class of Stochastic Processes, *Proc. Natl. Acad. Sci., U. S.*, vol. 41, no. 6. pp. 387—391, June, 1955.

409. Karlin, S., and. J. McGregor: Many Server Queueing Processes with Poisson Input and Exponential Service Times, *Pacific J. Math.*, vol. 8 no. 1, pp. 87—118, 1958.

410. Karlin, S., and J. McGregor: The Classification of Birth and Death Processes, *Trans. Am. Math. Soc.*, vol. 86, no. 2, pp. 366—400, 1957.

411. Karlin, S., and J. McGregor: The Differential Equations of Birth and Death Processes and the Stiltjes Moment Problem, *Trans. Am. Math. Soc.*, vol. 85, no. 2, pp. 489—546, July, 1957.

412. Karlin, S., and J. McGregor: Linear Growth Birth and Death Process, *J. Math. Mech.*, vol. 7, pp. 643—662, 1958.

413. Karlin, S., R. C. Miller, and N. V. Prabhu: Note on a Moving Single Server Problem, *Ann. Math. Statist.*, vol. 30, pp. 243—246, 1959.
414. Karlsson, S. A.: Konstgjord Telefontrafik som Hjalpmedel vid Behandling av Telefontrafik Problem, Helsingfors, 1945.
415. Karlsson, S. A.: Forbattrad Narmeformel fur Erlang's Ideela Gradering, *Kraft och Ljus* (Helsingfors), July—August, 1949.
416. Karlsson, S. A.: Ein Analysator für den Fernsprechverkehr, *Teleteknik* (English ed.), vol. 1. pp. 113—120, 1957 (German).
417. Karoly, S.: Mozdonyok Varakozasi Idejerol, *Magyar Tudomànyas Akad.*, vol. 3, 1954.
418. Karush, W.: A Queueing Model for an Inventory Problem, *Operations Research*, vol. 5, no. 5, 1957.
*419. Kaufmann, A.: „Méthodes et modèles de la recherche opérationnelle", Paris, Dunod, vol. 1, 1962; vol. 2, 1964.
420. Kawata, T.: A Problem in the Theory of Queues, *Repts. Statist. Application Research, Union Japanese Scientists and Engrs.*, vol. 3, pp. 122—129, 1955.
421. Keilson, J., and A. Kooharian: On the General Time Dependent Queue with a Single Server, *Sylvania Electronics Systems*, Applied Research Memo. 209, Waltham, Mass., Apr. 13, 1960.
422. Keilson, J.: On Time Dependent Queueing Processes, *Ann. Math. Statist.*, vol. 31, no. 1, March, 1960.
423. Keinonen, A. A.: Studien und Messungen unbegrenzter Wartezeiten mit Hilfe einer Speicherkunstschaltung, *Veröffentl. Geb. Nachr. Tech.*, vol. 8, pp. 665—680, 1938.
424. Keister, W., Ritchie, and S. Washburn: "The Design of Switching Circuits", D. Van Nostrand Company, Inc., Princeton, N. J., 1951.
425. Kellerer, H.: „Verkehrsstatistik", O. Elsner, Berlin, 1936 (German).
426. Kendall, D. G.: On Some Modes of Population Growth Leading to R. A. Fisher's Logarithmic Series, *Biometrika*, vol. 35, p. 6—15, 1948.
427. Kendall, D. G.: On the Role of Variable Generation Time in the Development of a Stochastic Birth Process, *Biometrika*, vol. 35, pp. 316—330, 1948.
428. Kendall, D. G.: On the Generalized "Birth-and-Death" Process, *Ann. Math. Statist.*, vol., 19, p. 1. 1948.
429. Kendall, D. G.: Stochastic Processes and Population Growth, *J. Roy. Statist. Soc., Ser. B*, no. 2, 1949.
430. Kendall, D. G.: Random Fluctuations in the Age-distribution of a Population Whose Development Is Controlled by the Simple Birth-and-Death Process, *J. Roy. Statist. Soc., Ser. B*, vol. 12, p. 278, 1950.
431. Kendall, D. G.: On Non-dissipative Markoff Chains with an Enumerable Infinity of States, *Proc. Cambridge Phil. Soc.*, vol. 47, p. 633, 1951.
432. Kendall, D. G.: Some Problems in the Theory of Queues, *J. Roy. Statist. Soc., Ser. B*, vol. 13, no. 2, pp. 151—185, 1951.
433. Kendall, D. G.: Stochastic Processes Occuring in the Theory of Queues and Their Analysis by the Method of the Imbedded Markov Chain, *Ann. Math. Statist.*, vol. 24, pp. 338—354, 1953.
Русский перевод: Кендалл Д. Стохастические процессы, встречающиеся в теории очередей, и их анализ методом вложенных цепей Маркова. Сб. переводов „Математика", 3:6, 1959, с. 97—111.
434. Kendall, D. G., and G. E. H. Reuter: Some Pathological Markov Processes with a Denumerable Infinity of States and the Associated Semi-groups of Operations on *l*, *Proc. Intern. Congr. Math.*, Amsterdam, 1954.
435. Kendall, D. G.: Some Further Pathological Examples in the Theory of Denumerable Markov Processes, *Quart. J. Math.*, Oxford Series, vol. 7, pp. 39—56, 1956.
436. Kendall, D. G.: Some Problems in the Theory of Dams, *J. Roy. Statist. Soc., Ser. B*, vol. 19, no. 2, 1957.
437. Kendall, D. G., and G. E. H. Reuter: The Calculation of the Ergodic Properties for Markov Chains and Processes with a Countable Infinity of States, *Acta Math.*, vol. 97, pp. 103—144, 1957.

438. K e n d a l l, D. G.: Geometric Ergodicity and the Theory of Queues, in K. J. Arrow, S. Karlin, and P. Suppes (eds.), „Mathematical Methods in the Social Sciences", chap. 12, Stanford University Press, Stanford, Calif., 1959.
439. K e n d a l l, M. G.: „The Advanced Theory of Statistics", vols. 1 and 2, Charles Griffin & Co., Ltd., London, 1948.
440. K e s t e n, H., and J. Th. R u n n e n b u r g: Some Elementary Proofs in Renewal Theory with Applications to Waiting Times, *Math. Centrum, Statist. Afdel. Rept.* S203, Amsterdam, 1956.
441. K e s t e n, H., and J. Th. R u n n e n b u r g: Priority in Waiting-line Problems, *Koninkl. Ned. Akad. Wetenschap., Proc., ser. A*, vol. 60, pt. I, pp. 312—324, and pt. II. pp. 325—336, 1957.
442. K i e f e r, J., and J. W o l f o w i t z: On the Theory of Queues with Many Servers, *Trans. Am. Math. Soc.*, vol. 78, pp. 1—18, 1955.
443. K i e f e r, J. and J. W o l f o w i t z: On the Characteristics of the General Queueing Process with Application to Random Walks, *Ann. Math. Statist.*, vol. 27, no. 1, pp. 147—161, 1956.
444. K i n z e r, J. P.: Application of the Theory of Probability to Problems in Highway Traffic, B. C. E. Thesis, Polytechnic Institute of Brooklyn, 1933.
445. K l e i n m a n n, H. A.: Les Phénomènes d'attente, *Univ. Paris, Inst. statist. bull. du séminaire de recherche opèrationnelle*, June, 1955.
446. K l e i n m a n n. H. A.: Simultaneous Determination of the Operating Time of 450 Machines, *Rev. statist. appl.*, vol. 3, pp. 11—22, 1955 (French).
447. K n i g h t, N. V.: Depreciation and Service Life of Telecommunication Plant, *P. O. Elec. Engrs. J.*, Paper 209, February, 1955.
448. K n i g h t, N. V.: Economic Principles of Telecommunications Plant Provision, *P. O. Elec. Engrs. J.*, Paper 210, November, 1955.
449. K n o d e l, W.: Mehrplatzarbeit, *Unternehmensforschung*, vol. 1, pp. 15—17, 1956 (German).
450. K n o w l t o n, S. L.: A Simplified Approach to Waiting Lines, *J. Ind. Enq.*, vol. 10, no. 6, pp. 423—425, November—December, 1959.
451. K o e n i g s b e r g, E.: Birth, Death and Waiting in Line, *Am. Math. Monthly*, vol. 62, p. 543, 1952.
452. K o e n i g s b e r g, E.: Queueing with Special Service, *Operations Research*, vol. 4, pp. 213—220, 1956.
453. K o e n i g s b e r g, E.: Cyclic Queues, *Operational Research Quart.*, vol. 9, no. 1, 1958.
454. K o e n i g s b e r g, E.: Finite Queues and Cyclic Queues, *Operations Research*, vol. 8, no. 2, p. 246, March—April, 1960.
455. K o m e t a n i, E.: On the Theoretical Solution of Highway Traffic Capacity under Mixed Traffic, *Mem. Fac. Eng., Kyoto Univ.*, vol. 17, pp. 79—88, 1955.
456. K o m e t a n i, E.: An Abridged Table for Infinite Queues, *Operations Research*, vol. 7, no. 3, pp. 385—393, May—June, 1959.
457. K o m e t a n i, E., and T. S a s a k i: A Safety Index for Traffic with Linear Spacing, *Operations Research*, vol. 7, no. 6, November—December, 1959.
458. K o o p, H.: Planung von Kabelkanalen in Großstadten, *Fernmelde—Praxis*, vol. 32, pp. 769—778, 1955.
459. K o o p m a n, B. O.: New Mathematical Methods in Operations Research, *J. Operations Research Soc. Am.*, vol. 1, pp. 3—9, 1952.
460. K o r n, F. A., and J. G. F e r g u s o n; Number 5 Grossbar Dial Telephone Switching System, *Trans. Am. Inst. Elec. Engrs.*, vol. 69, pp. 244—254 1950.
461. K o r t e, J. W., and W. L e u t z b a c h: Time-gap Distribution in Case of Interrupted Traffic Flow, *Verkehr Technik*, vol. 9, pp. 17—18, 42—43, 1956 (German).
462. K o s t e n, L.: Sur les problèmes de blocage dans les multiples graduées, *Ann. P. T. T.*, vol. 26, p. 1002, 1937.
463. K o s t e n, L.: Über Sperrungswahrscheinlichkeiten bei Staffelschaltungen, *Elek. Nachr.-Tech.*, vol. 14, pp. 5—12, 1937.

464. K o s t e n, L.: Over Blokkeerings-en Wachtproblemen, Dissertation, University of Delft, 1942 (Dutch, French and German summaries).

465. K o s t e n, L.: On the Frequency Distribution of the Number of Discharges Counted by a Geiger-Müller Counter in a Constant Interval, *Physica*, vol. 10, pp. 749—756, 1943.

466. K o s t e n, L.: On the Influence of Repeated Calls in the Theory of Probabilities of Blocking, *Ingenieur (Utrecht)*, vol. 59, pp. 1—25, 1947 (Dutch).

467. K o s t e n, L.: On the Measurement of Congestion Quantities by Means of Fictitious Traffic, *Het P. T. T. Bedriff*, vol. 2, pp. 15—25, 1948.

468. K o s t e n, L.: On the Validity of the Erlang and Engset Loss Formulae, *Het P. T. T. Bedriff*, vol. 2, pp. 42—45, 1948.

469. K o s t e n, L., J. R. M a n n i n g, and F. G a r w o o d: On the Accuracy of Measurements of Probabilities of Loss in Telephone Systems, *J. Roy. Statist. Soc., Ser. B*, vol. 11, pp. 54—67, 1949.

470. K o s t e n, L.: Beschrijving van een Machine voor Kunstmatig Verkeer op Electronische Basis, *Ingenieur (Utrecht)*, vol. 63, p. 62, 1951.

471. K o s t e n, L.: Waarschijnlijkheidsrekening in de Telecommunicatietechniek, *Voord. Inst. Ing. Utrecht*, vol. 11, 1951.

472. K o s t e n, L.: On the Accuracy of Measurements of Probabilities of Delay in Telecommunication Systems, Part I: Estimates of Probabilities of Delay, *Appl. Sci Research, Ser. B*, vol. 2, pp. 108—130, 1952.

473. K o s t e n, L.: On the Accuracy of Measurements of Probabilities of Delay in Telecommunication Systems, Part II: Estimates of Average Times of Delay, *Appl. Sci. Research, Ser. B*, vol. 2, pp. 401—415, 1952.

474. K o s t e n, L.: Tafel van de Stagnatiekans van de volledige Bundel met Wachtgelgenheid, *Centr. Lab. P. T. T. (Neth.)*, no. 69, 1954.

475. K o s t e n, L.: The Historical Development of the Theory of Probability in Telephone Traffic Engineering in Europe, *Teleteknik* (English ed.), vol. 1, pp. 32—40, 1957.

476. K o s t e n L.: Application of Artificial Traffic Methods to Telephone Problems, *Teleteknik* (English ed.), vol. 1, pp. 107—110, 1957.

477. K r e m e r, H.: Das Verteilungsgesetz der Belegunslagen in der Wahltechnik, *Arch. Elektrotech.*, vol. 6, pp. 195—198, 1952 (German).

478. K r e m e r, H.: The Statistics of the „Fully Available Group" as Applied to Telephone Network Theory, *Arch. Electrotech.*, vol. 6, pp. 469—472, 1952.

479. K r e p e l i e n, H. Y.: The Influence of Telephone Rates on Local Traffic, *Ericsson Tech.*, no. 2, 1958.

480. K r o e s, J. L. de: Calculation of the Number of Direct Junction Lines in Telephony, *Commun. News*, vol. 12, p. 132, 1952.

481. K r o e s, J. L. de: Optimum Crouping in Uniselector Trunking, *Commun. News*, vol. 14, pp. 70—76, January, 1954.

482. K r o n i g, R.: On Time Losses in Machinery Undergoing Interruptions, part I, *Physics*, vol. 10, pp. 215—224, 1943.

483. K r o n i g, R. and H. M o n d r i a: On Time Losses in Machinery Undergoing Interruptions, part II, *Physics*, vol. 10, pp. 331—336, 1943.

484. K r u i t h o f, J.: Telefoonsverkeersrekening, *Ingenieur (Utrecht)*, vol. 52, pp. 15—25, 1937 (Dutch).

485. K r u i t h o f, J.: Rotary Traffic Machine, *Elec. Commun.*, vol. 23, p. 192, 1946.

486. K u h n, S.: „Traffic Problems in Automatic Telephony", p. 384, P. W. N., Warszawa, Poland, 1957 (Polish).

487. K u r b a t o v, J. D., and H. B. M a n n: Correction of Geiger—Müller Counter Data, *Phys. Rev.*, vol. 68. pp. 40—43, 1945.

488. L a m b e r t, F.: Les Problèmes d'attente, *Cahiers centre d'ètudes recherche opérationnelle*, no. 2, pp. 5—28, 1959.

