МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Пензенский государственный университет» (ПГУ)

Т. В. Черушева, Н. В. Зверовщикова

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ

Учебное пособие

Пенза
Издательство ПГУ
2021

УДК 519.872+519.85
Ч-50

Р е ц е н з е н т ы :

кандидат технических наук,
заведующий кафедрой «Прикладная информатика»
Пензенского государственного технологического университета
А. П. Ремонтов;

кандидат технических наук,
доцент кафедры «Информационные вычислительные системы»
Пензенского государственного университета
А. П. Писарев

Черушева, Татьяна Вячеславовна.
Ч-50 Теория массового обслуживания : учеб. пособие / Т. В. Черушева, Н. В. Зверовщикова. – Пенза : Изд-во ПГУ, 2021. – 224 с.

ISBN 978-5-907456-66-2

В издании, состоящем из двух глав и посвященном изучению теории случайных процессов, подробно изложены основные модели теории массового обслуживания и методы их исследования. В первой главе вводятся потоки, их преобразования и основные свойства. Вторая глава посвящена одноканальным и многоканальным системам массового обслуживания с накопителем и без накопителя. Системы рассматриваются как марковские процессы. Приводится описание системы, кодирование состояний случайного процесса, строятся размеченные графы переходов случайного процесса и диаграммы функционирования системы. Составляются матрицы интенсивностей переходов, системы уравнений и идет расчет характеристик с анализом свойств. Анализ математических моделей и дискретных систем различных классов ведется с использованием аналитических, численных и имитационных методов. Аналитические методы исследования основаны на аппарате теории массового обслуживания, численные – на аппарате марковских случайных процессов, статистические – на методах имитационного моделирования. Также рассматриваются системы с приоритетами. Материал пособия сопровождается большим количеством примеров и соответствует программе курса «Теория массового обслуживания».

Предназначено для обучающихся по направлению подготовки 01.03.04 «Прикладная математика», профиль подготовки «Математическое моделирование в экономике и технике».

УДК 519.872+519.85

ISBN 978-5-907456-66-2

© Пензенский государственный
университет, 2021

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Предмет, цель и задачи теории массового обслуживания

Во многих областях производства, бытового обслуживания, экономики и финансов важную роль играют системы специального вида, реализующие многократное выполнение однотипных задач. Подобные системы называют *системами массового обслуживания* (СМО). В качестве примеров СМО в финансово-экономической сфере можно привести системы, представляющие собой банки, страховые организации, налоговые инспекции, аудиторские службы. В сфере производства и обслуживания примерами СМО могут служить: различные системы связи (в том числе телефонные станции), погрузочно-разгрузочные комплексы (порты, товарные станции), автозаправочные станции, магазины, парикмахерские, билетные кассы, пункты обмена валюты, ремонтные мастерские, больницы и т.д. Такие системы, как компьютерные сети, системы сбора, хранения и обработки информации, транспортные системы, автоматизированные производственные участки, а в военной области системы противовоздушной или противоракетной обороны, также могут рассматриваться как своеобразные СМО.

Заявкой (или ***требованием***) называется спрос на удовлетворение какой-либо потребности (далее потребности предполагаются однотипными). Выполнение заявки называется ***обслуживанием*** заявки.

Предмет теории массового обслуживания – установление зависимости между характером потока заявок, производительностью отдельного канала, числом каналов и успешностью (эффективностью) обслуживания.

Цель теории массового обслуживания – выработка рекомендаций по рациональному построению СМО, рациональной организации их работы и регулированию потока заявок для обеспечения высокой эффективности функционирования СМО.

Для достижения этой цели ставятся ***задачи теории массового обслуживания***, состоящие в установлении зависимостей эффективности функционирования СМО от ее организации (параметров): характера потока заявок, числа каналов и их производительности и правил работы СМО.

В качестве характеристик эффективности функционирования СМО можно выбрать три основные группы (обычно средних) показателей:

1. Показатели эффективности использования СМО:

1.1. Абсолютная пропускная способность СМО – среднее число заявок, которое сможет обслужить СМО в единицу времени.

1.2. Относительная пропускная способность СМО – отношение среднего числа заявок, обслуживаемых СМО в единицу времени, к среднему числу поступивших за это же время заявок.

1.3. Средняя продолжительность периода занятости СМО.

1.4. Коэффициент использования СМО – средняя доля времени, в течение которого СМО занята обслуживанием заявок, и т.п.

Время T_0, оставшееся до завершения обслуживания заявки, находящейся в приборе, от момента поступления некоторой заявки в систему, и учитывающее, что на момент поступления в системе может и не оказаться заявок, т.е. учитывающее простои системы, называется ***временем дообслуживания***. Математическое ожидание этого времени рассчитывается по формуле

$$M[T_0] = \lambda b^2 (1 + v_b^2),$$

где λ – интенсивность *простейшего* потока заявок, поступающих в систему; v_b – коэффициент вариации длительности обслуживания.

2. Показатели качества обслуживания заявок:

2.1. Среднее время ожидания заявки в очереди.

2.2. Среднее время пребывания ($b = t_{\overline{\text{об}}}$) заявки в СМО (часто длительность обслуживания заявок предполагается распределенной по *экспоненциальному закону*, что существенно упрощает аналитические выкладки. Это обусловлено тем, что процессы, протекающие в системах с экспоненциальным распределением интервалов времени, являются *марковскими*). Величина, обратная средней длительности обслуживания b, характеризует среднее число заявок, которое может быть обслужено за единицу времени, и называется ***интенсивностью обслуживания***: $\mu = \dfrac{1}{b}$.

2.3. Вероятность отказа заявке в обслуживании без ожидания.

2.4. Вероятность того, что вновь поступившая заявка немедленно будет принята к обслуживанию.

2.5. Закон распределения времени ожидания заявки в очереди.

2.6. Закон распределения времени пребывания заявки в СМО.

2.7. Среднее число заявок, находящихся в очереди.

2.8. Среднее число заявок, находящихся в СМО, и т.п.:

– абсолютная пропускная способность системы (A или λ'), т.е. среднее число заявок, обслуживаемых в единицу времени;

– относительная пропускная способность (Q или λ''), т.е. средняя доля поступивших заявок, обслуживаемых системой.

3. *Показатели эффективности функционирования пары «СМО – клиент»,* где под «клиентом» понимают всю совокупность заявок или некий их источник. К числу таких показателей относится, например, средний доход, приносимый СМО в единицу времени, и т.п.

Системой массового обслуживания называется любая система для выполнения заявок, поступающих в нее в случайные моменты времени.

Поступление заявки в СМО называется ***событием***. Последовательность событий, заключающихся в поступлении заявок в СМО, называется ***входящим потоком заявок***.

Последовательность событий, заключающихся в выполнении заявок в СМО, называется ***выходящим потоком заявок***.

Случайный процесс, протекающий в системе массового обслуживания, состоит в том, что система в случайные моменты времени переходит из одного состояния в другое: меняется число занятых каналов, число заявок, стоящих в очереди, и т. п. Система массового обслуживания представляет собой физическую систему дискретного типа с конечным (или счетным) множеством состояний, а переход системы из одного состояния в другое происходит скачком, в момент, когда осуществляется какое-то событие (приход новой заявки, освобождение канала, уход заявки из очереди и т. п.).

Классификация систем массового обслуживания

Системы массового обслуживания делятся на типы (или классы) по ряду признаков.

По числу каналов СМО подразделяют на *одноканальные* (когда имеется один канал обслуживания) и *многоканальные*, точнее n-канальные (когда количество каналов $n \geq 2$). Здесь и далее будем полагать, что каждый канал одновременно может обслуживать только одну заявку и, если не оговорено специально, каждая находящаяся на обслуживании заявка обслуживается только одним каналом. Многоканальные СМО могут состоять из однородных каналов либо из разнородных, отличающихся длительностью обслуживания одной заявки. Практически время обслуживания каналом одной заявки $T_{об} = b$ является непрерывной случайной величиной. Однако при условии абсолютной однородности поступающих заявок и каналов время обслуживания может быть и величиной постоянной ($T_{об} = b = \text{const}$).

По дисциплине обслуживания СМО подразделяют на три класса:

1. СМО *с отказами*, в которых заявка, поступившая на вход СМО в момент, когда все каналы заняты, получает «отказ» и покидает СМО («пропадает»). Чтобы эта заявка все же была обслужена, она должна снова поступить на вход СМО и рассматриваться при этом как заявка, поступившая впервые. Примером СМО с отказами может служить работа АТС: если набранный телефонный номер (заявка, поступившая на вход) занят, то заявка получает отказ, и, чтобы дозвониться по этому номеру, следует его набрать еще раз (заявка поступает на вход как новая).