489. L a n g e r, M.: Calculation of Switches Required in Automatic Exchanges, *Elektrotech. Z.*, no. 1, 1924.

490. L a n i n g, J. H., Jr., and R. H. B a t t i n: "Random Processes in Automatic Control", McGraw-Hill Book Company, Inc., New York, 1956.
491. L a v a l e e, R. S.: The Scheduling of Traffic Signals by Linear Programming, *Highway Research Board, Proc.*, vol. 35, pp. 534—542, 1956.
492. Lectures in Queueing Theory given by Univac Engineering Division, Remington Rand, PX 1301, St. Paul, Minn., May, 1959.
493. L e d e r m a n n, W., and G. E. R e u t e r: Spectral Theory for the Differential Equations of Simple Birth and Death Processes, *Phil. Trans. Roy. Soc. London, Ser. A*, vol. 246, pp. 321—369, 1954.
494. L e e, A. M.: Some Aspects of a Control and Communication System, *Operational Research Quart.*, vol. 10, no. 4, pp. 206—216, December, 1959.
495. L e e, A. M., and P. A. L o n g t o n: Queueing Processes Associated with Airline Passenger Check-in, *Operational Research Quart.*, vol. 10, no. 1, pp. 56—71, March, 1959.
496. L e e, C. Y.: Analysis of Switching Networks, *Bell System Tech. J.*, vol. 34, pp. 1287—1315, November, 1955.
497. Le G a l l, R.: Etude du blocage dans les systèmes de commutation téléphonique automatique utilisant des commutateurs électroniques du type crossbar, *Ann. télécommun.*, vol. 11, pp. 159—171. 1956; pp. 180—194, 1956; p. 197 (Erratum).
498. Le G a l l, P.: Méthode de calcul de l'encombrement dans les systèmes téléphoniques automatiques á marquage, *Ann. télécommun.*, vol. 12, pp. 374—386, 1957.
499. Le G u i l l a n t et al.: La Neurose des téléphonistes, *Presse méd.*, vol. 64, p. 274, 1956.
500. L e i b o w i t z, M. A.: An Approximate Method for Treating a Class of Multiqueue Problems, *IBM J. Research Develop.*, vol. 5, no. 3, pp. 204—209, 1961.
501. L e l y, U. P.: Waarschijnlijkheidsrekening bij Automatische Telephonie, The Hague, 1918.
502. L e l y, U. P.: On Ekelöf's Calculation of Delays in an Automatic Telephone System, *Ericsson Rev.*, pp. 7—12, 1930.
503. L e r o y, R., and A. E. V a u l o t: Sur la proportion d'appels perdus dans certains systèmes de téléphonie automatique ne permettant dans un groupe d'organes qu'une seule exploration simultanée, *Compt. rend.*, vol. 220, pp. 84—85, 1945.
504. Le R o y, J.: Recherches en deux ordres inverses dans la sélection en téléphonie automatique, *Ann. P. T. T.*, vol. 5, pp. 366—374, 1950 (French).
505. Le R o y, J.: Formules matricielles du calcul du délai d'attente dans le cas des appels desseruis au hasard, *Ann. télécommun.*, vol. 12, 1957.
506. L e s o u r n e, J.: „Technique économique et gestion industrielle", Dunod, Paris, 1958.
507. L e u n b a c h, G.: Illustration of the Application of Statistical Decision Functions in a Telephone Plant (International Teletraffic Congress, Copenhagen, 1955), *Nord. Tidsk. Tek. Okonomi*, p. 146, 1955.
508. L e v e n s o n, B. D., and R. B. T a s k e r: High Density Military Traffic Control System, *Air Navigation Lab. Progress Repts.* 387 (1—6); *Final Eng. Rept.* 6, July, 1950—July, 1951.
509. L e v e r t, C., and W. L. S c h e e n: Probability Fluctuations of Discharges in a Geiger—Müller Counter Produced by Cosmic Radiation, *Physica*, vol. 10, pp. 225—238, 1943.
510. L e v i n s o n, M. S.: The Sluggishness or Queueing Factor, *Traffic Eng.*, vol. 27, pp. 401—402, 1956.
511. L e v y, P.: Sur la division d'un segment par des pointes choises au hasard, *Compt. rend.*, vol. 208, pp. 147—149, 1939.
512. L e v y, P.: Convergence des séries aléatoires et loi normale, *Compt. rend.*, vol. 234, pp. 72 422—72 424, 1952.
513. L e w i s, N. H.: Notes on the Exponential Distribution in Statistics, *P. O. Elec. Engrs. J.*, vol. 41, pp. 10—12, 1948.
514. L e w i s, R., F. N e e l a n d, and M. G o u r a r y: An Inventory Control Bibliography, *Naval Research Logistics Quart.*, vol. 3, pp. 295—303, 1956.

515. Lighthill, M. J., and G. B. Witham: On Kinematic Waves — A Theory of Traffic Flow on Long Crowded Roads, *Proc. Roy. Soc. (London), Ser. A*, vol. 229, pp. 281—316; Part II, pp. 317—345, 1955.

516. Lind, G.: Statistical Supervision of Telephone Plant, *Ericsson Tech.*, no. 2, 1958.

517. Lindelöf, E.: „Calcul des résidus," Chelsea Publishing Company, New York, 1947.

518. Lindley, D. V.: Mathematical Theory — Marshalling & Queueing, *Operational Research Quart.*, vol. 3, no. 1; 1952.

519. Lindley, D. V.: The Theory of Queues with a Single Server, *Proc. Cambridge Phil. Soc.*, vol. 48, pp. 277—289, 1952.

520. Little, J. D. C.: The Use of Storage Water in a Hydroelectric System, *J. Operations Research Soc. Am.*, vol. 3. pp. 187—197, 1955.

521. Little, J. D. C.: Approximate Expected Delays for Several Maneuvers by a Driver in Poisson Traffic, *Operations Research*, vol. 9, no. 1, p. 39, 1961.

522. Loeve, M.: „Probability Theory", D. Van Nostrand Company, Inc., Princeton, N. J., 1955.

Русский перевод: Лоэв, М. Теория вероятностей. Изд-во иностранной литературы, 1962.

523. Longley, H. A.: The Efficiency of Gradings, *P. O. Elec. Engrs. J.*, vol. 41, pp. 45—67, 1948—1949.

524. Lotka, A. J.: A Contribution to the Theory of Self-renewing Aggregates with Special Reference to Industrial Replacement, *Ann. Math. Statist.*, vol. 10, pp. 1—25, 1939.

525. Lotka, A. J.: Application of Recurrent Series in Renewal Theory, *Ann. Math. Statist.*, vol. 19, pp. 190—206, 1948.

526. Lotze, A.: Theorie der Gefahrzeit volkommener Bündel in Fernsprechwahlsysteme, Dissertation, Technische. Hochschule, Stuttgart, 1953 (German).

527. Lotze, A.: Gefahrzeit oder Verlust als Maß für die Betriebsgüte im Fernspechverkehr, *FTZ — Fernmeldetech. Z.*, vol. 6. pp. 564—570, 1953.

528. Lotze, A.: Berechnung der Verkehrsgrößen im Wartezeitsystem aus den Verkehrsgrößen eines Verlustsystems, *FTZ—Fernmeldetech. Z.*, vol. 7, pp. 443—453, 1954.

529. Lotze, A.: Erwiderung zur Stellungnahme von J. W. Cohen, *NTZ — Nachrichtentech Z.*, vol. 8, pp. 139—140, 1955.

530. Lotze, A., and E. Schwiderski: Wartezeitprobleme im Fernsprechverkehr, *NTZ — Nachrichtentech. Z.*, vol. 8, pp. 646—649, 1955.

531. Lubberger, F.: Beobachtungen, Vorschriften und Theorien der Schwankungen im Fernsprechverkehr, in F. Lubberger (ed.), „Wahrscheinlichkeiten und Schwangungen", pp. 41—65, Springer — Verlag, Berlin, 1937.

532. Luchak, G.: The Solution of the Single-channel Queueing Equation Characterized by a Time-dependent Poisson-distribution Arrival Rate and a General Class of Holding Times, *Operations Research*, vol. 4, pp. 711—732, 1956.

533. Luchak, G.: The Distribution of the Time Required to Reduce to Some Pre-assigned Level a Single-channel Queue Characterized by a Time-dependent Poisson-distributed Arrival Rate ana a General Class of Holding Times, *Operations Research*, vol. 5, pp. 205—209, 1957.

534. Luchak, G.: The Continuous Time Solution of the Equations of the Single Channel Queue with a General Class of Service-time Distributions by the Method of Generating Function, *J. Roy. Statist. Soc., Ser. B*, vol. 20, pp. 176—181, 1958.

535. Lukaszewicz, J,. and H. Steinhaus: On Determining the „Centre of Copper" of a Telephone Network, *Zastos Mat.*, vol. 1, pp. 299—307, 1954 (Polish, English and Russian summaries).

536. Lundkvist, K.: Calculation of the Grade of Service in Automatic Telephone Systems, *Ericsson Tech.*, vol. 4, pp. 75—81, 1936.

537. Lundkvist, K.: Bestamning av Sannolikheten for Sparrning vid s. k. Stel Koppling, *Tek. Medd. Från. Kungl. Teleg.*, nos. 4—6, 1942.

538. Lundkvist, K.: Method of Computing the Grade of Service in a Selection Stage Composed of Primary and Secondary Switches, *Ericsson Rev.*, pp. 11—17, 1948.

539. Lundkvist, K.: General Theory for Telephone Traffic, *Ericsson Tech.*, vol. 9, pp. 11—40, 1953.

540. Lundkvist, K.: Analysis of General Theory for Telephone Traffic, *Ericsson Tech.*, vol. 11, pp. 3—32, 1955.

541. Lunger, G. F.: Bibliography on Queueing Theory, Sperry Rand Corporation, Univac Division, 1959.

542. Mack, C.: The Efficiency of *N* Machines Unidirectionally Patrolled by One Operator When Walking Time Is Constant and Repair Times Are Variables, *J. Roy. Statist. Soc., Ser. B*, vol. 19, pp. 173—178, 1957.

543. Mack, C., T. Murphy, and N. L. Webb: The Efficiency of *N* Machines Unidirectionally Patrolled by One Operative When Walking Time and Repair Time Are Constants, *J. Roy. Statist Soc., Ser. B*, vol. 19 no. 1, pp. 166—172, 1957.

544. Maier-Leibnitz, M.: Die Koinzidenzmethode und ihre Anwendung auf Kernphysicalische Probleme, *Phys. Z.*, vol. 43, pp. 333—362, 1942 (German).

545. Malcolm, D. C.: Queueing Theory in Organization Design, *J. Ind. Eng.*, vol. 6, pp. 19—26, 1955.

546. Mangelsdorf, T. A.: Applications of Waiting-line Theory to Manufacturing Problems, Master's Thesis, Harvard University, 1955.

547. Mann, H. B.: A Note on the Correction of Geiger—Müller Counter Data, *Quart. Appl. Math.*, vol. 4. pp. 307—309, 1946.

548. Marshall, A. W., and E. Reich: Study of the Characteristics of Queues in Tandem, *Operations Research*, vol. 4, p. 386, 1956.

549. Marshall, B. O., Jr.: Queueing Theory, in Joseph F. McCloskey and Florence N. Trefethen (eds.), „Operations Research for Management," vol. I, pp. 134—148, Johns Hopkins Press, Baltimore, 1954.

550. Martin, N. H.: A Note on the Theory of Probability Applied to Telephone Traffic Problems, *P. O. Elec. Engrs. J.*, vol. 16, pp. 237—341, 1923.

551. Massé, D.: Les Réserves et la régulation de l'avenir dans la vie économique (2 vols.). Vol. I: Avenir determinè; vol. II: Avenir aléatoire, Actualités Scientifiques et Industrielles, nos. 1007 and 1008, Hermann & Cie, Paris, 1954 (French).

552. Mathewson, J. H., D. L. Trautman, and D. L. Gerlough: Study of Traffic Flow by Simulation, presented at Thirty-fourth Annual Meeting of the Highway Research Board, Washington, D. C., 1955.

553. Mayne, A. J.: Some Further Results in the Theory of Pedestrians and Road Traffic, *Biometrica*, vol. 41, pp. 375—389, 1954.

554. McCloskey, J., and F. Trefethen (eds.): „Operations Research for Management", vol. 1, Johns Hopkins Press, Baltimore, 1954.

555. McCloskey, J., and J. M. Coppinger (eds.): „Operations Research for Management," vol. 2, Johns Hopkins Press, Baltimore, 1956.

556. McGuire, C. B., and C. B. Winsten: A Theoretical Study of Traffic Congestion, *J. Operations Research Soc. Am.*, vol. 1, pp. 149—150, 1953.

557. McMillan, B., and J. Riordan: A Moving Single Server Problem, *Ann. Math. Statist.*, vol. 28, 1957.

558. Mehlis, A.: Ist der Einsatz II Vorwahlern in der Vorwählstufe Gerecht Fertigt? *FTZ — Fernmeldetech. Z.*, vol. 4, p. 545, 1951 (German).

559. Mehlis, A.: Wahler oder Schalter als Verbindungsorgane in der Fernsprechvermittlungstechnik, *FTZ — Fernmeldetech. Z.*, vol. 5, pp. 293—296, 1952 (German).

560. Meinesz, M.: The Problem of the Gambler's Ruin, *Statist. Need.*, vol. 10, no. 2, pp. 87—97, 1956.

561. Meisling, T.: Discrete-time Queueing Theory, *Operations Research*, vol. 6, no. 1, pp. 96—105, 1958.

562. Meisling, T.: All-epoch Queueing Theory, Stanford Research Institute, English Division, July, 3, 1958 (mimeo. notes).

563. Mellor, S. D.: Delayed Call Formulae When Calls Are Served in Random Order, *P. O. Elec. Engrs. J.*, vol. 35, pt. 2, pp. 53—56, 1942.

564. M e l l o r, S. D.: A Method of Assessing the Effect of Flexibility in Distribution Networks, *P. O. Elec. Engrs J.*, vol. 45, pp. 125—128, 1952.

565. M e r c e r, A.: A Queueing Problem in Which the Arrival Times of the Customers Are Scheduled, *J. Roy. Statist. Soc.*, *Ser. B*, vol. 22, no. 1, pp. 108—113, 1960.

566. M e r k e r, M.: Some Notes on the Use of the Probability Theory to Determine the Number of Switches in an Automatic Telephone Exchange, *P. O. Elec. Engrs. J.*, vol. 16—17, pp. 27—50, 347—362, 1924.

567. M i l l e r, R. G., Jr.: Priority Queues, *Ann. Math. Statist.*, vol. 31, no. 1, March, 1960.

568. M i l l e r, R. G., Jr.: A Contribution to the Theory of Bulk Queues, *J. Roy. Statist. Soc.*, vol. 21, no. 2, pp. 320—337, 1959.

569. Ministry of Transportation and Civil Aviation: Theoretical Researches into Traffic Congestion, *Operations Research Memo* 18.

570. M i t c h e l l, H.: Machine Interference, Cotton Board Work Study Course, Tutorial Notes.

571. M. I. T. Summer Short Course in Operations Research, Technology Press, M. I. T., Cambridge, Mass., 1953.

572. M. I. T. Interim Report 2, Fundamental Investigations in Methods of Operations Research, Apr. 1, 1954—Nov. 30, 1954.

573. M o d e e, G.: Undersökning av Kombinations Gruppernas Utnyttjande vid det Amerikanska Koordinatväljarsystemet, *Tek. Medd. Fran Kungl. Tele.*, pp. 7—9, 1943 (Swedish).

574. M o l i n a, E. C.: The Theory of Probabilities Applied to Telephone Trunking Problems, *Bell System Tech. J.*, vol. 1, p. 69, 1922.

575. M o l i n a, E. C.: Application of the Theory of Probability to Telephone Trunking Problems, *Bell System Tech. J.*, vol. 6, pp. 461—494, 1927.

576. M o l i n a, E. C.: "Poisson's Exponential Binomial Limit", D. Van Nostrand Company, Inc., Princeton, N. J., 1947.

577. M o l n a r, I.: Delay Probability Charts for Telephone Traffic Where the Holding Times Are Constant, Auto. Elec. Co., *Engrs. Notes* 2032, 1952.

578. M o l n a r, I.: A Study of Traffic in Automatic Toll Telephone Systems, Ph. D. Thesis, Northwestern University, 1950.

579. M o l n a r, I.: Toll Answering Delays and Their Measurement, *Auto. Elec. Tech. J.*, vol. 2, pp. 11—21, 1951.

580. M o l n a r, I.: On the Accuracy of Holding-time Measurements, *Auto. Elec. Tech. J.*, vol. 3, pp. 63—72, 1952.

581. M o l n a r I.: The Appraisal of Delays in Gate-type Operation, *Trans. Am. Inst. Elec. Engrs.*, vol. 74. pp. 475—485, 1955.

582. M o r a n, P. A. P.: A Probability Theory of Dams and Storage Systems, *Australian J. Appl. Sci.*, vol. 5, pp. 116—124, 1954.

583. M o r a n, P. A. P.: A Probability Theory of Dams and Storage Systems — Modification of the Release Rule, *Australian J. Appl. Sci.*, vol. 6, pp. 117—130, 1955.

584. M o r a n, P. A. P.: A Probability Theory of a Dam with a Continuous Release, *Quart. J. Math.*, Oxford Series, vol. 2, no 7, pp. 130—137, 1956.

585. M o r a n, P. A. P.: „The Theory of Storage", Methuen's Monographs, Methuen & Co., Ltd., London, 1959.

586. M o r g a n, T. J., „Telecommunication Economics," p. 452, Macdonald & Company, Ltd., London, 1958.

587. M o r l a t, G.: Sur une généralisation de la loi de Poisson, *Compt. rend.*, vol. 234, p. 933, 1952.

588. M o r s e, P. M., and M. L. F r n s t: „Waiting Lines — and Examples and Applications — from Notes from M. I. T. Summer Course on Operations Research, June 16 to July 3, 1953", pp. 93—101, 102—108, Technology Press, M. I. T., Cambridge, Mass., 1953.

589. M o r s e, P. M., H. N. G a r b e r, and M. L. E r n s t: A Family of Queueing Problems, *J. Operations Research Soc. Am.*, vol. 2, pp. 444—445, 1954.

590. M o r s e, P. M.: Stochastic Properties of Waiting Lines, *J. Operations Research Soc. Am.*, vol. 3, pp. 255—261, 1955.

591. M o r s e, P. M.: „Queues, Inventories and Maintenance," John Wiley & Sons, Inc., New York, 1958.

592. M o r t i m e r, W. J.: Moving Vehicle Method Estimating Traffic Volumes and Speeds, *Traffic Eng.*, vol. 28, pp. 539—544, 1956.

593. M o s k o w i t z, K.: Waiting for a Gap in a Traffic Stream, *Highway Research Board, Proc.*, vol. 33, pp. 385—394, 1954.

594. M u l l e r, H. L.: Wachtjdenproblem bij Bedienind van een Groep Machines, Doctoral Thesis, University of Delft, 1951 (Dutch).

595. M u r d o c h, J.: Congestion Problems in Industry, *Appl. Statist.*, vol. 3, p. 200, 1954.

596. M u r r a y, L. J.: Crossbar Systems, *A. T. E. Tech. Soc. Paper*, Nov. 8, 1948.

597. M u r r a y, L. J.: Electronics in the Switching of Telephone Calls, *Engineering*, vol. 176, pp. 413—415, Sept. 25; pp. 445—447, Oct. 2, 1953.

598. M u r r a y, L. J.: Development of Automatic Telephony in Great Britain, *A. T. E. Journal*, vol. 10, no. 4, pp. 271—293, 1954.

599. M y e r s, O.: Common Control Telephone Switching Systems, *Bell System Tech. J.*, vol. 31, pp. 1086—1120, 1952.

600. M y e r s, O.: Automatic Alternate Routing of Telephone Traffic, *Bell Labs. Record*, vol. 32, p. 51, 1954.

601. N a o r, P.: On Machine Interference, *J. Roy. Statist. Soc., Ser. B.* vol. 18, pp. 280—287, 1956.

602. N a o r, P.: Normal Approximation to Machine Interference with Repair Men, *J. Roy. Statist. Soc., Ser. B*, vol. 19, no 2, 1957.

603. N a o r, P.: Some Problems of Machine Interference, *Proc. Intern. Conf. Operations Research*, Sept. 2—6, 1957.

604. N e l s o n, R. T.: Waiting Time Distribution for Application to a Series of Service Centers, *Operations Research*, vol. 6, pp. 856—862, 1958.

605. N e l s o n, R. T.: An Extension of Queueing Theory Results to a Series of Service Centers, *Univ. Calif. Management Sci. Research Project, Notes and Discussion Paper* 68, Apr. 29, 1958.

606. N e o v i u s, G.: Artificial Traffic Trials Using Digital Computers, *Ericsson Tech.*, no. 2, 1955.

607. N e v e u, J.: Théorie des semi groupes de Markov, *Univ. Calif., Publs. in Statist.*, vol. 2, no. 14, pp. 319—394, 1958.

608. N e w e l l, A.: The Capacity of a Railroad Freight Yard (A Survey of the Problem — Not a Solution), *George Washington Univ. Logistics Research Project, App. I to Quart. Progr. Rept., Logistics Papers, Issue* 3, May 16 — Aug. 15, 1950.

609. N e w e l l, G. F.: Mathematical Models for Freely-flowing Highway Traffic, *J. Operations Research Soc. Am.*, vol. 3, pp. 176—186, 1955.

610. N e w e l l, G. F.: Statistical Analysis of the Flow of Highway Traffic through a Signalized Intersection, *Appl. Math.*, vol. 13., no 4, January, 1956.

611. N e w e l l, G. F.: Queues for a Fixed-cycle Traffic Light, *Ann. Math. Statist.*, vol. 31, no 3, p. 589, 1960.

612. N e w l a n d, W. F.: A Method of Approach and Solution to Some Fundamental Traffic Problems, *P. O. Elec. Engrs. J.*, vol. 25, pp. 119—131, 1932.

613. N e y m a n, J.: „Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability", University of California Press, Berkeley. Calif., 1949. Also „Proceedings of the Second Symposium, 1951," and „Proceedings of the Third Symposium, 1955—1956".

614. N i c o l i e c h i a, P. E.: Sui metodi contabili del trafico telefonico, *Poste e telecomun.*, vol. 8, pp. 435—451, 1955.

615. O' B r i e n, G. G.: The Solution of Some Queueing Problems, *J. Soc. Ind. Appl. Math.*, vol. 2, pp. 133—142, 1954.

616. O' C o n n o r, T. F.: „Productivity and Probability — A Treatise on Time Study and the Improvement of Industrial Efficiency", chap. VI, Mech. World Monographs 65, Emmott & Co., Ltd., London, 1952.

617. O' D e l l, G. F.: The Influence of Traffic in Automatic Exchange Design, *P. O. Elec. Engrs. Inst.*, no. 85, 1920.

618. O'Dell, G. F.: Theoretical Principles of the Traffic Capacity of Automatic Switches, *P. O. Elec. Engrs. J.*, vol. 13, pp. 209—223, 1920.
619. O'Dell, G. F., and W. W. Gibson: Automatic Trunking in Theory and Practice, *P. O. Elec. Engrs. Inst., Profess. Paper 107*, 1926 (41 pp.).
620. O'Dell, G. F: An Outline of the Trunking Aspect of Automatic Telephony, *J. Inst. Elec. Engrs. (London)*, vol. 65, p. 185, 1927.
621. Olcott, E. S.: The Influence of Vehicular Speed and Spacing of Tunnel Capacity, *J. Operations Research Soc. Am.*, vol. 3, pp. 147—167, 1955 (also „Operations Research for Management," vol. 2. chap. 3, Johns Hopkins Press, Baltimore, 1956).
622. Olshevsky, D. E.: A Study of Flight Time Estimate Errors, *Cornell Aeronaut. Lab. Rept.* JA-693-1, May, 1951.
623. Oostrum, H. R. Van: Verkeersmetingen, Verkeersprognose en Contactbankindeligen, *Telegram en Telefoon*, vol. 57, pp. 3—23, 1956.
624. Orden, A.: The Transshipment Problem, *Management Sci*, vol. 2. pp. 276—285, 1956.
625. Ore, O.: Studies on Directed Graphs, part I, *Ann. Math.*, vol. 63, no 3, pp. 383—406, May, 1956.
626. Palasti, I., A. Renyi, T. Szentmartony, and L. Takacs: Ergänzung des Lagervorrates, I, *Magyar Tudomànyos Akad. Alkalm. Mat. Int. Közlemènyei*, vol, 2, pp. 187—201, 1953 (Hungarian; German and Russian summaries). (Also see Ref. 896).
627. Palm, C.: Calcul exact de la perte dans le groups de circuits èchelonnes, *Ericsson Tech.*, vol. 4, 1936.
628. Palm, C.: Some Investigations into Waiting Times in Telephone Plants, *Tek. Medd. Från. Kungl. Teleg.*, no. 7—9, 1937 (Swedish).
629. Palm, C.: Inhomogeneous Telephone Traffic in Full Availability Groups, *Ericsson Tech.*, vol. 5, pp. 3—36, 1937.
630. Palm, C.: Étude des délais d'attente, *Ericsson Tech.*, vol. 5, p. 39, 1937.
631. Palm, C.: Analysis of the Erlang Traffic Formulae for Busy-signal Arrangements. *Ericsson Tech.*, vol. 6, pp. 39—58, 1938.
632. Palm, C.: Mätnoggrannhet vid Bestämning av Trafikmängd Enligt Genomsökningsförfarandet, *Tek. Medd. Från. Kungl. Teleg.*, nos. 7—9, 1941 (Swedish).
633. Palm, C.: A Form Factor for Judging Delay Time Distribution, *Tek. Medd. Från. Kungl. Teleg.*, nos. 1—3, 1943 (Swedish).
634. Palm, C.: Intensity Fluctuations in Telephone Traffic, *Ericsson Tech.* vol. 1, no. 44, pp. 1—18S, 1943.
635. Palm, C.: Några Anmarkningar över de Erland'ska Formlerna for Upptaget-System, Specialnummer for Teletrafikteknik, *Tek. Medd. Från. Kungl. Teleg.*, pp 1—110, 1946 (Swedish).
636. Palm, C.: The Distribution of Repairmen in Servicing Automatic Machines, *Industritidn. Norden*, vol. 75, pp. 75—80, 90—94, 119—123, 1947 (Swedish; an account is given in W. Feller, Ref. 236).
637. Palm, C.: „Tables of the Erlang Loss Formula," Telefon Aktiebolaget, L. M. Ericsson, Stockholm, 1947 (2d ed., 1954).
638. Palm, C.: Waiting Times When Traffic Has Variable Mean Intensity, *Ericsson Rev.*, vol. 24, pp. 102—107, 1947.
639. Palm, C.: Methods of Judging Annoyance Caused by Congestion, *Tele.* (English ed.), vol. 2, pp. 1—20, 1953.
640. Palm, C.: Research on Telephone Traffic Carried by Full Availability Groups, *Tele.* (English ed.), no. 1, 1957.
641. Pearcey, T.: Delays in Landing of Air Traffic, *J. Roy. Aeronaut. Soc.*, vol. 52, pp. 799—812, 1948.
642. Peck, L. G., and R. N. Hazelwood: „Finite Queueing Tables", John Wiley & Sons, Inc., New York, 1958.

643. Phillips, C. R., Jr.: A Variable-channel Queueing Model with a Limited Number of Channels, Thesis, Georgia Institute of Technology, 1960.
644. Phipps, T. E., Jr.: Machine Repair as a Priority Waiting Line Problem, *Operations Research*, vol. 4, pp. 76—85, 1956 (Comments by W. R. Van Voorhis, *ibid.*, p. 86).
645. Piesch, J.: Die Beanspruchung der Bundel im modernen Fernspechverkehr, *Arch. Elektrotech.*, vol. 8, pp. 324—328, 353—362, 411—429, 1954 (German).
646. Pilé, G.: Étude des délais d'attente des aéronefs á l'atterissage, *Rev. statist. appl.*, vol. 3, pp. 73—84, 1955 (French).
647. Pilliod, J. J.: Fundamental Plans for Toll Telephone Plant, *Bell System Tech. J.*, vol. 31, pp. 832—850, September, 1952 (also in *Trans. Am. Inst. Elec. Engrs.*, vol. 71, no. 1, p. 248, 1952).
648. Pipes, L. A.: A Proposed Dynamic Analogy of Traffic, University of California, Special Study Institute of Transport and Traffic Engineering, vol. 11 July, 1950.
649. Pipes, L. A.: A Mathematical Analysis of Traffic Dynamics, *J. Operations Research Soc. Am.*, vol. 1, p. 151, 1953.
650. Pipes, L. A.: An Operational Analysis of Traffic Dynamics, *J. Appl. Phys.*, vol. 24, pp. 274—281, 1953.
651. Pitt, H. R.: A Theorem on Random Functions with an Application to a Theory of Provisioning, *J. London Math. Soc.*, vol. 21, pp. 16—22, 1946.
652. Pollaczek, F.: Über eine Aufgabe der Wahrscheinlichkeitstheorie, part I, *Math. Z.*, vol. 32, pp. 64—100, 1930; part II, *ibid.*, vol. 32, pp. 729—750, 1930.
653. Pollaczek, F: Theorie des Warten vor Schaltern, *Telegraphen u. Fernsprechtechnik*, vol. 19, pp. 71—78, 1930.
654. Pollaczek, F.: Über zwei Formeln aus der Theorie des Wartens von Schaltergruppen, Gesprächverluste und Wartezeiten, *Elek. Nachr. Tech.*, vol. 8, pp. 256—268, 268—279, 1931.
655. Pollaczek, F.: Zur Theorie des Wartens von Schaltergruppen, *Elek. Nachr. Tech.*, vol. 9, pp. 434—454, 1932.
656. Pollaczek, F.: Lösung eines geometrischen Wahrscheinlichkeitsproblemes, *Math. Z.*, vol. 35, pp. 230—278, 1932.
657. Pollaczek, F.: Kurven für die Gesprächverlust in Volkommenen Leitung-Shiendel, *Elek. Nachr. Tech.*, vol. 10, 1933.
658. Pollaczek, F.: Über das Warteproblem, *Math. Z.*, vol. 38 pp. 492—537, 1934.
659. Pollaczek, F.: *Elek. Nachr. Tech.*, vol. 11, pp. 396—399, 1934.
660. Pollaczek, F.: *Ann. P. T. T.*, vol. 24, pp. 997—1001, 1935.
661. Pollaczek, F.: Sur quelques lois asymptotiques de la thèorie de l'encombrement des réseaux téléphoniques, *Ann. univ. Lyon, Sec. A*, vol. 5, pp. 21—35, 1942.
662. Pollaczek, F.: Résolution de certaines èquations intégrales de deuxiéme espace, *J. math. pures appl.*, vol. 24, pp. 73—93, 1945.
663. Pollaczek, F.: La loi d'attente des appels téléphoniques, *Compt. rend.*, vol. 222, pp. 353—355, 1946.
664. Pollaczek, F.: Sur l'application de la thèorie des fonctions au calcul de certaines probabilités continues utilisées dans la théorie des réseaux téléphoniques, *Ann. inst. Henri Poincaré*, vol. 10. pp. 1—54, 1946.
665. Pollaczek, F.: Sur un problème de calcul des probabilités qui se rapporte á la téléphonie, *J. math. pures appl.*, vol. 25, pp. 307—334, 1946.
666. Pollaczek, F.: Sur la probabilité de perte d'un appel téléphonique dans le cas d'un seul groupe de lignes avec blocage temporaire, *Compt. rend.*, vol. 226, pp. 2045—2047, 1948.
667. Pollaczek, F.: Application d'opérateurs intégrocombinatoires dans la théorie des intégrales multiples de Dirichlet, *Ann. inst. Henri Poincaré*, vol. 11, pp. 113—133, 1949.
668. Pollaczek, F.: Réductions de différents problémes concernant la probabilité d'attente au téléphone, à la résolution de systèmes d'équations intégrales, *Ann. inst. Henri Poincaré*, vol. 11, pp. 135—173, 1949.