2. СМО *с ожиданием* (*неограниченным ожиданием* или *очередью*). В таких системах заявка, поступившая в момент занятости всех каналов, становится в очередь и ожидает освобождения канала, который примет ее к обслуживанию. Каждая заявка, поступившая на вход, в конце концов будет обслужена. Такие СМО часто встречаются в торговле, в сфере бытового и медицинского обслуживания, на предприятиях (например, обслуживание станков бригадой наладчиков).

3. СМО *смешанного типа* (*с ограниченным ожиданием*). Это такие системы, в которых на пребывание заявки в очереди накладываются некоторые ограничения.

Эти ограничения могут накладываться на *длину очереди*, т.е. максимально возможное число заявок, которые одновременно могут находиться в очереди. В качестве примера такой системы можно привести мастерскую по ремонту автомобилей, имеющую ограниченную по размерам стоянку для неисправных машин, ожидающих ремонта.

Ограничения ожидания могут касаться *времени пребывания заявки в очереди*, по истечении которого она выходит из очереди и покидает систему, либо касаться *общего времени пребывания заявки* в СМО (т.е. суммарного времени пребывания заявки в очереди и обслуживании).

В СМО с ожиданием и в СМО смешанного типа применяются различные схемы обслуживания заявок из очереди. Обслуживание может быть *упорядоченным*, когда заявки из очереди обслуживаются в порядке их поступления в систему, и *неупорядоченным*, при котором заявки из очереди обслуживаются в случайном порядке. Иногда применяется *обслуживание с приоритетом*, когда некоторые заявки из очереди считаются приоритетными и поэтому обслуживаются в первую очередь.

По ограничению потока заявок СМО делятся на *замкнутые* и *открытые*.

Если поток заявок ограничен и заявки, покинувшие систему, могут в нее возвращаться, то СМО является *замкнутой*, в противном случае – *открытой*. Классическим примером замкнутой СМО служит работа бригады наладчиков в цеху. Станки являются источниками заявок на обслуживание, и их количество ограничено, наладчики – каналы обслуживания. После проведения ремонтных работ вышедший из строя станок снова становится источником заявок на обслуживание. В открытой СМО характеристики потока заявок не зависят от того, в каком состоянии сама СМО (сколько каналов занято). В замкнутой СМО – зависят. Так, в рассмотренном выше примере интенсивность потока «заявок» со стороны станков (т.е. количество заявок в единицу времени) зависит от того, сколько их неисправно и ждет наладки.

По количеству этапов обслуживания СМО делятся на *однофазные* и *многофазные* системы.

Если каналы СМО однородны, т.е. выполняют одну и ту же операцию обслуживания, то такие СМО называются *однофазными*. Если каналы обслуживания расположены последовательно и они неоднородны, так как выполняют различные операции обслуживания (т.е. обслуживание состоит из нескольких последовательных этапов или фаз), то СМО называется *многофазной*. Примером работы многофазной СМО является обслуживание автомобилей на станции технического обслуживания (мойка, диагностирование и т.д.).

ГЛАВА 1

ПОТОКИ СОБЫТИЙ. ПРЕОБРАЗОВАНИЕ ПОТОКОВ

1.1. Обозначения систем массового обслуживания (символика Кендалла)

Схема СМО изображена на рис. 1.1.

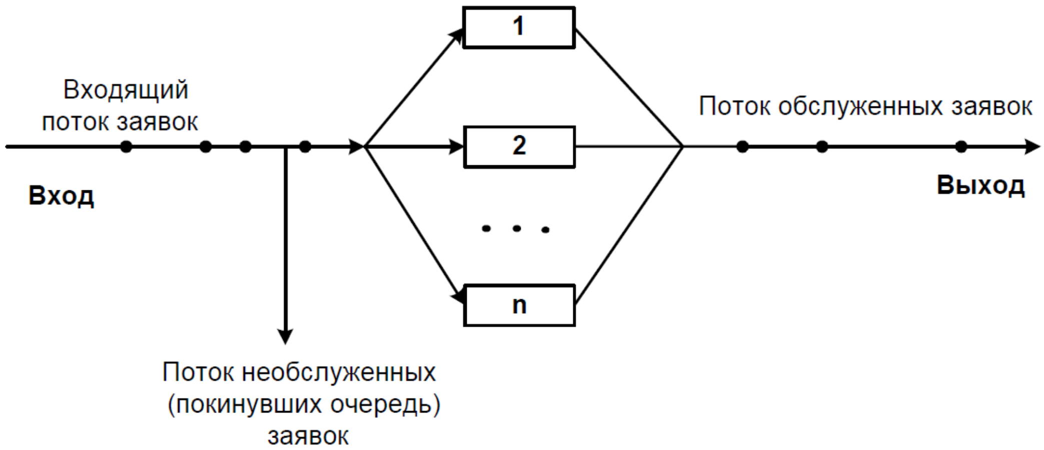


Рис. 1.1. Система массового обслуживания

Для компактного описания систем массового обслуживания часто используются обозначения, предложенные Д. Кендаллом, в виде