669. Pollaczek, F.: Application du calcul des probabilités au phénomène de blocage temporaire des lignes téléphoniques, *Ann. télécommun.*, vol. 6, pp. 49—53, 1951.
670. Pollaczek, F.: Problèmes de calcul des probabilités relatifs à des systèmes téléphoniques sans possibilité d'attente, *Ann. inst. Henri Poincaré* vol. 12, pp. 57—96, 1951.
671. Pollaczek, F.: Répartition des délais d'attente des avions arrivant à un aéroport qui possédes pistes d'atterrissage, *Compt. rend.*, vol. 232, pp. 1901—1903, 2286—2288, 1951.
672. Pollaczek, F.: Délais d'attente des avions atterrissant selón leur ordre d'arrivée sur un aéroport à sistes, *Compt. rend.*, vol. 234, pp. 1246—1248, 1952.
673. Pollaczek, F.: Sur la répartition des périodes d'occupation ininterrompue d'un guichet, *Compt. rend.*, vol. 234, pp. 2042—2044, 1952.
674. Pollaczek, F.: Fonctions caractéristiques de certaines répartitions définies au moyen de la notion d'ordre application á la théorie des attentes, *Compt. rend.*, vol. 234, pp. 2334—2336, 1952.
675. Pollaczek, F.: Sur une generalisation de la theorie des attentes, *Compt. rend.*, vol. 236, pp. 578—580, 1953.
676. Pollaczek, F.: Généralisation de la théorie probabiliste des systèmes téléphoniques sans dispositif d'attente, *Compt. rend.*, vol. 236, pp. 1469—1470, 1953.
677. Pollaczek, F.: Sur la théorie stochastique des compteurs électroniques, *Compt. rend.*, vol. 238, pp. 766—768, 1954.
678. Pollaczek F.: Dévelopment de la théorie stochastique des lignes téléphoniques pour un état initial quelconque, *Compt. rend.*, vol. 239, pp. 1764—1766, 1954.
679. Pollaczek, F.: "Problèmes stochastiques posés par le phénomène de formation d'une queue d'attente á un guichet et par des phénomènes apparentes", Memorial des Sciences Mathématiques, Gauthier-Villars, Paris, 1957.
680. Pollaczek, F.: Application de la théorie des probabilités posées par l'encombrement des réseaux téléphoniques, *Ann. télécommun.*, vol. 14, nos. 7—8, pp. 165—183, 1959.
681. Pomey, F. L.: Téléphone et statistique, *Rev. gén. élec.*; vol. 9, pp. 133—138, 1921.
682. Popovic, Ż.: The Efficiency of Grading with Cyclic Arrangement, *Telekomunikacije*, vol. 9, pt. 4, 1955.
683. Port of New York Authority, Operations Standards Division: A Collection of Delay Probability Formulas and Curves, 1956.
684. Prabhu, N. U.: Some Results for the Queue with Poisson Arrivals, *J. Roy. Statist. Soc., Ser. B.* vol. 22, no. 1, pp. 104—107, 1960.
685. Prabhu, N. V.: Application of Storage Theory to Queues with Poisson Arrivals, *Ann. Math. Statist.*, vol. 31, no. 2, p. 475, 1960.
686. Prager, W.: Problems of Traffic and Transportation, *Proc. Symposium on Operations Research in Business and Industry*, pp. 105—113, Midwest Research Institute, Kansas City, Mo., 1954.
687. Prigogine, I., and F. S. Andrews: A Boltzmann-like Approach for Traffic Flow, *Operations Research*, vol. 8, no. 6, p. 789, 1960.
688. Prim, R. C.: Shortest Connection Networks and Some Generalizations, *Bell System Tech. J.*, vol. 36, pp. 1389—1401, 1957.
689. Prince, R. K.: Queueing Up, *Proc. Eng.*, pp. 188—192, November, 1955.
690. Proceedings of the First International Congress on the Application of the Theory of Probability in Telephone Engineering and Administration, Copenhagen, June 20—23, 1955 [issued as vol. 1, no. 1, of *Teleteknik* (English ed.), 1957].
691. Pyke, R.: On Renewal Processes Related to Type I and Type II Counter Models, *Ann. Math. Statist.*, vol. 29, pp. 737—754, 1958.
692. Pyke, R.: Markov Renewal Processes: Definitions and Preliminary Properties [this research at Columbia University was supported by the Office of Naval Research under Contract Nonr 266 (59) Proj. No. 042—205], October, 1959.
693. Rabe, F. W.: Considerations on the Correctness of the Erlang Formula for Automatic Telephone Traffic, *Ingenieur (Utrecht)*, vol. 58, p. 12, 1946.

694. R a b e, F. W.: Variations of Telephone Traffic, *Elec. Commun.*, vol. 26, pp. 243—248, 1949.

695. R a f f, M. S.: Space-time Relationships at „Stop“ Intersections, *Proc. Inst. Traffic Eng.*, vol. 20, pp. 42—49, 1949.

696. R a f f, M. S.: „A Volume Warrant for Urban Stop Signs“, chap. VI, *Eno Foundation for Highway Traffic Control*, Saugatuck, Conn., 1950.

697. R a f f, M. S.: The Distribution of Blocks in an Uncongested Stream of Automobile Traffic, *J. Am. Statist. Assoc.*, vol. 46, no. 253, March, 1951.

698. R a f f, M. S.: Answer to Question 35 — Waiting Time for Public Transportation, *Am. Statist.* vol. 7, no. 2, pp. 26—27, April—May, 1953: also *ibid.*, vol. 7. no. 3, p. 20, 1953.

699. R a m a k r i s h n a n, A.: Some Simple Stochastic Processes, *J. Roy. Statist. Soc.*, vol. 13(B), p. 131, 1951.

700. R a m a k r i s h n a n, A., and P. M. M a t h e w s: A Stochastic Problem Relating to Counters, *Phil. Mag.*, vol. 44, no. 7, pp. 1122—1128, 1953.

701. R a m m, B. T.: The Determination of Branch Staff Establishments in a Multiple Shop Organization, *Appl. Statist.*, vol. 4, pp. 295—298, 1955.

702. R a p p, Y.: The Economic Optimum in Urban Telephone Network Problems, *Ericsson Tech.*, no. 49, pp. 1—132, 1950.

703. R a p p l e y e, S. C.: A Study of the Delays Encountered by Toll Operators in Obtaining an Idle Trunk, *Bell System Tech. J.*, vol. 25, pp. 53—62, 1946.

704. R a w d i n, E.: A Measure of Effectiveness for Queueing Problems, *Operations Research*, vol. 8, no. 2, p. 278, March-April, 1960.

705. R a y, D.: Stable Processes with an Absorbing Barrier, *Trans. Am. Math. Soc.*, vol. 89, pp. 16—24, 1958.

706. R e e d, I. S.: Some Queue Problems Associated with Buffer-storage Systems, *Mass. Inst. Technol.*, *Lincoln Lab. Tech. Rept.* 73, Dec. 2, 1954.

707. R e i c h, E.: Birth-Death Process and Tandem Queues, The Rand Corporation, *Paper* P863, 1956.

708. R e i c h, E.: Waiting Times When Queues are in Tandem, *Ann. Math. Statist.*, vol. 28, no. 3, p. 768, September, 1957.

709. R e i c h, E.: On an Integrodifferential Equation of Takács, part I, *Ann. Math. Statist.*, vol. 29, pp. 563—570, 1958.

710. R e i c h, P. G.: Instability in Some Flow Systems of Civil Aviation, *Operational Research Quart.*, vol. 8, pp.126—132, 1957.

711. R e n i e r, P.: Criteria for Maximum Economy in Planning Local Telephone Networks, *Electro-Tech.*, vol. 40, pp. 160—163, 1953 (Italian).

712. R e n y i, A., and T. S z e n t m a r t o n y: Wahrscheinlichkeitstheoretisch Bestimmung von Reserven von Maschinenteilen unter Ausrüstongsgegenstanden, *Mat. Lapok.*, vol. 3, pp. 129—139, 1952.

713. R e n y i, A., and L. T a k á c s: Sur les processus d'événements dérivés par un processus de Poisson et sur leurs applications techniques et physiques, *Magyar Tudomànyos Akad. Alkalm. Mat. Inst. Közleményei*, vol. 1, pp. 139—146, 1952 (Hungarian; French and Russian summaries).

714. R e n z, A.: Graphisches Verfahren für die Addition ungleichen Verkehrswerte, *Fernmelde-Praxis*, vol. 33, pp. 243—254, 1956 (German).

715. R e n z, A.: Teilung und Zusammenfassung ungleichgrosser Verkehrswerte, *Fernmelde-Praxis*, vol. 34, pp. 170—173, 1957.

716. Reports presented at the Second International Course in Traffic Engineering, Switzerland, Sept. 20—25, 1954. (W. T. & A. O. and P. I. A. R. C.) London, W. T. A. O. 1954 (Methods for Measuring Distribution of Traffic Volume).

717. R e t t e n b e r g e r, J.: Die Bestimmung der optimale Leitungszahlen bei Fernsprechnetzen mit Überlaufverkehr, *NTZ—Nachr. Tech. Z.*, vol. 10, pp. 53—59, 1957.

718. R e u s c h e l, A.: The Movement of a Column of Vehicles When the Leading Vehicle Is Uniformly Accelerated or Decelerated, *Z. österr. Ing. Archit. Ver.*, vol. 95, pp. 59—62, 73—77, 1950.

719. R e u t e r, G. E. H.: Denumerable Markov Processes and the Associated Contraction Semigroups on 1, *Acta Math.*, vol. 97—98, pp. 1—46, 1957.

720. Reuter, G. E. H., and W. Ledermann: On Differential Equations for the Transition Probabilities of Markov Processes with Enumerably Many States, *Proc. Cambridge Phil. Soc.*, vol. 49, 1953.
721. Richards, P. I.: Shock Waves on the Highway, *Operations Research*, vol. 4, pp. 42—51, 1956.
722. Riley, V.: Bibliography of Queueing Theory, Appendix A, in „Operations Research for Management", vol. 5, pp. 541—556, Johns Hopkins Press, Baltimore, 1956.
723. Riordan, J., and C. E. Shannon: The Number of Two-terminal Series-Parallel Networks, *J. Math. and Phys.*, vol. 21, pp. 83—93, 1942.
724. Riordan, J.: Telephone Traffic Time Averages, *Bell System Tech. J.*, vol. 30, p. 1129, 1951.
725. Riordan, J.: Delay Curves for Calls Served at Random, *Bell System Tech. J.*, vol. 32, pp. 100—119, 1266, 1953.
726. Riordan, J.: Derivation of Moments of Overflow Traffic, in R. I. Wilkinson, Theories for Toll Traffic Engineering in U. S. A., App. I, *Bell. System Tech. J.*, vol. 35, p. 514, 1956.
*727. Riordan, J.: "Stochastic Service Systems", John Wiley & Sons, N. Y., 1962.
728. Rios, S.: "Introduction a los metodos de la estadistica", Instituto Estadistica, Madrid, 1957.
729. Robinson, F. D., and W. E. Duckworth: An Application of Queueing Theory to the Speed of Estimating, *Proc. Intern. Conf. Operations Research*, Sept. 2—6, 1957.
730. Rodenburg, D.: Alternative Routing of Junction Traffic, *Commun. News*, vol. 10, p. 30, 1949.
731. Rodenburg, N.: Some Problems Relating to a Telephone System Employing Nonhoming Selectors, *Commun. News*, vol. 13, pp. 69—114, May, 1953.
732. Rohde, K., and H. Störmer: "Passage Probabilites" in Switching Facilities of Communication Engineering, *Mitt. Bl. Math. Statist.*, vol. 5, nos. 2—3, 1953.
733. Rohde, K., and Bretchneider: Econòmic Traffic Standards in Telephone Systems, *Frequency*, vol. 8, no. 8, pp. 233—239, August, 1954 (German).
734. Rohde, K.: Considerations on the Significance of Random Sampling Process in Determining the Magnitude of Telephone Traffic (International Teletraffic Congress, Copenhagen, 1955), *Teleteknik*, vol. 1, pp. 86—90, 1957 (German).
735. Romani, J.: La Teoria de las colas aplicàda a un problema de produccion industrial, *Trabajos de estadistica*, vol. 6, cuaderno III, 1955.
736. Romani, J.: Distribution of the Algebraic Sum of Poisson Variables, *Trabajos de estadistica*, vol. 7, no. 2, pp. 175—181, 1956.
737. Romani, J.: Un Modelo de la teoria de colas con numero variable de canales, *Trabajos de estadistica*, vol. 8, 1957.
738. Romani, J.: A Queueing Model with a Variable Number of Channels, *Trabajos de estadistica*, vol. 8, no. 3, pp. 175—189, 1957.
739. Roy, J. le: Recherche en deux ordres inverses dans la sélection en téléphonie automatique, *Ann. P. T. T*, vol. 5, p. 366, 1950.
740. Rowdin, E.: Correction and Addendum to a New Measure of Effectiveness for Queueing Problems, *Operations Research*, vol. 8, no. 6, p. 868, 1960.
741. Rückle, G., and F. Lubberger: "Der Fernsprechverkehr als Massenerscheinung mit Starken Schwankungen", Springer-Verlag, Berlin, 1924.
742. Runnenburg, J. Th: Machines Served by a Patrolling Operator, *Math. Centrum, Statist. Afdel. Rept.* S221 (VP13), Amsterdam, 1957 (13 pp.).
743. Ryll-Nardzewski, C.: On the Homogeneous Poisson Process, *Studia Math.*, vol. 14, pp. 124—128, 1953.
744. Saaty, T. L.: Résumé of Useful Formulas in Queueing Theory, *Operations Research*, vol. 5, no. 2, pp. 161—200, April, 1957.
745. Saaty, T. L.: Five Papers by Conny Palm, *Operations Research*, vol. 6, no. 3, pp. 456—460, May—June, 1958.
746. Saaty, T. L.: "Mathematical Methods of Operation Research", Mc Graw-Hill Book Company, Inc., New York, 1959.
Русский перевод: Саати Т. Математические методы исследования операций, Воениздат, 1963.

747. S a a t y, T. L.: Some Stochastic Processes with Absorbing Barriers, *J. Roy. Statist. Soc.*, October, 1961.
748. S a a t y, T. L.: Time Dependent Solution of the Many Server Poisson Queue, *Operations Research*, November—December, 1960.
749. S a a t y, T. L.: Approximate Solution to Semiconductor Noise as a Queueing Problem, *Proc. IRE*, June, 1961.
750. S a a t y, T. L.: On Jockeying, Collusion, Scheduling, Optimization and Graph Theoretic Queues, Office of Naval Research, Washington, 1961.
751. S a c k s, J.: Ergodicity of Queues in Series, *Ann. Math. Statist.*, vol. 31, no. 3, p. 579, 1960.
752. S a l v e s o n, M. E.: A Problem in Optimal Machine Loading, *Management Sci.*, vol. 2, p. 232, 1956; also Note, *ibid.*, vol. 3, p. 114, 1957.
753. S a r k a d i, K.: On the Waiting Time of Locomotives, *Magyar Tudományos Akad. Alkalm. Math. Int. Közleményei*, vol. 3, pp. 191—194, 1954 (Hungarian; English and Russian summaries).
754. S c h i l l e r, D. H., and M. M. L a v i n: The Determination of Requirements for Warehous Dock Facilities, *Operations Research*, vol. 4, pp. 231—243, April, 1956.
755. S c h n e i d e r, E., and B. J e s s e n: Absatz, Production und Lagerhaltung bei einfacher Production, *Arch. Mat. Wert. Sozialforsch.*, vol. 4, no. 1, 1938 (German).
756. S c h o f i e l d, H.: Traffic Congestion. An Assessment of the Problem and Measures for Relief, *J. Inst. Munic. Eng.*, vol. 83, no. 2, pp. 49—62, 1956; Discussion, *ibid.*, pp. 62—63, 81—82.
757. S c h o u t e n, J. P., and J. G i l t a y: Oplessing van een Problem uit de Waarschijnlijkheidsrekening van Belang vaar de Automatische Telephonie, *Ingenieur (Utrecht)*, vol. 55, pp. 67—75, 1940 (Dutch).
758. S c h u h l, A.: Probability Calculations and Vehicle Movement on a Two Lane Road, *Ann. ponts et chaussées*, vol. 125, pp. 631—663, 1955 (French; English summary).
759. S e g e r d a h l, C. O.: On Homogeneous Random Processes and Collective Risk Theory, Uppsala, 1939.
760. S e s p a n i a k, L. J.: An Application of Queueing Theory for Determining Manpower Requirements for an In-line Assembly Inspection Department, *J. Ind. Eng.*, vol. 4, pp. 265—267, July—August, 1959.
761. S h a n n o n, C. E.; A Symbolic Analysis of Relay and Switching Circuits, *Trans. Am. Inst. Elec. Engrs.*, vol. 57, pp. 713—723, 1938.
Русский перевод: «Символический анализ релейных и переключательных схем». В кн. К. Ш е н н о н. Работы по теории информации и кибернетике. Изд-во иностранной литературы, 1963, с. 9—45.
762. S h a n n o n, C. E.: The Synthesis of Two-terminal Networks, *Bell System Tech. J.*, vol. 28, pp. 59—98, 1949.
Русский перевод: «Синтез двухполюсных переключательных схем». Там же, с. 59—105.
763. S h a n n o n, C. E., and W. W e a v e r: "The Mathematical Theory of Communication", University of Illinois Press, Urbana, Ill., 1949.
Русский перевод: «Математическая теория связи». Там же, с. 243—332.
764. S h a n n o n, C. E.: Memory Requirements in a Telephone Exchange *Bell System Tech. J.*, vol. 29, pp. 343—349, July, 1950.
Русский перевод: «Требования, предъявляемые к объему памяти телефонного коммутатора». Там же, с. 106—113.
765. S i e m e n s B. & Co. Ltd.: "Traffic Calculations for Trunk and Local Telephone Exchanges", p. 103, London, 1955.
766. S i m o n d.: La Distribution hyperexponentielle, *Rev. franç. recherche opérationnelle*, vol. 2, no. 9, pp. 196—215, 1958.
767. S i t t i g, J.: De statistische Bases van de Planning in een Machine Fabriek, *Statist. Need.*, vol. 9, pp. 47—69, 1955.
768. S k e l l a m, J. G.: The Frequency Distribution of the Difference between Two Poisson Variates Belonging to Different Populations, *J. Roy. Statist. Soc.*, vol. 109, p. 296, 1946.

769. Smeed, R. J., and G. Bennett: Research on Road Safety and Traffic Flow, *Inst. Civil Engrs., Road Paper* 29, 1949.
770. Smith, H. A. B.: Some Aids to Traffic Flow, *Proc. Inst. Civil Engrs.*, II, vol. 2, pp. 416—429, 1953.
771. Smith, W. L.: On the Distribution of Queueing Times, *Proc. Cambridge Phil. Soc.*, vol. 49, pp. 449—461, 1953.
772. Smith, W. L.: Stochastic Sequences of Events, Ph. D. Thesis, Cambridge University, 1953.
773. Smith, W. L.: Asymptotic Renewal Theorems, *Proc. Roy. Soc. Edinburgh, Sec. A*, vol. 64, p. 9, 1954.
774. Smith, W. L.: Regenerative Stochastic Processes, *Proc. Roy. Soc. (London), Ser. A*, vol. 232, p. 6, 1955.
775. Smith, W. L.: Renewal Theory, Couter Problems, and Quasi-Poisson Processes, *Proc. Cambridge Phil. Soc.*, vol. 53, pp. 175—193, 1957.
776. Smith, W. L.: Renewal Theory and Its Ramifications, *J. Roy. Statist. Soc., Ser. B*, vol. 20, no. 2, pp. 243—302, 1958.
Русский перевод: Смит В. Л. Теория восстановления и смежные с ней вопросы. Сб. переводов «Математика», 5 : 3, 1961, с. 95—150.
777. Smith, W. L.: On the Cumulants of Renewal Processes, *Biometrika*, vol. 46, pts. 1—2, June, 1959.
778. Smith, W. L.: Infinitesimal Renewal Processes, *Univ. North Carolina, Inst. of Statist., Mimeo. Ser.* 237, August, 1959.
779. Sneddon, I. N.: "Elements of Partial Differential Equations", Mc Graw-Hill Book Company, Inc., New York, 1957.
780. Sochor, Z.: Choice of a Suitable Size of Subscribers' Distribution Boxes in Local Telephone Networks, *Slaboproudý obsor*, vol. 14, pp. 316—321, 1953 (Czech.).
781. Stadler, F.: On the Characteristics of Telecommunication Traffic, *Slaboproudý Obsor*, vol. 14, 1953 (Chech.).
782. Steffensen, J. F.: Om Sandsynligheden for at Afkommet Uddor, *Mat. Tidsskr., Ser. B*, vol. 19, 1930.
783. Steffensen, N. J.: Deux problèmes du calcul des probabilités, *Ann. inst. Henri Poincaré*, vol. 3, p. 319, 1933.
784. Stephan, F. F.: Two Queues with a Priority Rule and a Single Queue with Interrupted Service, *Operations Research*, vol. 4, p. 386, 1956.
785. Storet, H. P.: Synchronization and Its Effect on Machine Efficiency, *J. Textile Inst.*, vol. 41, pp. T225—T235, 1950.
786. Störmer, H., and F. L. Bauer: Berechnung von Wartezeiten in Vermittlungseinrichtungen mit kleinen Zubringerbündeln, *Arch. Electrotech.*, vol. 9, p. 69, 1955.
787. Störmer, H.: Wartezeitlenkung in Handbedienten Vermittlungsanlagen, *Arch. Electrotech.*, vol. 10, pp. 58—64, 1956.
788. Stribling, J. W.: Work Load Calculations for Winding Machines, *Textile Ind.*, vol. 116, pp. 135—150, 1952.
789. Swan, A. W.: A Method of Graphing Sales Figures for Executive Use, *Appl. Statist.*, vol. 4, pp. 15—21, 1955.
790. Swensson, O.: An Approach to a Class of Queueing Problems, *Operations Research*, vol. 6, no. 2, pp. 276—295, 1958.
791. Syski, R.: The Theory of Congestion in Lost-call Systems, *A. T. E. Journal*, vol. 9, pp. 182—215, 1953.
792. Syski, R.: Analogies between the Congestion and Communication Theories, *A. T. E. Journal*, vol. 11, pp. 220—243, 1955.
793. Syski, R.: Determination of Waiting Time in the Simple Delay System, *A. T. E. Journal*, vol. 13, pp. 281—286, 1957.
794. Syski, R., and J. W. Cohen: Second List of References to Publications on the Application of Probability Theory to Telephone Trunking and Related Problems, P. T. T. publication, paper presented at the International Teletraffic Congress at the Hague, 1958.
795. Syski, R.: "Introduction to Congestion Theory in Telephone Systems", Oliver & Boyd, Ltd., Edinburgh and London, 1960.

796. T a k á c s, L.: Probabilistic Treatment of the Simultaneous Stoppage of Machines with Considerations of the Waiting Times, *Magyar Tudományos Akad. III (Mat. Fiz.) Oszt. Közleményei*, vol. 1, pp. 228—234, 1951 (Hungarian).

797. T a k á c s, L.: Occurence and Coincidence Phenomena in Case of Happenings with an Arbitrary Distribution Law of Duration, *Acta Math. Acad. Sci. Hung.*, vol. 2, pp. 275—298, 1951 (revised English translation of *Magyar Tudományos Akad. III*, vol. 1, pp. 371—386, 1951).

798. T a k á c s, L.: A New Method for Discussing Recurrent Stochastic Processes, *Magyar Tudomànyos Akad. Alkalm. Mat. Int. Közleményei*, vol. 2, pp. 135—151, 1953 (Hungarian; Russian and English summaries).

799. T a k á c s, L.: Coincidence Problems Arising in the Theory of Counters, *Magyar Tudományos Akad. Alkalm. Mat. Int. Közleményei*, vol. 2, pp. 153—163, 1953 (Hungarian; Russian and English summaries).

800. T a k á c s, L.: Investigation of Waiting Time Problems by Reduction to Markov Processes, *Acta Math. Acad. Sci Hung.*, vol. 6, pp. 101—129, 1955.

801. T a k á c s, L.: On Processes of Happenings Generated by Means of a Poisson Process, *Acta Math.*, vol. 6, nos. 1—2, 1955.

802. T a k á c s, L.: On a Probability Problem Arising in Some Traffic Investigations, *Publs. Math. Inst. Acad. Sci. Hung.*, vol. 1, pp. 99—107, 1956 (Hungarian; Russian and English summaries).

803. T a k á c s, L.: On Some Probabilistic Problems Concerning the Counting of Particles (note in paper of A. Bekessy), *Publs. Math. Inst. Acad. Sci. Hung.*, vol. 1, pp. 93—98, 1956 (Hungarian; Russian and English summaries).

804. T a k á c s, L.: О последовательности событий, выбранных счетчиком из рекуррентного процесса событий. «Теория вероятностей и ее применения», т. 1, вып. 1, 1956, с. 90—101 (на англ. яз., рез. на русск. яз.).

805. T a k á c s, L.: On a Probability Problem Arising in Some Traffic Investigations, *Publs. Math. Inst. Acad. Sci. Hung.*, vol. 1, pp. 99—107, 1956 (Hungarian; Russian and English summaries).

806. T a k á c s, L.: On the Probability Problem Arising in the Theory of Counters, *Proc. Cambridge Phil. Soc.*, vol. 52, pp. 488—498, 1956.

807. T a k á c s, L.: On Limiting Distributions Concerning a Sojourn Time Problem, *Acta Math. Acad. Sci. Hung.*, vol. 8, pp. 279—294, 1957.

808. T a k á c s, L.: Задача о времени пребывания. «Теория вероятностей и ее применения», т. 3, вып. 1, 1958, с. 61—69 (на англ. яз., рез. на русск. яз.).

*809. T a k á c s, L.: О стохастическом процессе, связанном с некоторыми задачами о времени ожидания. «Теория вероятностей и ее применения», т. 2, вып. 1, 1957, с. 92—105 (на англ. яз., рез. на русск. яз.).

810. T a k á c s, L.: On a Secondary Stochastic Process Generated by a Multidimensional Poisson Process, *Hung. Acad. Sci.*, vol. 2, pp. 1—2, 1957.

811. T a k á c s, L.: On a Probability Problem Concerning Telephone Traffic, *Acta Math. Acad. Sci. Hung.*, vol. 8, pp. 3—4, 1957.

812. T a k á c s, L.: On a Queueing Problem Concerning Telephone Traffic, *Acta Math. Acad. Sci. Hung.*, vol. 8, pp. 325—395, 1957.

813. T a k á c s, L.: On the Generalization of Erlang's Formula, *Acta Math. Acad. Sci. Hung.*, vol. 7, pp. 419—433, 1957.

814. T a k á c s, L.: On a Coincidence Problem Concerning Telephone Traffic, *Acta Math.*, vol. 9, pp. 1—2, 1958.

815. T a k á c s, L.: On a Combined Waiting Time and Loss Problem Concerning Telephone Traffic, *Ann. Budapest Sect. Math.*, vol. 1, 1958.

816. T a k á c s. L.: On a General Probability Theorem and Its Applications in the Theory of the Stochastic Processes, reprint from *Proc, Cambridge Phil. Soc.*, vol. 54, pt. 2, pp. 219—224, 1958.

817. T a k á c s, L.: Some Probability Question in the Theory of Telephone Traffic, *Magyar Tudományos Akad. Mat. Fiz. Oszl. Közleményei*, vol. 8, pp. 151—210. 1958 (Hungarian).

Русский перевод: Т а к а ч Л. Некоторые вероятностные задачи в телефонии. Сб. переводов «Математика», 4 : 6, 1960, с. 93—144.

818. T a k á c s, L.: On a Sojourn Time Problem in the Theory of Stochastic Processes, *Trans. Am. Math. Soc.*, vol. 93, no. 3, pp. 531—540, December, 1959.

819. T a k á c s, L.: On the Limiting Distribution of the Number of Coincidences Concerning Telephone Traffic, *Ann. Math. Statist.*, vol. 30, no. 1, pp. 134—142, March, 1959.

820. T a k á c s, L.: On Certain Stochastic Processes Arising in the Theory of Telephone Traffic, Department of Statistics, Columbia University, New York, 1959.

821. T a k á c s, L.: On the Probability Law of the Busy Period for Two Types of Queueing Processes, *Office of Naval Research, Sponsored Research Rept.* CU-22-61-Nonr-266 (59) MS, December, 1960.

822. T a k á c s, L.: On the Transient Behaviour of Single-server Queueing Processes with Recurrent Input and Exponentially Distributed Service Time, *Operations Research.* vol. 8, no. 2, p. 231, March—April, 1960.

823. T a k á c s, L.: "Stochastic Processes, Problems and Solution", Methuen's Monographs, Methuen & Co., Ltd., London, 1960.

824. T a k á c s, L.: Transient Behavior of Single-server Queueing Process with Erlang Input, *Office of Naval Research, Sponsored Research Rept.* CU-40-60 Nonr-266 (33) MS, March, 1960.

*825. T a k á c s, L.: "Introduction to the Theory of Queues." N. Y., Oxford University Press, 1962.

826. T a c k l i n d, S.: Fourieranalytische Behandlung von Erneuerungsproblem, *Skand. Actuar Tidskr.*, vol. 28, pp. 68—105, 1945 (German).

827. T a n g e, I.: Optimum Methods for Determining Routes and Number of Lines in a Telephone Network with Alternative Traffic Facilities, *Tele.*, vol. 1, pp. 1—21, 1957 (Swedish); see also Document 19d, 6-ième et 7-ème Commissions d'etude du C. C. I. F. Genève, 1952—1954.

828. T a n n e r, J. C.: The Delay to Pedestrians Crossing a Road, *Biometrika*, vol. 38, pp. 383—392, 1951.

829. T a n n e r, J. C.: A Problem of Interference between Two Queues' *Biometrika*, vol. 40, pp. 58—69, 1953 (Summary in *Operational Research Quart.*, vol. 4, p. 30, June, 1953).

830. T a n n e r, J. C.: A Simplified Model for Delays in Overtaking on a Two-lane Road, *J. Roy. Statist. Soc.*, vol. 20, no. 2, 1958.

831. T a y l o r, A. E. N., and others: Studies of Traffic and the Conditions Which Influence It, Ninth Meeting of Permanent International Association of Road Congresses, Lisbon, *Permanent Intern. Assoc. Road Congr. Sect.* 2, *Rept.* 37, 1951.

832. T a y l o r, J., and R. R. P. J a c k s o n: An Application of the Birth and Death Process to the Provision of Spare Machines, *Operational Research Quart.*, vol. 5, pp. 95—108, 1954.

833. T a y l o r, R. J.: Queues with Non-Poisson Inputs and Service Times (Including Constant Holding Times), May 13, 1958 (mimeo. copy).

834. T e m p l e, G.: The Use of Partition Functions in Problems of Traffic Congestion, *Brit. Ministry Civil Aviation, Operations Research Sect.*, *ORS/MCA Rept.* 2, June, 1948.

835. T h o r n d y k e, F.: Applications of Poisson's Probability Summation, *Bell System Tech. J.*, vol. 5, p. 610, 1926.

836. T i t c h m a r s h, E. C.: "Theory of Functions", Oxford University Press, London, 1938.

Русский перевод: Т и т ч м а р ш Э. Теория функций. Гостехиздат, М.—Л., 1951.

837. T o c h e r, K. D.: *J. Roy. Statist Soc.*, vol. 13, no. 2, p. 181, 1951.

838. T o c h e r, K. D.: Some Unsolved Problems in Queueing, *Operational Research Quart.*, vol. 3, pp. 11—13, 1952.

839. T o f t, F. J., and H. B o o t h r o y d: A Queueing Model for Spare Coal Faces, *Operational Research Quart.*, vol. 10, no. 4, pp. 245—251, December, 1959.

840. T r a u t m a n, D. L., M. D a v i s, J. H e i l f r o n, E. C. H o, and A. R o s e n b l o o m: Analysis and Simulation of Vehicular Traffic Flow, *Univ. Calif., Inst. Transportation and Traffic Eng., Research Rept.* 20, 1954 (74 pp.).

841. T r u i t t, C. J.: Traffic Engineering Techniques for Determining Trunk Requirements in Alternate Routing Trunk Network, *Bell System Tech. J.*, vol. 33, pp. 277—302, 1954.

842. Turner, W. O.: Estimation of Requirements in Dial Telephone Central Offices, *Proc. Operations Research Conf.*, Case Institute of Technology, Cleveland, January, 1953.

843. Van Den Burg, A. R.: Waiting for Transport, *Sigma*, vol. 1, p. 43, 1955 (Dutch).

844. Van Dobben de Bruyn, M.: The Condition for Equilibrium in Automatic Telephone Traffic, *Ingenieur (Utrecht)*, vol. 59, pp. 1—12, 1947.

845. Van Dobben de Bruyn, M.: Détermination et prévision du rendement de groupes hommes—machines, *Cahiers bur. temps élémentaires*, sér. 5, no. 503—01, 1955 (9 pp.; French).

846. Van Voorhis, W. R.: Waiting Line Theory as a Management Tool, *Operations Research*, vol. 4, pp. 221—231, 1956.

847. Vaulot, A. E.: Application du calcul des probabilités a l'exploitation téléphonique, *Rev. gén. élec.*, vol. 16, pp. 411—418, 1924.

848. Vaulot, A. E.: Application du calcul des probabilités à l'exploitation téléphonique, *Ann. P. T. T.*, vol. 14, p. 138, 1925.

849. Vaulot, A. E.: Extension des formules d'Erlang au cas où les durées des conversations suivent une loi quelconque, *Rev. gén. élec.*, vol. 22, pp. 1164—1171, 1927.

850. Vaulot, A. E.: Application du calcul des probabilités à l'exploitation téléphonique, *Rev. gén. élec.*, vol. 30, pp. 173—175, 1931.

851. Vaulot, A. E.: Étude du trafic téléphonique reçu par des lignes explorées dans un ordre déterminé, *Rev. gén. élec.*, vol. 38, pp. 651—665, 1935.

852. Vaulot, A. E.: Sur l'application du calcul des probabilités à la théorie du trafic téléphpnique, *Compt. rend.*, vol. 200, pp. 1815—1818, 1935.