$$A/B/N/L,$$

где A и B задают законы распределений соответственно интервалам времени между моментами поступления заявок в систему и длительностью обслуживания заявок в приборе; N – число обслуживающих приборов в системе (N = 1, 2, ..., ∞); L – число мест в накопителе, которое может принимать значения 0, 1, 2, … (отсутствие L означает, что накопитель имеет неограниченную емкость).

Для задания законов распределений A и B используются следующие обозначения:

G (*General*) – произвольное распределение общего вида;

M (*Markovian*) – экспоненциальное (показательное) распределение;

D (*Deterministik*) – детерминированное распределение;

U (*Uniform*) – равномерное распределение;

E_k (*Erlangian*) – распределение Эрланга k-го порядка (с k последовательными одинаковыми экспоненциальными фазами, или $Э_k$);

h_k (*hipoexponential*) – гипоэкспоненциальное распределение k-го порядка (с k последовательными разными экспоненциальными фазами);

H_r (*Hiperexponential*) – гиперэкпоненциальное распределение порядка r (с r параллельными экспоненциальными фазами);

g (*gamma*) – гамма-распределение;

P (*Pareto*) – распределение Парето и т.д.

Примеры:

***M*/*M*/1** – одноканальная СМО с накопителем неограниченной емкости, в которую поступает однородный поток заявок с экспоненциальным распределением интервалов времени между последовательными заявками (простейший поток) и экспоненциальной длительностью обслуживания заявок в приборе.

***M*/*G*/3/10** – трехканальная СМО с накопителем ограниченной емкости, равной 10, в которую поступает однородный поток заявок с экспоненциальным распределением интервалов времени между последовательными заявками (простейший поток) и длительностью обслуживания заявок, распределенной по закону общего вида.

$D/E_2/7/0$ – семиканальная СМО без накопителя (емкость накопителя равна 0), в которую поступает однородный поток заявок с детерминированными интервалами времени между последовательными заявками (детерминированный поток) и длительностью обслуживания заявок в приборе, распределенной по закону Эрланга 2-го порядка.

Для обозначения более сложных СМО дополнительно могут использоваться обозначения, описывающие неоднородный поток заявок и приоритеты между заявками разных классов.

1.2. Случайные процессы с дискретными состояниями

Рассмотрим физическую систему X со счетным множеством состояний

$$x_1, x_2, \ldots, x_n.$$

В любой момент времени t система X может быть в одном из этих состояний. Обозначим $p_k(t)$ $(k = 1, 2, \ldots, n, \ldots)$ – вероятность того, что в момент t система будет находиться в состоянии x_k. Очевидно, для любого t

$$\sum_{k=1}^{\infty} p_k(t) = 1. \qquad (1.1)$$

Совокупность вероятностей $p_k(t)$ для каждого момента времени t характеризует данное сечение случайного процесса, протекающего в системе. Эта совокупность не является исчерпывающей характеристикой процесса (она, например, совсем не отражает зависимости между сечениями), но все же достаточно хорошо описывает процесс и для ряда практических применений оказывается достаточной.

Случайные процессы со счетным множеством состояний бывают двух типов: с дискретным или непрерывным временем. Первые отличаются тем, что переходы из состояния в состояние могут происходить только в строго определенные, разделенные конечными интервалами моменты времени t_1, t_2, ... Случайные процессы с непрерывным временем отличаются тем, что переход системы из состояния в состояние возможен в любой момент времени t.

В качестве примера дискретной системы X, в которой протекает случайный процесс с непрерывным временем, рассмотрим группу из n самолетов, совершающих налет на территорию противника, обороняемую истребительной авиацией. Ни момент обнаружения группы, ни моменты подъема по ней истребителей заранее не известны. Различные состояния системы соответствуют различному числу пораженных самолетов в составе группы:

x_0 – не поражено ни одного самолета,

x_1 – поражен ровно один самолет,

…………

x_k – поражено ровно k самолетов,

…………

x_n – поражены все n самолетов.