853. Vaulot, A. E.: Délais d'attente des appels téléphoniques, traités au hasard, *Compt. rend.*, vol. 222, pp. 268—269, 1946.

854. Vaulot, A. E.: Application du calcul des probabilités á la téléphonie, École Nationale Supérieur des Télécommunication, 1947.

855. Vaulot, A. E.: Sur la proportion d'appels perdus dans la système de téléphonie automatique où chaque organe de contrôle est commun à tous les appareils d'un groupe, *Adm. franç. P. T. T. étude* 9*RS*, Service des recherches et du contrôle techniques, November, 1947.

856. Vaulot, A. E.: Les Formules d'Erlang et leur calcul pratique, *Ann. télécommun.*, vol. 6, pp. 279—286, 1951.

857. Vaulot, A. E.: Délais d'attente des appels téléphoniques dans l'ordre inverse de leur arrivée, *Compt. rend.*, vol. 238, pp. 1188—1189, 1954.

858. Vaulot, A. E.: Délais d'attente des appels téléphoniques traités au hasard, *Ann. télécommun.*, vol. 9, pp. 9—14, 1954.

859. Vaulot, A. E.: Valeurs asymptotiques des expessions rencontrées dans les applications du calcul des probabilités á l'exploitation téléphonique, *Adm. franç P. T. T. étude* 140 *RS*, Service des recherches et du contrôle techniques.

860. Vaulot, A. E.: Sur les formules d'Erlang et leurs expessions asymptotiques, *Adm. franç. P. T. T. étude* 31 *RS*, Service des recherches et du contrôle techniques (17 pp.)

861. Ventura, E.: Sur l'utilisation des intégrales de contour dans les problèmes de stocks et de délais d'attente, *Management Sci.*, vol. 6, no. 4, p. 423, 1960.

862. Von Sydow, L.: Some Aspects on the Variations in Traffic Intensity, *Teleteknik*, pp. 58—64, 1958.

863. Wagner, H. W.: Economies of Telephone Relay Applications, *Bell System Tech. J.*, vol. 33, pp. 218—275, 1954.

864. Wallstrom, B.: Artificial Traffic Trials on a Two-stage Link System Using a Digital Computer, *Ericsson Tech.*, no. 2, 1958.

865. Walz, E.: Verlustzeiten und Wartezeiten bei mehr Maschinenbedienung, *Maschinenbau*, vol. 16, pp. 11—14, 1937.

866. Wardop, J. G.: Some Theoretical Aspects of Road Traffic Research, *Proc. Inst. Civil Engrs. (London)*, part II, vol. 1, 1952.

867. Wardop, J. G.: Traffic Capacity of Town Streets, *Roads and Eng. Construct.*, vol. 30, no. 350, pp. 39—42; no. 351, pp. 68—71, 1952.

868. Wardop, J. G., and G. Charlesworth: A Method of Estimating Speed and Flow of Traffic from a Moving Vehicle, *Proc. Inst. Civil Engrs. (London)*, pt. II, vol. 3, pp. 158—171, 1954.

869. Wardop, J. G.: The Capacity of Roads, *Operational Research Quart.*, vol. 5, pp. 14—24, 1954.

870. Waugh, W. A. O'N.: An Age-dependent Birth and Death Process, *Biometrika*, vol. 42, pts. 3—4, 1955.

871. Wax, N. (ed.): "Selected Papers on Noise and Stochastic Processes", Dover Publications, New York, 1954.

872. Webster, F. V.: Traffic Signal Settings and Expected Delay, presented at International Study Week in Traffic Engineering, Stresa, Italy, Oct. 1—5, 1956; published in *World Touring and Automobile Organization*, London, 1956.

873. Weir W. F.: Figuring Most Economical Machine Assignment, *Factory Management and Maintenance*, vol. 102, pp. 100—102, 1944.

874. Westgarth, D. R.: The Effect of Congestion in the Operation of Weaving Machines, M. Sc. Thesis, part II, London University, 1948.

875. Wheeler, R. C., Jr.: Transient Sequencing Delays as Applied to Air Traffic Control, Airborne Instrument Laboratory Inc., Mineola, N. Y., 1958.

876. White, H. C., and L. S. Christie: Queueing with Preemptive Priorities or with Breakdown, *Operations Research*, vol. 6, no. 1, pp. 79—95, 1958.

877. Whittaker, E. T., and G. N. Watson: "A Course of Modern Analysis", 4th ed., reprinted by Cambridge University Press, London, 1952.

Русский перевод: Уиттекер Э. Т. и Ватсон Дж. Н. Курс современного анализа. Изд. 2-е, ч. 1—2, Физматгиз, 1962—1963.

878. Whittaker, E. T., and G. Robinson: "The Calculus of Observations", 4th ed. Blakie & Son, Ltd., Glasgow, 1940.

Русский перевод: Уиттекер Э. Т. и Робинсон Г. Математическая обработка результатов наблюдений. Гостехиздат, М.—Л., 1933.

879. Widder, D. V.: „The Laplace Transform", Princeton University Press Princeton, N. J., 1946.

880. Wilkins, C. A.: On Two Queues in Parallel, *Biometrika*, vol. 47, pts. 1 and 2, p. 198, 1960.

881. Wilkinson, R. T.: The Reliability of Holding-time Measurements, *Bell System Tech. J.*, vol. 20, pp. 365—404, 1941.

882. Wilkinson, R. I.: Working Curves for Delayed Exponential Calls Served in Random Order, *Bell System Tech. J.*, vol. 32, pp. 360—383, 1953.

883. Wilkinson, R. I.: Theories for Toll-traffic Engineering in the U. S. A. (with an Appendix by J. Riordan), *Bell System Tech. J.*, vol. 35, pp. 421—514, 1956.

884. Wilkinson, R. I.: The Beginnings of Switching Theory in the United States (International Teletraffic Congress, Copenhagen, 1955), *Teleteknik*, vol. 1, pp. 14—31, 1957.

885. Wilkinson, R. I.: Queueing Theory and Some of Its Industrial Uses, *Natl. Conv. Trans. Am. Soc. Quality Control*, pp. 313—330, 1958.

886. Winsten, C. B.: A Stochastic Model of Traffic Congestion, The Rand Corporation, P-459A, Nov. 19, 1953.

887. Winsten, C. B.: Geometric Distributions in the Theory of Queues, *J. Roy. Statist. Soc., Ser. B.*, vol. 21, no. 1, 1959.

888. Wishart, D. M. G.: "On a Queueing Problem", Princeton University, Department of Mathematics, Princeton, N. J.,1953.

889. Wishart, D. M. G.: A Queueing System with χ^2 Service-time Distribution, *Ann. Math. Statist.*, vol. 27, 1956.

890. Wishart, D. M. G.: A Queueing System with Service Time Distribution of Mixed Chi-squared Type, *Operations Research*, vol. 7, no. 2, pp. 174—179, March—April, 1959.

891. Wishart, D. M. G.: Queueing Systems in Which the Discipline is „Last-come, First-served", *Operations Research*, vol. 8, no. 5, pp. 591—599, 1960.

892. Wolfowitz, J., and J. Kiefer: On the Characteristics of the General Queueing Process with Applications to Random Walks, *Ann. Math. Statist.*, vol. 27, pp. 147—161, 1956.

893. W r i g h t, E. P. G.: Behavior of Telephone Exchange Traffic Where Non-equivalent Choice Outlets Are Commoned, *Elec. Commun.*, vol. 24, pp. 42—54, 1947.

894. W r i g h t, E. P. G., and J. R i c e: Probability Studies Applied to Telecommunications Systems with Storage, *Elec. Commun.*, vol. 34, 1956.

895. W r i g h t, W. R., W. G. D u v a l l, and H. A. F r e e m a n: Machine Interference: Two Solutions of Problem Raised by Multiple Machine Units, *Mech. Eng.*, vol. 58, pp. 510—514, 1936.

896. Z i e r m a n n, M.: Ergänzung des Lagervorrates, II. Nachbestellung, *Magyar Tudományos Akad. Alkalm. Mat. Int. Közleményei,* vol. 2, pp. 203—216, 1953 (Hungarian; Russian and German summaries).

*897. Z i t e k, F.: Заметка к одной теореме Королюка. «Чехосл. математический журнал», 8 (83), 1958, с. 448—459 (на русск. яз.).

*898. Z i t e k, F.: К теории ординарных потоков. «Чехосл. математический журнал», 7 (82), 1957, с. 318—319 (на русск. яз.).

*899. А з л а р о в Т. А. Обобщение одной теоремы А. Я. Хинчина. Труды Ташкентского университета, вып. 189, 1961, с. 113—118.

*900. А л и е в Г. А. Моделирование на ЭВМ работы проектируемой автоматической линии. Труды Вычисл. центра АН Азерб. ССР, с. 2, 1963, с. 14—16.

*901. А л и е в Г. А. Моделирование одной задачи массового обслуживания. «Доклады АН Азерб. ССР», т. 19, № 3, 1963, с. 3—6.

*902. А л и е в Г. А. Некоторые вопросы многофазных систем массового обслуживания. Кандидатская диссертация. МГУ, 1964.

*903. А л и е в Г. А., Б у с л е н к о Н. П., К л и м о в Г. П., Н а з а р е н к о А. И. Моделирование производственного процесса автоматизированного стана печной сварки труб. Сб. «Проблемы кибернетики», вып. 9, 1963, с. 211—240.

*904. А л и е в Г. А., К л и м о в Г. П. Принципы построения математической модели задачи теории массового обслуживания. (На примере моделирования технологического процесса участка холодной отделки автоматизированного стана печной сварки труб.) Труды Вычисл. центра АН Азерб. ССР, т. 1, 1962, с. 28—37.

*905. А л и е в Г. А., К л и м о в Г. П. Моделирование производственных процессов на ЭВМ. Сб. «Кибернетику — на службу коммунизму», т. 2, изд-во «Энергия», М.—Л., 1964, с. 354—364.

*906. Б а б и ц к и й И. А. Теория телефонного сообщения. Трансжелдориздат, 1937.

*907. Б а б и ц к и й И. А. Формула ступенчатого включения. «Вестник связи», вып. Техника связи, № 7, 1949, с. 15—17.

*908. Б а б и ц к и й И. А. О работе ступенчатого включения. «Вестник связи», № 10, 1954, с. 8—10.

*909. Б а б и ц к и й И. А. К расчету ступенчатого включения на АТС. Связьиздат, 1956.

*910. Б а з и л е в и ч К. В., Г о в о р к о в В. А. Трафик и работа приборов соединения автоматических телефонных станций. Связьтехиздат, 1933.

*911. Б а л о н и ш н и к о в М. С., Т и т о в Н. И. Методы решения задач теории массового обслуживания на аналоговых вычислительных машинах. IV Всесоюзная конференция-семинар по теории и методам матем. моделирования. Тезисы докладов. Киев, 1964, с. 215.

*912. Б а ш а р и н Г. П. Финальные вероятности многомерного марковского процесса, описывающего действие некоторых двухкаскадных телефонных систем с отказами. «Теория вероятностей и ее применения», т. 3, вып. 4, 1958, с. 452—458.

913. Б а ш а р и н Г. П. Теоретико-вероятностное исследование двухкаскадной телефонной системы с отказами, работающей в режиме свободного искания. «Доклады АН СССР», т. 121, № 1, 1958, с. 101—104.

914. Б а ш а р и н Г. П. О многомерном предельном распределении чисел занятых линий в коммутаторах второго каскада телефонной системы с отказами. «Доклады АН СССР», т. 121, № 2, 1958, с. 280—283.

*915. Б а ш а р и н Г. П. О применении многомерной локальной предельной теоремы к вычислению вероятностей состояний двухкаскадных телефонных систем с отказами. Сб. «Проблемы передачи информации», вып. 4, 1959, с. 19—26.

*916. Б а ш а р и н Г. П. О выводе систем уравнений состояния для двухкаскадных телефонных систем с потерями. «Электросвязь», № 1, 1960, с. 56—64.
*917. Б а ш а р и н Г. П. О новом приближенном методе вычисления вероятностей потерь в двухкаскадных схемах. «Электросвязь», № 9, 1960, с. 52—63.
*918. Б а ш а р и н Г. П. О предельном распределении времени занятости полнодоступного пучка линий. «Теория вероятностей и ее применения», т. 5, вып. 2, 1960, с. 246—252.
*919. Б а ш а р и н Г. П. Об аналитическом определении и методах вычисления вероятностей потерь в коммутационных схемах. Сб. «Проблемы передачи информации», вып. 9, 1961, с. 5—47.
*920. Б а ш а р и н Г. П. О точном и приближенном методах вычисления вероятностей потерь в двухкаскадных схемах (резюме). Труды VI Всесоюзного совещания по теории вероятностей и математической статистике 1960 г., Вильнюс, 1962.
*921. Б а ш а р и н Г. П. О статистических оценках вероятности потерь и других характеристик коммутационных систем. «Электросвязь», № 3, 1962, с. 41—48.
*922. Б а ш а р и н Г. П. Таблицы вероятностей и средних квадратических отклонений потерь на полнодоступном пучке линий. Изд. АН СССР, 1962.
*923. Б а ш а р и н Г. П. О коммутационных системах в режиме группового искания. «Электросвязь», № 2, 1963, с. 58—67.
*924. Б а ш а р и н Г. П. О сложных системах массового обслуживания с несколькими конечными очередями и нетерпеливыми заявками. Сб. «Кибернетику — на службу коммунизму», т. 2. Изд-во «Энергия», М.—Л., 1964, с. 274—302.
*925. Б а ш а р и н Г. П., Ш в а л ь б В. П. О моделировании действия коммутационных систем методом Монте-Карло на ЭЦВМ. «Известия АН СССР», ОТН, Энергетика и автоматика, № 3, 1962, с. 143—153.
*926. Б а ш а р и н Г. П., Ш н е п с М. А. Обзор некоторых новейших работ в области теории телефонного сообщения. «Электросвязь», 1963, № 5, с. 41—48, № 6, с. 43—48.
*927. Б а ш а р и н о в А. Е., Ф л е й ш м а н Б. С. Некоторые кибернетические задачи статистического различения информационных потоков. Труды VI Всесоюзного совещания по теории вероятностей и математической статистике 1960 г., Вильнюс, 1962, с. 195—199.
*928. Б е л я е в Ю. К. Линейчатые марковские процессы и их приложение к задачам теории надежности. Труды VI Всесоюзного совещания по теории вероятностей и математической статистике 1960 г., Вильнюс, 1962, с. 309—323.
*929. Б е л я е в Ю. К. Предельные теоремы для редеющих потоков. «Теория вероятностей и ее применения», т. 8, вып. 2, 1963, с. 175—184.
*930. Б е л я е в Ю. К. Производительность при наличии двух типов отказов. Сб. «Кибернетику — на службу коммунизму», т. 2. Изд-во «Энергия», М.—Л., 1964, с. 303—309.
*931. Б л ю м е н ф е л ь д В. Н. О формулах Эрланга для произвольного закона распределения длительности занятий. Сб. трудов НИИТС, т. 3, 1959, с. 76—88.
*932. Б л ю м е н ф е л ь д В. Н. Потери в телефонных системах с отказами для потока вызовов с ограниченным последействием. Сб. трудов НИИТС, № 5, 1959, с. 80—87.
*933. Б о л о т и н В. А. Расчет двухкаскадных схем свободного искания. Сб. трудов НИИТС, № 12, 1963, с. 73—104.
*934. Б о л ь ш е в Л. Н. Сравнение интенсивностей простейших потоков (резюме доклада). «Теория вероятностей и ее применения», т. 7, вып. 3, 1962, с. 352—354.
*935. Б о р о в к о в А. А. О дискретных системах массового обслуживания. «Теория вероятностей и ее применения», т. 8, вып. 3, 1963, с. 251—263.
936. Б р о д и С. М. Об интегро-дифференциальном уравнении для систем с τ-ожиданием. «Доклады АН УССР», № 6, 1959, с. 571—573 (на укр. яз.; резюме на русск. яз.).
*937. Б р о д и С. М. Об одной задаче теории массового обслуживания. Труды Всес. совещания по теории вероятностей и математической статистике 1958 г., Ереван, 1960, с. 143—147.

*938. Броди С. М. Однолинейная система с τ-ожиданием в случае эрланговского входящего потока, «Доклады АН УССР», № 11, 1962, с. 1425—1428 (на укр. яз., резюме на русск. яз.).