Схема возможных состояний системы и возможных переходов из состояния в состояние показана на рис. 1.2.

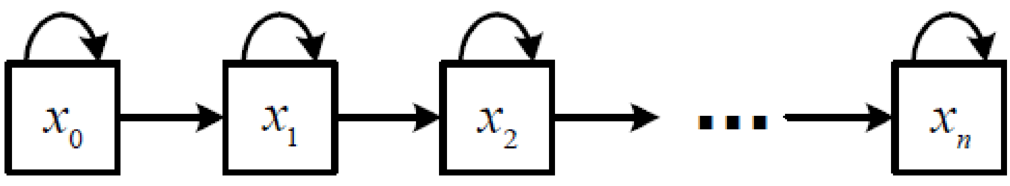


Рис. 1.2. Схема возможных состояний

Стрелками показаны возможные переходы системы из состояния в состояние. Закругленная стрелка, направленная из состояния x_k в него же, означает, что система может не только перейти в соседнее состояние, но и остаться в прежнем. Для данной системы характерны необратимые переходы (пораженные самолеты не восстанавливаются); в связи с этим из состояния x_n никакие переходы в другие состояния уже невозможны.

Отметим, что на схеме возможных переходов (см. рис. 1.2) показаны только переходы из состояния в соседнее состояние и не показаны «перескоки» через состояние: эти перескоки отброшены как практически невозможные. Действительно, для того чтобы система «перескочила» через состояние, нужно, чтобы строго одновременно были поражены два или более самолета, а вероятность такого события равна нулю.

Случайные процессы, протекающие в системах массового обслуживания, как правило, представляют собой процессы с непрерывным временем. Это связано со случайностью потока заявок. В противоположность системе с необратимыми переходами, рассмотренной в предыдущем примере, для системы массового обслуживания характерны обратимые переходы: занятый канал может освободиться, очередь может «рассосаться».

В качестве примера рассмотрим одноканальную систему массового обслуживания (например, одну телефонную линию), в которой заявка, заставшая канал занятым, не становится в очередь, а покидает систему (получает «отказ»). Это дискретная система с непрерывным временем и двумя возможными состояниями:

x_0 – канал свободен,

x_1 – канал занят.

Переходы из состояния в состояние обратимы. Схема возможных переходов показана на рис. 1.3.

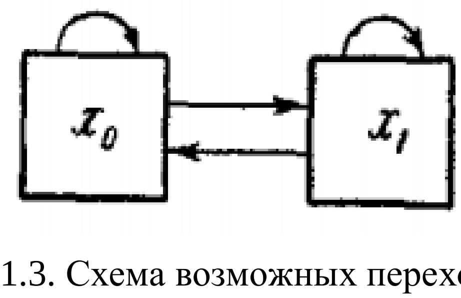


Рис. 1.3. Схема возможных переходов

Для n-канальной системы такого же типа схема возможных переходов показана на рис. 1.4.

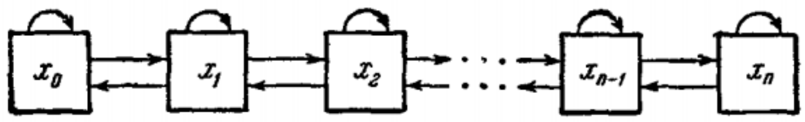


Рис. 1.4. Схема возможных переходов для *n*-канальной системы

Состояние x_0 – все каналы свободны; x_1 – занят ровно один канал, x_2 – занято ровно два канала и т.д.

Рассмотрим еще один пример дискретной системы с непрерывным временем: одноканальную систему массового обслуживания, которая может находиться в четырех состояниях:

x_0 – канал исправен и свободен,

x_1 – канал исправен и занят,

x_2 – канал неисправен и ждет ремонта,

x_3 – канал неисправен и ремонтируется.

Схема возможных переходов для этого случая показана на рис. 1.5.

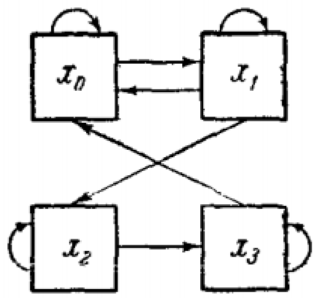


Рис. 1.5. Схема возможных переходов

Переход системы из x_3 непосредственно в x_1, минуя x_0, можно считать практически невозможным, так как для этого нужно, чтобы окончание ремонта и приход очередной заявки произошли строго в один и тот же момент времени.