*939. Броди С. М. Задачи с ограничениями и некоторые предельные теоремы для однолинейных систем массового обслуживания. Кандидатская диссертация. АН УССР, Киев, 1962.

*940. Броди С. М. Об одной предельной теореме теории массового обслуживания. «Украинский математический журнал», т. 15, № 1, 1963, с. 76—79.

*941. Бронштейн О. И. О вероятности приема сообщений телеконтроля в случае, когда аппаратура диспетчерского пункта может выходить из строя и восстанавливаться. «Автоматика и телемеханика», т. 24, № 9, 1963, с. 1267—1271.

*942. Булинская Е. В. Некоторые задачи оптимального управления запасами. «Теория вероятностей и ее применения», т. 9, вып. 3, 1964, с. 431—447.

*943. Булинская Е. В. Стационарное решение в задачах оптимального управления запасами. «Теория вероятностей и ее применения», т. 9, вып. 3, 1964, с. 556—560.

*944. Бусленко Н. П. Решение задач теории массового обслуживания методом моделирования на электронных цифровых вычислительных машинах. Сб. «Проблемы передачи информации», вып. 9, 1961, с. 48—69.

*945. Бусленко Н. П. О суперпозиции стационарных ординарных потоков с ограниченным последействием. Сб. «Проблемы передачи информации», вып. 9, 1961, с. 79—82.

*946. Бусленко Н. П. О решении методом Монте-Карло задач, связанных с массовым обслуживанием. Труды IV Всесоюзного математического съезда, т. 2. Изд-во «Наука», Л., 1964, с. 328—331.

*947. Бусленко Н. П. О решении методом Монте-Карло задач, связанных с массовым обслуживанием. Сб. «Кибернетику — на службу коммунизму», т. 2. Изд-во «Энергия», М.—Л., 1964, с. 365—368.

*948. Бусленко Н. П., Голенко Д. И., Соболь И. М., Срагович В. Г., Шрейдер Ю. А. Метод статистических испытаний (метод Монте-Карло). Физматгиз, 1962.

*949. Бусленко Н. П., Шрейдер Ю. А. Метод статистических испытаний Монте-Карло и его реализация на цифровых вычислительных машинах. Физматгиз, 1961.

*950. Бухман Е. Н. К проблеме скученности. «Прикладная математика и механика», т. 3, № 4, 1939, с. 189—194.

*951. Бухман Е. Н. О надбавках при дроблении на группы потока трафика АТС. «Электросвязь», № 5, 1941, с. 69—75.

952. Бухман Е. Н. Проблемы скученности в телефонии. Сб. «Прикладная математика и механика», т. 6, вып. 2—3, 1942, с. 247—256.

953. Бухман Е. Н. Задача о времени ожидания. Сб. «Прикладная математика и механика», т. 11, вып. 4, 1947, с. 475—484.

*954. Бухман Е. Н., Подгорецкий И. А. Статистические методы расчета средств связи. Сб. «Статистика связи». Связьиздат, 1948, с. 242—289.

*955. Бухман Е. Н. Проблемы расчета средств обслуживания в условиях неравномерной нагрузки. Докторская диссертация. Моск. экон.-стат. ин-т, 1948.

*956. Бухман Е. Н., Кузнецова М. В. Статистическое изучение телефонной нагрузки при проектировании автоматических телефонных станций. «Ученые записки по статистике», т. II. Изд-во АН СССР, 1956, с. 263—270.

*957. Васильев П. И. Стационарный поток с ограниченным последействием. «Ученые записки Кишиневского университета», т. 50, 1962, с. 123—138.

*958. Васильев П. И., Коваленко И. Н. Замечание о стационарных потоках однородных событий. «Украинский математический журнал», т. 16, № 3, 1964, с. 374—375.

*959. Вдовин А. А., Швальб В. П. Исследование действия коммутационных схем в режиме группового искания методом статистических испытаний на ЭЦВМ. Сб. «Проблемы передачи информации», вып. 11, 1962, с. 77—85.

*960. Вентцель Е. С. Обобщение уравнений и формул Эрланга на случай системы массового обслуживания смешанного типа с ограниченным временем ожидания. «Морской сборник», № 1, 1961, с. 43—49.

*961. Вентцель Е. С. Теория вероятностей, изд. 2. Физматгиз, 1962. Гл. 19. «Элементы теории массового обслуживания», с. 504—550.

*962. Висков О. В. Некоторые предельные теоремы теории массового обслуживания. Кандидатская диссертация, Математический институт им. В. А. Стеклова, 1964.

*963. Висков О. В. Две асимптотические формулы теории массового обслуживания. «Теория вероятностей и ее применения», т. 9, вып. 1, 1964, с. 177—178.

*964. Висков С. В., Прохоров Ю. В. Вероятность потери вызова при большой интенсивности потока. «Теория вероятностей и ее применения», т. 9, вып. 1, 1964, с. 99—104.

965. Владзиевский А. П. Вероятностный закон работы и внутренние запасы автоматических линий. «Автоматика и телемеханика», т. 13, № 4, 1952, с. 227—281.

*966. Волконский В. А. Эргодическая теорема для распределения длины выброса. «Теория вероятностей и ее применения», т. 5, вып. 3, 1960. с. 357—360.

967. Вольберг О. А. Задача об ожидании. «Известия Военной Электротехнической академии РККА», т. 17, 1939, с. 93—124.

968. Вольберг О. А. Задача о стационарной и нестационарной очередях. «Доклады АН СССР», т. 24, № 7, 1939, с. 656—661.

*969. Генин Л. С. Расчет числа каналов для немедленной системы эксплуатации каналов междугородной связи. «Вестник связи», вып. Электросвязь, № 7, 1949, с. 15—17.

*970. Георгалин Р. А., Гусарев В. С. К постановке задач анализа структурных схем автоматических линий методами теории вероятностей. «Научные записки Одесского политехнического института», т. 35, 1961, с. 32—42.

*971. Гнеденко Б. В. Вычисление среднего перехода между станками. Иваново. «Бюллетень ИВНИТИ», № 1—2, 1934, с. 117—122.

*972. Гнеденко Б. В. Вычисление среднего перехода между станками. Иваново. «Бюллетень ИВНИТИ», № 11—12, 1934, с. 118—122.

*973. Гнеденко Б. В. О среднем простое станков при многостаночной работе. «Известия хлопчатобумажной промышленности», № 11, 1934, с. 15—18.

974. Гнеденко Б. В. К теории счетчиков Гейгера—Мюллера. «Журнал экспериментальной и теоретической физики», т. 11, вып. 1, 1941, с. 101—106.

*975. Гнеденко Б. В. Курс теории вероятностей, изд. 3-е. Гостехиздат, 1962.

*976. Гнеденко Б. В. О некоторых задачах теории вероятностей. «Украинский математический журнал», т. 9, № 4, 1957, с. 377—388.

*977. Гнеденко Б. В. Об одной задаче массового обслуживания. «Доклады АН УССР», 1958, с. 477—479 (на укр. яз.; резюме на русск. яз.).

*978. Гнеденко Б. В. Об одном обобщении формул Эрланга. «Доклады АН УССР», № 4, 1959, с. 347—350 (на укр. яз.; резюме на русск. яз.).

979. Гнеденко Б. В. Несколько замечаний к двум работам Д. И. Баррера. *Buletinul institutuli din Iasi*, т. 5, № 1—2, 1959, с. 111—118.

*980. Гнеденко Б. В. Об одной задаче теории массового обслуживания. *Trans. 2nd Prague Conf., Liblice*, 1959. Прага, 1960, с. 177—183.

981. Гнеденко Б. В. Über einige Aspecte der Entwicklung der Theory der Warteschlangen. *Mathematik Technik Wirtschaft*, H. 3, 1960, S. 162—166 (на нем. яз.).

*982. Гнеденко Б. В. О некоторых задачах теории массового обслуживания. Труды Всесоюзного совещания по теории вероятностей и математической статистике 1958 г., Ереван, 1960, с. 15—24.

*983. Гнеденко Б. В. Теория вероятностей и некоторые вопросы ее применения. «Морской сборник», № 9, 1961, с. 31—41.

*984. Гнеденко Б. В. Замечание к статье С. И. Петухова «Решение одной задачи теории массового обслуживания». «Морской сборник», № 2, 1962, с. 44—47.

*985. Гнеденко Б. В. О некоторых постановках задач и результатах теории массового обслуживания. В кн.: Хинчин А. Я. «Работы по математической теории массового обслуживания», Физматгиз, 1963, с. 221—233.

*986. Гнеденко Б. В. О теории массового обслуживания. «Математика в школе», № 3, 1964, с. 10—20.

*987. Гнеденко Б. В. О ненагруженном дублировании. «Известия АН СССР», Техническая кибернетика, № 4, 1964, с. 3—12.

*988. Гнеденко Б. В. О дублировании с восстановлением. «Известия АН СССР». Техническая кибернетика, № 5, 1964, с. 111—118.

*989. Гнеденко Б. В., Беляев Ю. К., Коваленко И. Н. Основные направления исследований в теории массового обслуживания. Труды VI Всесоюзного совещания по теории вероятностей и математической статистике 1960 г., Вильнюс, 1962, с. 341—355.

*990. Гнеденко Б. В., Зубков М. Н. Об определении оптимального числа причалов. «Морской сборник», № 6, 1964, с. 30—39.

*991. Гнеденко Б. В., Колмогоров А. Н. Предельные распределения сумм независимых случайных переменных. Гостехиздат, М.—Л., 1949.

*992. Годлевский В. А. Распределение времени ожидания при случайном обслуживании вызовов. Сб. научных трудов ЦНИИИС, вып. 2, 1949, с. 127—138.

*993. Григелионис Б. Об одной предельной теореме теории восстановления. «Лит. математический сборник», т. 2, № 1, 1962, с. 25—34.

*994. Григелионис Б. И. О центральной предельной теореме для сумм процессов восстановления. «Лит. математический сборник», т. 4, № 2, 1964, с. 197—201.

*995. Григелионис Б. И. Предельные теоремы для сумм процессов восстановления. Сб. «Кибернетику — на службу коммунизму», т. 2. Изд-во «Энергия», М.—Л., 1964, с. 246—265.

*996. Гринберг Э. Я., Шнепс М. А. О некоторых свойствах потерянного потока телефонного сообщения. «Ученые записки Латв. университета», № 47, 1963, с. 253—260.

*997. Гу Цзи-фа. Обслуживающие системы с диспетчером. Кандидатская диссертация. Математический институт им. В. А. Стеклова, 1963.

*998. Добрушин Р. Л. О законе Пуассона для распределения частиц в пространстве. «Украинский математический журнал», т. 8, № 2, 1956, с. 127—134.

*999. Думлер С. А. Составление календарного графика загрузки оборудования на электронной вычислительной машине. «Механизация и автоматизация производства», № 4, 1961, с. 53—57.

*1000. Дынкин Е. Б. Об одной задаче из теории вероятностей. «Успехи математических наук», т. 4, вып. 5 (33), 1949, с. 183—197.

*1001. Ежов И. И. Об одной задаче Такача. «Вестник Киевского университета», серия матем. и механ., № 5, вып. 2, 1962, с. 161—163 (на укр. яз.; резюме на русск. яз.).

1002. Зайончковский Д. А. Étude sur la question du nombre des circuits dans le service automatique, C. C. I. F., Documents 20, 8th C. E., 1955.

*1003. Зубова А. Ф. О холодном дублировании с восстановлением при любом законе распределения потока отказов и времени восстановления. «Известия АН СССР», Техническая кибернетика, № 5, 1964, с. 107—110.

*1004. Ивакин-Тревогин Г. Н. О решении системы уравнений Хинчина. «Известия АН Узб. ССР», серия физ.-матем. наук, № 2, 1960, с. 55—60.

*1005. Ицхоки Я. С. Вероятность n-зацепления хаотически следующих импульсов случайной длительности и распределение длительности их зацепления. «Радиотехника и электроника», т. 7, № 1, 1962, с. 16—24.

*1006. Климов Г. П. Моделирование на электронных цифровых машинах некоторого класса систем массового обслуживания. «Журнал вычислительной математики и математической физики», т. 1, № 5, 1961, с. 935—940.

*1007. Климов Г. П. Об экстремальных потоках в теории массового обслуживания. «Успехи математических наук», т. 17, № 5, 1961, с. 193—195.

*1008. Климов Г. П. Экстремальные задачи в теории массового обслуживания. Сб. «Кибернетику — на службу коммунизму», т. 2, Изд-во «Энергия», М.—Л., 1964, с. 310—324.

*1009. Климов Г. П., Алиев Г. А. Решение на вычислительных машинах одной задачи теории массового обслуживания методом Монте-Карло. «Журнал вычислительной математики и математической физики», т. 1, № 5, 1961, с. 933—935.

*1010. Коваленко И. Н. Определение корреляционных функций некоторых процессов, связанных с задачами обслуживания. «Доклады АН УССР», № 5, 1958, с. 480—481 (на укр. яз.; резюме на русск. яз.).

*1011. Коваленко И. Н. Исследование многолинейной системы массового обслуживания с очередью и ограниченным временем пребывания в системе. «Украинский математический журнал», т. 12, № 4, 1960, с. 471—476.

*1012. Коваленко И. Н. Некоторые задачи массового обслуживания с ограничением. «Теория вероятностей и ее применения», т. 6, вып. 2, 1961, с. 222—228.

*1013. Коваленко И. Н. Об одном методе в теории массового обслуживания. Труды VI Всесоюзного совещания по теории вероятностей и математической статистике. 1960 г., Вильнюс, 1962, с. 357—358.

*1014. Коваленко И. Н. Sur la condition pour que, en régime stationnaire, la distribution soit indépendante des lois des dureés de conversation. *Annales des télécommunications*, t. 17, № 7—8, 1960, p. 190—191.

*1015. Коваленко И. Н. Об условии независимости вероятностей состояний системы обслуживания от вида распределения времени обслуживания. Сб. «Проблемы передачи информации», вып. 11, АН СССР, 1962, с. 147—151.

*1016. Коваленко И. Н. Некоторые аналитические методы в теории массового обслуживания. Сб. «Кибернетику — на службу коммунизму», т. 2, Изд-во «Энергия», М.—Л., 1964, с. 325—337.

*1017. Коваленко И. Н. Некоторые вопросы теории надежности сложных систем. Сб. «Кибернетику — на службу коммунизму», т. 2, Изд-во «Энергия», М.—Л., 1964, с. 194—205.

1018. Колмогоров А. Н. Sur le problème d'attente, „Математический сборник", № 38, вып. 1—2, 1931, с. 47—50 (на фр. яз.).

1019. Колмогоров А. Н. Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zustanden. „Математический сборник", т. 1 (43), № 4, 1936, s. 607—610 (на нем. яз.).

*1020. Ланин М. И. О вычислении вероятности потери сообщения в системах централизованного контроля. «Автоматика и телемеханика», т. 23, № 3, 1962, с. 321—330.

*1021. Лебедева Э. Н. Решение одной системы бесконечных линейных уравнений. «Известия АН Узб. ССР», серия физико-математических наук, № 2, 1959, с. 7—13.

*1022. Лезерсон В. К. Основы теории и расчета неполнодоступного пучка. Кандидатская диссертация. Московский электротехнический институт, 1950.

*1023. Лившиц Б. С. Потери. Сб. трудов НИИТС, № 3, 1959, с. 64—75.

*1024. Лившиц Б. С. Определение телефонной нагрузки при объединении и разъединении телефонных потоков. «Электросвязь», № 10, 1959, с. 61—70.

*1025. Лившиц Б. С. Методы расчета телефонной нагрузки и потерь. Сб. трудов НИИТС, № 7, 1960.

*1026. Лившиц Б. С. Методы расчета телефонной нагрузки и потерь. Докторская диссертация, НИИТС, Л., 1960.

*1027. Лившиц Б. С. Расчет потерь в схемах ступенчатых включений. Сб. трудов НИИТС, № 11, 1963, с. 54—78.

*1028. Лившиц Б. С., Блюменфельд В. Н., Парилов В. П. Современное состояние теории телефонного сообщения. Сб. трудов НИИТС, № 3, 1959, с. 5—64.

*1029. Лившиц Б. С., Блюменфельд В. Н., Фидлин Я. В. Терминология теории телефонного сообщения. Сб. трудов НИИТС, № 9, 1962, с. 5—18.

*1030. Лившиц Б. С., Родзянко В. Е. Расчет неравнодоступного звеньевого включения пучка линий. Сб. трудов НИИТС, № 5, 1959, с. 12—26.

*1031. Лившиц Б. С., Фидлин Я. В. Средняя величина потерь на телефонных сетях. «Электросвязь», № 7, 1962, с. 46—55.

*1032. Лившиц Б. С., Фидлин Я. В. Суммарные потери нескольких ступеней искания. «Электросвязь», № 12, 1962, с. 43—55.

*1033. Лившиц Б. С., Фидлин Я. В. Поток с простым последействием и его воздействие на схемы обслуживания. Сб. трудов НИИТС, № 11, 1963, с. 3—53.

*1034. Линковский Г. Б. Об использовании экспериментальных данных в расчете надежности радиоэлектронной аппаратуры, основанной на распределении Пуассона. «Известия вузов», Приборостроение, т. 4, № 5, 1961, с. 43—46.

*1035. Ляпунов А. А., Фандюшина С. М. К вопросу о повторяемости землетрясений. «Известия АН СССР», серия географическая и геофизическая, т. 6, 1950, с. 547—552.

*1036. Максименко В. Ф., Ляшенко И. М. Один алгоритм составления некоторых схем массового обслуживания. Сб. «Обчислювальна математика і техника», АН УССР, Киев, 1963, с. 66—76 (на укр. яз.; резюме на русск. яз.).

*1037. Маргулис Х. Ш. О «потоке пар». Сб. «Кибернетику — на службу коммунизму», т. 2, Изд-во «Энергия», М.—Л., 1964, с. 266—273.

*1038. Мархай Е. В., Рогинский В. Н., Харкевич А. Д. Автоматическая телефония. Связьиздат, 1960.

*1039. Марьянович Т. П. Обобщение формул Эрланга на случай, когда приборы могут выходить из строя и восстанавливаться. «Украинский математический журнал», т. 12, № 3, 1960, с. 279—286.

*1040. Марьянович Т. П. Надежность системы со смешанным резервом. «Доклады АН УССР», № 8, 1961, с. 964—997 (на укр. яз.; резюме на русск. яз.).

*1041. Марьянович Т. П. Обслуживание с учетом выхода приборов из строя (резюме). Труды VI Всесоюзного совещания по теории вероятностей и математической статистике 1960 г., Вильнюс, 1962, с. 363—364.

*1042. Марьянович Т. П. Однолинейная система массового обслуживания с ненадежным прибором. «Украинский математический журнал», т. 14, № 4, 1962, с. 417—422.

*1043. Марьянович Т. П. Некоторые вероятностные задачи теории надежности систем. Кандидатская диссертация, АН УССР, Киев, 1963.

*1044. Медведев Г. А. О помехоустойчивости приемника с конечным временем восстановления. Труды СФТИ при Томском университете, 1961, вып. 40, с. 29—38.

*1045. Медведев Г. А. Помехоустойчивость приемника с конечным временем восстановления. Случай простейшего потока мешающих сигналов. «Известия вузов», Радиотехника, 1962, т. 5, № 2, с. 200—207.

*1046. Медвидь М. В. К теории автоматического ориентирования деталей в бункерных устройствах. «Научные записки Львовского политехн. института», Автоматизация в машиностроении, вып. XV, изд. Львовского университета, 1958, с. 40—66.

*1047. Медвидь М. В. К расчету основных размеров бункерных разгрузочных устройств. Там же, с. 67—92.

*1048. Медвидь М. В. Теоретические основы проектирования вибрационных бункерных загрузочных устройств. Сб. «Автоматизация в машиностроении», № 2. Изд. Львовского политехн. института, 1959, с. 35—62.

*1049. Мельников К. П. Возможности решения задач звеньевых включений на электронных вычислительных цифровых машинах. Сб. трудов НИИТС, № 10, 1962, с. 35—47.

*1050. Мельников К. П. Метод моделирования телефонной нагрузки на универсальных вычислительных цифровах машинах (УЦВМ). Сб. трудов НИИТС, № 11, 1963, с. 79—88.

*1051. Мудров В. И. К вопросу об определении вероятности отказа в однолинейных системах массового обслуживания смешанного типа. Сб. «Проблемы кибернетики», вып. 4, 1960, с. 45—52.

*1052. Мудров В. И. Очередь с «нетерпеливыми» клиентами и переменным временем обслуживания, линейно зависящим от времени пребывания клиента в очереди. Сб. «Проблемы кибернетики», вып. 5, 1961, с. 283—285.

*1053. Мудров В. И. Математика и решение задач организации обслуживания. «Вестник ПВО», № 8, 1964, с. 72—76.

*1054. Насирова Т. И. Об одном обобщении задачи Эрланга. Труды Вычислительного центра АН Азерб. ССР, т. 2, 1963, с. 3—13.

*1055. Насирова Т. И. Некоторые задачи теории массового обслуживания. Кандидатская диссертация. Вычислительный центр АН Азерб. ССР, Баку, 1964.

*1056. Нечепуренко М. И., Хайлов И. К. Об одной задаче обслуживания. Труды МФТИ, вып. 9, 1962, с. 105—111.

*1057. Новиков О. А., Висков О. В. О влиянии управления на пропускные способности системы массового обслуживания с отказами. «Морской сборник», № 9, 1964, с. 33—37.

1058. Ососков Г. А. Одна предельная теорема для потоков однородных событий. Сб. «Теория вероятностей и ее применения», т. 1, вып. 2, 1956, с. 274—282.

*1059. Ососков Г. А. Предельные теоремы для потоков однородных случайных событий. Кандидатская диссертация, МГУ, 1957.

*1060. Парилов В. П. Обслуживание телефонной нагрузки с комбинированными потерями при ограниченном числе источников нагрузки. Сб. трудов НИИТС, № 5, 1959, с. 88—97.

*1061. Парилов В. П. Исследование систем обслуживания телефонных вызовов с комбинированными потерями при равнодоступных, неравнодоступных и звеньевых включениях. Кандидатская диссертация, ЛИИЖТ, Л., 1961.

*1062. Петухов С. И. Решение одной задачи теории массового обслуживания. «Морской сборник», № 2, 1962, с. 39—44.

*1063. Погожев И. Б. Оценка отклонений потока отказов в аппаратуре многоразового использования от пуассоновского потока. Сб. «Кибернетику — на службу коммунизму», т. 2, Изд-во «Энергия». М.—Л., 1964, с. 228—245.

*1064. Прохоров Ю. В. Переходные явления в теории массового обслуживания. «Лит. математический сборник», т. 3, № 1, 1963, с. 199—206.

*1065. Прохоров Ю. В. Предельные явления в процессах массового обслуживания. Резюме доклада, прочитанного на Всесоюзном коллоквиуме по предельным теоремам теории вероятностей. Фергана, 17—22 сент. 1962 г. Сб. «Предельные теоремы теории вероятностей», Ташкент, 1963, с. 150.

*1066. Псарев С. А. Теория расчета приборов управления (маркеров). Сб. трудов НИИТС, № 5, 1959, с. 33—49.

*1067. Псарев С. А. Вероятность ожидания при полнодоступном блокируемом включении приборов. Сб. трудов НИИТС, № 12, 1963, с. 32—38.

*1068. Родзянко В. Е., Вишнякова М. К. Приближенная формула расчета потерь в двухзвеньевом включении с концентрацией на звене *A* и звене *B*. Сб. трудов НИИТС, № 5, 1959, с. 5—11.

*1069. Розенберг В. Я., Прохоров А. И. Что такое теория массового обслуживания. Изд-во «Советское радио», 1962.

*1070. Россиков Б. В. Применение обходных направлений на телефонных сетях. Сб. трудов НИИТС, № 11, 1963, с. 89—110.

*1071. Самандаров Э. Г. О периоде вхождения системы в стационарный режим работы. «Теория вероятностей и ее применения», т. 8, вып. 3, 1963, с. 123.

*1072. Самандаров Э. Г. Очереди при неординарном потоке поступления вызовов. «Предельные теоремы теории вероятностей». Ташкент, 1963, с. 108—117.

*1073. Самандаров Э. Г. О периоде занятости системы обслуживания при групповом входящем потоке требований. «Доклады АН Тадж. ССР», т. 6, № 4, 1963, с. 3—7.

*1074. Самандаров Э. Г. Исследование некоторых задач массового обслуживания аналитическими методами. Кандидатская диссертация. АН Тадж. ССР, Душанбе, 1963.

*1075. Самандаров Э. Г. Системы обслуживания в условиях большой нагрузки. «Теория вероятностей и ее применения», т. 8, вып. 3, 1963, с. 327—330.

*1076. Самандаров Э. Г. Стационарное распределение в некоторых системах массового обслуживания. «Доклады АН Тадж. ССР», т. 6, № 3, 1963, с. 7—10.

*1077. Сванидзе Г. Г. Применение метода Монте-Карло к задачам регулирования речного стока. «Доклады АН СССР», 1961, т. 140, № 6, с. 1394—1396.

1078. Севастьянов Б. А. Эргодическая теорема для марковских процессов и ее приложение к телефонным линиям с отказами. «Теория вероятностей и ее применения», т. 2, вып. 1, 1957, с. 106—116.

*1079. Севастьянов Б. А. Формулы Эрланга в телефонии при произвольном законе распределения длительности разговора. Труды III Всесоюзного математического съезда 1956 г., т. 4, АН СССР, 1959, с. 68—70.

*1080. Севастьянов Б. А. Задача о влиянии емкости бункеров на среднее время простоя автоматической линии станков. «Теория вероятностей и ее применения», т. 7, вып. 4, 1962, с. 438—447.

*1081. Сливняк И. М. Некоторые свойства стационарных потоков однородных случайных событий. «Теория вероятностей и ее применения», т. 7, вып. 3, 1962, с. 347—352.

*1082. Соловьев А. Д. Об определении резервов для систем многократного действия. «Известия АН СССР», ОТН, Энергетика и автоматика, № 2, 1962, с. 124—129.

*1083. Соловьев А. Д. Надежность системы с восстановлением. Сб. «Кибернетику — на службу коммунизму», т. 2. Изд-во «Энергия», М.—Л., 1964, с. 189—193.

*1084. Туманян. С. Х Об одной схеме массового обслуживания. Труды VI Всесоюзного совещания по теории вероятностей и математической статистике 1960 г., Вильнюс, 1962, с. 367—369.

*1085. Фидлин Я. В. Расчет потерь в звеньевых включениях при осуществлении маркером ограниченного числа попыток установить соединение. Сб. «Проблемы передачи информации», вып. 11, 1962, с. 124—132.

*1086. Фидрих И. Описание одного из алгоритмов решения задач теории массового обслуживания методом статистических испытаний. «Доклады АН СССР», т. 153, № 4, 1963, с. 779—782.

*1087. Фидрих И. Один из алгоритмов решения задач теории массового обслуживания методом статистических испытаний. Кандидатская диссертация, ЛГУ, 1964.

*1088. Флейшман Б. С. О переходном режиме загрузки памяти кибернетического устройства статистического воспроизводства образов. «Радиотехника и электроника», т. 6, № 7, 1961, с. 1041—1048.

*1089. Харкевич А. Д. Метод приближенной оценки пропускной способности двухзвеньевой коммутационной системы. «Доклады АН СССР», т. 114, № 2, 1957, с. 308—309.

1090. Хинчин А. Я. Математическая теория стационарной очереди. «Математический сборник», т. 39, 1932, с. 73—84.

1091. Хинчин А. Я. О среднем времени простоя станков. «Математический сборник», т. 40, 1933, с. 119—123.

*1092. Хинчин А. Я. Korrelationstheorie der stationären stochastischen Prozesse, *Math. Ann.*, Bd. 109, 1934, S. 604—605.

1093. Хинчин А. Я. Математические методы теории массового обслуживания. Труды математического института им. В. А. Стеклова, т. 49, 1935.

1094. Хинчин А. Я. О пуассоновских потоках без последействия. «Теория вероятностей и ее применения», т. 1, вып. 1, 1956, с. 3—18.

1095. Хинчин А. Я. О пуассоновских потоках случайных событий. «Теория вероятностей и ее применения», т. 1, вып. 3, 1956, с. 320—327.

1096. Хинчин А. Я. Mathematical Methods in the Theory of Queueing, Charles Griffin & Co., Ltd., London, 1960.

*1097. Хинчин А. Я. О формулах Эрланга в теории массового обслуживания. Сб. «Теория вероятностей и ее применения», т. 7, вып. 3, 1962, с. 330—335.

*1098. Хинчин А. Я. Работы по математической теории массового обслуживания. Физматгиз, 1963.

*1099. Шахбазов А. А. Об одной задаче обслуживания неординарного потока требований. «Доклады АН СССР», т. 145, № 2, 1962, с. 289—292.

*1100. Шахбазов А. А. О некоторых задачах теории массового обслуживания (резюме доклада). «Теория вероятностей и ее применения», т. 7, вып. 4, 1962, с. 475—477.

*1101. Шахбазов А. А. Обслуживание приборами разной производительности. «Ученые записки Азерб. университета», серия физико-математических и химических наук, № 3, 1962, с. 107—113.

*1102. Шахбазов А. А. О некоторых задачах теории массового обслуживания. Кандидатская диссертация, МГУ, 1962.

*1103. Шахбазов А. А., Самандаров Э. Г. Об обслуживании неординарного потока. Сб. «Кибернетику — на службу коммунизму», т. 2. Изд-во «Энергия», М.—Л., 1964, с. 338—353.

*1104. Швальб В. П. О рекуррентных формулах для вычисления дисперсии времени занятости полнодоступного пучка. Сб. «Проблемы передачи информации», вып. 9, 1961, с. 83—86.

*1105. Швальб В. П. О матрице вторых моментов распределения времени пребывания для многолинейной системы с ограниченным числом ожидающих. Сб. «Проблемы передачи информации», вып. 11, 1962, с. 110—116.

*1106. Шнепс М. А. Изучение однокаскадных неполнодоступных телефонных систем на электронной вычислительной машине. Сб. «Проблемы передачи информации», вып. 12, 1963, с. 109—123.

*1107. Шнепс М. А. О применении цепей Маркова для изучения телефонных систем с потерями. Сб. «Проблемы передачи информации», вып. 12, 1963, с. 124—134.

*1108. Шнепс М. А. О применении метода вложенных цепей Маркова к моделированию систем массового обслуживания. «Ученые записки Латв. университета», № 47, 1963, с. 261—266.

*1109. Шор Я. Б. Статистические методы анализа и контроля качества и надежности. Изд-во «Советское радио», 1962.

*1110. Яровицкий Н. В. О выходящем потоке однолинейной системы обслуживания с потерями. «Доклады АН УССР», № 1, 1961, с. 1251—1254 (на укр. яз.; резюме на русск. яз.).

*1111. Яровицкий Н. В. О векторных односвязно-зависимых потоках. «Доклады АН УССР», № 6, 1962, с. 715—719 (на укр. яз.; резюме на русск. яз.).

*1112. Яровицкий Н. В. Выходящие потоки и многоэтапное обслуживание. Кандидатская диссертация, АН УССР, Киев, 1962.

*1113. Яровицкий Н. В. О некоторых свойствах односвязно-зависимых потоков. «Украинский математический журнал», т. 14, № 2, 1962, с. 170—179.

*1114. Яровицкий Н. В. О задаче многоэтапного обслуживания с потерями. «Украинский математический журнал», т. 16, № 3, 1964, с. 421—427.

*1115. Ярошенко В. М. Об одной задаче теории массового обслуживания. «Доклады АН УССР», 1962, № 2, с. 153—156 (на укр. яз.; резюме на русск. яз.).

Предметный указатель

Оглавление

ТОМАС Л. СААТИ

ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ
И ЕЁ ПРИЛОЖЕНИЯ

Редактор *И. М. Волкова*
Техн. редактор *В. В. Беляева*
Переплет художника *В. М. Позднякова*

Сдано в набор 21/X 1964 г. Подп. к печати 24/II 1965 г.
Формат $60\times90^1/_{16}$. Объём 32 печ. л. Уч.-изд. л. 33,869.
Тираж 5000 экз. Заказ 2076. Цена в пер. № 5 2 р. 47 к.
Темплан 1964 г. № 61

Ленинградская типография № 6
Главполиграфпрома Государственного комитета
Совета Министров СССР по печати,
Ленинград, ул. Салтыкова-Щедрина, 54

Цена 2 р. 47 к.