Для того чтобы описать случайный процесс, протекающий в дискретной системе с непрерывным временем, прежде всего нужно проанализировать причины, вызывающие переход системы из состояния в состояние. Для системы массового обслуживания основным фактором, обусловливающим протекающие в ней процессы, является поток заявок. Поэтому математическое описание любой системы массового обслуживания начинается с описания потока заявок.

Под потоком событий в теории вероятностей понимается последовательность событий, происходящих одно за другим в какие-то моменты времени. События, образующие поток, в общем случае могут быть различными, но мы будем рассматривать лишь поток однородных событий, различающихся только моментами появления. Такой поток можно изобразить как последовательность точек $t_1, t_2, \ldots, t_k, \ldots$ на числовой оси (рис. 1.6), соответствующих моментам появления событий.

Рис. 1.6. Поток однородных событий

Поток событий называется ***регулярным***, если события следуют одно за другим через строго определенные промежутки времени. Такой поток сравнительно редко встречается в реальных системах, но представляет интерес как предельный случай. Типичным для системы массового обслуживания является случайный поток заявок. Для полного описания регулярного потока заявок достаточно задать интенсивность потока λ или значение интервала $t = 1/\lambda$.

Рассмотрим потоки событий, обладающие некоторыми особенно простыми свойствами. Для этого введем ряд определений.

Поток заявок называется ***простейшим***, если он удовлетворяет следующим условиям:

1) ***отсутствие последействия***, т.е. заявки поступают независимо друг от друга;

2) ***стационарность***, т.е. вероятность поступления данного числа заявок на любом временнóм отрезке $[t_1, t_2]$ зависит лишь от величины этого отрезка и не зависит от значения t_1, что позволяет говорить о среднем числе заявок за единицу времени λ, называемом ***интенсивностью потока заявок***;

3) ***ординарность***, т.е. в любой момент времени в СМО поступает лишь одна заявка, а поступление одновременно двух и более заявок пренебрежимо мало.

Интенсивность и параметр потока

Пусть $P_1(t_0, t)$ – вероятность того, что интервале $[t_0, t]$ наступит ровно одно событие.

Величина $\lim\limits_{t \to t_0} \dfrac{P_1(t_0, t)}{t - t_0} = \lambda(t_0)$ называется параметром потока.

Обозначим через $m(t_0, t)$ среднее число событий, наступивших на интервале $[t_0, t]$. Величина $\lim\limits_{t \to t_0} \frac{m(t_0, t)}{t - t_0} = \mu(t_0)$ называется интенсивностью потока.

Для ординарного потока интенсивность и параметр потока совпадают.

В дальнейшем для ординарного потока не будем делать разницы между интенсивностью потока и его параметром. Заметим лишь, что для неординарных потоков $\lambda(t) \le \mu(t)$, так как $m(t_0, t) \ge P_1(t_0, t)$.

Для простейшего потока вероятность $p_i(t)$ поступления в СМО ровно i заявок за время t вычисляется по формуле

$$p_i(t) = \frac{(\lambda i)^t}{i!} e^{-\lambda t}, \tag{1.2}$$

т.е. вероятности распределены по закону Пуассона с параметром λt. По этой причине простейший поток называется также ***пуассоновским потоком***.

Прежде чем выводить формулы, введем обозначения и сделаем некоторые следствия из ординарности. Пусть $P_k(t_0, t)$ есть вероятность того, что на интервале $[t_0, t]$ наступит ровно k событий. Тогда, в силу определения интенсивности, функция $\lambda(t) = \lim\limits_{\Delta t \to 0} \frac{P_1(t, t + \Delta t)}{\Delta t}$ будет интенсивностью пуассоновского потока. Это же соотношение можно записать как

$$P_1(t, t + \Delta t) = \lambda(t)\Delta t + o(\Delta t);$$

сила ординарности потока:

$$\lim_{\Delta t \to 0} \frac{P_k(t, t + \Delta t)}{\Delta t} = 0 \text{ при } k \ge 2,$$

и поэтому при $k \ge 2$

$$P_k(t, t + \Delta t) = o(\Delta t).$$

Так как выполнено условие нормировки

$$P_0(t, t + \Delta t) + P_1(t, t + \Delta t) + P_{>1}(t, t + \Delta t) = 1,$$

то

$$P_0(t, t + \Delta t) = 1 - P_1(t, t + \Delta t) - P_{>1}(t, t + \Delta t) = 1 - \lambda(t)\Delta t + \mathrm{o}(\Delta t).$$

Рассмотрим подробнее условия 1–3, посмотрим, чему они соответствуют для потока заявок и за счет чего они могут нарушаться.

1. Условие отсутствия последействия – наиболее существенное для простейшего потока – означает, что заявки поступают в систему независимо друг от друга. Например, поток пассажиров, входящих на станцию метро, можно считать потоком без последействия потому что причины, обусловившие приход отдельного пассажира именно в тот, а не другой момент, как правило, не связаны с аналогичными причинами для других пассажиров. Однако условие отсутствия последействия может быть легко нарушено за счет появления такой зависимости. Например, поток пассажиров, покидающих станцию метро, уже не может считаться потоком без последействия, так как моменты выхода пассажиров, прибывших одним и тем же поездом, зависимы между собой.

Вообще нужно заметить, что выходной поток (или поток обслуженных заявок), покидающий систему массового обслуживания, обычно имеет последействие, даже если входной поток его не имеет. Чтобы убедиться в этом, рассмотрим одноканальную систему массового обслуживания, для которой время обслуживания одной заявки вполне определено и равно $t_{об}$. Тогда в потоке обслуженных заявок минимальный интервал времени между заявками, покидающими систему, будет равен $t_{об}$. Нетрудно убедиться, что наличие такого минимального интервала неизбежно приводит к последействию. Действительно, пусть стало известно, что в какой-то момент t_1 систему покинула обслуженная заявка. Тогда можно утверждать с достоверностью, что на любом участке времени τ, лежащем в пределах $(t_1, t_1 + t_{об})$, обслуженной заявки не появится; значит, будет иметь место зависимость между числами событий на неперекрывающихся участках.

Последействие, присущее выходному потоку, необходимо учитывать, если этот поток является входным для какой-либо другой СМО («многофазовое обслуживание», когда одна и та же заявка последовательно переходит из системы в систему).

Отметим, между прочим, что самый простой на первый взгляд регулярный поток, в котором события отделены друг от друга равными интервалами, отнюдь не является «простейшим» в нашем смысле слова, так как в нем имеется ярко выраженное последействие: моменты появления следующих друг за другом событий связаны жесткой, функциональной зависимостью. Именно из-за наличия последействия анализ процессов, протекающих в СМО при регулярном потоке заявок, гораздо сложнее, чем при простейшем.

2. Условию стационарности удовлетворяет поток заявок, вероятностные характеристики которого не зависят от времени. В частности, для стационарного потока характерна постоянная плотность (среднее число заявок в единицу времени). На практике часто встречаются потоки заявок, которые (по крайней мере, на ограниченном отрезке времени) могут рассматриваться как стационарные. Например, поток вызовов на городской телефонной станции на участке времени от 12 до 13 часов может считаться стационарным. Тот же поток в течение целых суток уже не может считаться стационарным (ночью плотность вызовов значительно меньше, чем днем). Заметим, что так обстоит дело и со всеми физическими процессами, которые мы называем «стационарными»: в действительности все они стационарны лишь на ограниченном участке времени, а распространение этого участка до бесконечности – лишь удобный прием, применяемый в целях упрощения анализа. Во многих задачах теории массового обслуживания представляет интерес анализ работы системы при постоянных условиях; тогда задача решается для стационарного потока заявок.

3. Условие ординарности означает, что заявки приходят поодиночке, а не парами, тройками и т.д. Например, поток атак, которому подвергается воздушная цель в зоне действия истребительной авиации, будет ординарным, если истребители атакуют цель поодиночке, и не будет ординарным, если истребители идут в атаку парами. Поток клиентов, входящих в парикмахерскую, может считаться практически ординарным, чего нельзя сказать о потоке клиентов, направляющихся в ЗАГС для регистрации брака.

Если в неординарном потоке заявки поступают только парами, только тройками и т.д., то неординарный поток легко свести к ординарному; для этого достаточно вместо потока отдельных заявок рассмотреть поток пар, троек и т.д. Сложнее будет, если каждая заявка случайным образом может оказаться двойной, тройной и т.д. Тогда уже приходится иметь дело с потоком не однородных, а разнородных событий.

Простейший поток играет среди потоков событий особую роль, аналогичную роли нормального закона среди других законов распределения. При суммировании (взаимном наложении) большого числа ординарных, стационарных потоков с практически любым последействием получается поток, сколь угодно близкий к простейшему. Условия, которые должны для этого соблюдаться, аналогичны условиям центральной предельной теоремы, а именно: складываемые потоки должны оказывать на сумму приблизительно равномерно малое влияние: при наложении (суперпозиции) достаточного большого числа n независимых, стационарных и ординарных потоков (сравнимых между собой по интенсивностям λ_i (i = 1, 2, ..., n)) получается поток, близкий к простейшему с интенсивностью λ, равной сумме интенсивностей входящих потоков, т.е. $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$.

Проиллюстрируем сказанное элементарными рассуждениями. Пусть имеется ряд независимых потоков $П_1$, $П_2$, …, $П_n$ (рис. 1.7). «Суммирование» потоков состоит в том, что все моменты появления событий сносятся на одну и ту же ось $0t$, как показано на рис. 1.7.

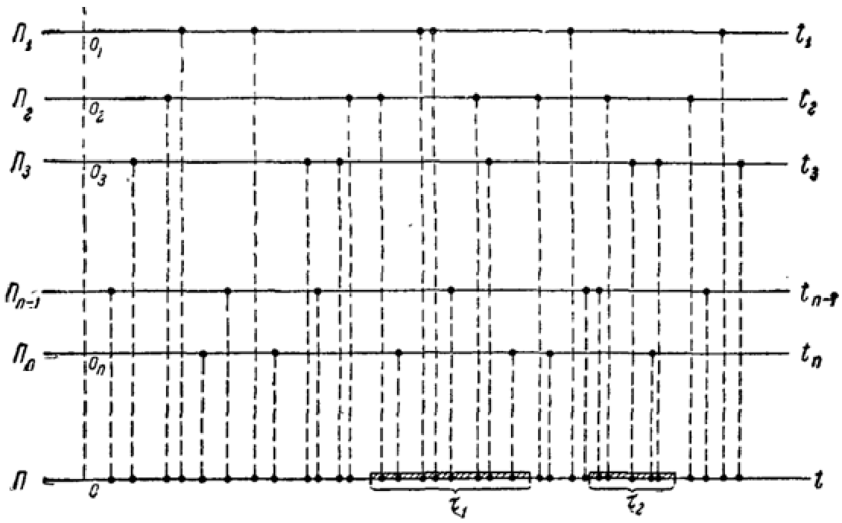


Рис. 1.7. Суммирование» потоков

Предположим, что потоки $П_1$, $П_2$,… сравнимы по своему влиянию на суммарный поток (т.е. имеют плотности одного порядка), а число их достаточно велико. Предположим, кроме того, что эти потоки стационарны и ординарны, но каждый из них может иметь последействие, и рассмотрим суммарный поток

$$П = \sum_{k=1}^{n} П_k \tag{1.4}$$

на оси $0t$ (рис. 1.7). Очевидно, что поток П должен быть стационарным и ординарным, так как каждое слагаемое обладает этим свойством и они независимы. Кроме того, при увеличении числа слагаемых последействие в суммарном потоке, даже если оно значительно в отдельных потоках, должно постепенно слабеть. Действительно, рассмотрим на оси $0t$ два неперекрывающихся отрезка τ_1 и τ_2 (рис. 1.7). Каждая из точек, попадающих в эти отрезки, случайным образом может оказаться принадлежащей тому или иному потоку, и по мере увеличения n удельный вес точек, принадлежащих одному и тому же потоку (и, значит, зависимых), должен уменьшаться, а остальные точки принадлежат разным потокам и появляются на отрезках $[\tau_1,\tau_2]$ независимо друг от друга. При увеличении n суммарный поток будет терять последействие и приближаться к простейшему. На практике оказывается обычно достаточно сложить 4–5 потоков, чтобы получить поток, с которым можно оперировать как с простейшим.

Простейший поток играет в теории массового обслуживания особенно важную роль. Во-первых, простейшие и близкие к простейшим потоки заявок часто встречаются на практике. Во-вторых, даже при потоке заявок, отличающемся от простейшего, часто можно получить удовлетворительные по точности результаты, заменив поток любой структуры простейшим с той же плотностью. Поэтому займемся подробнее простейшим потоком и его свойствами.

1.3. Простейший поток

Рассмотрим на оси $0t$ простейший поток событий П (рис. 1.8) как неограниченную последовательность случайных точек.

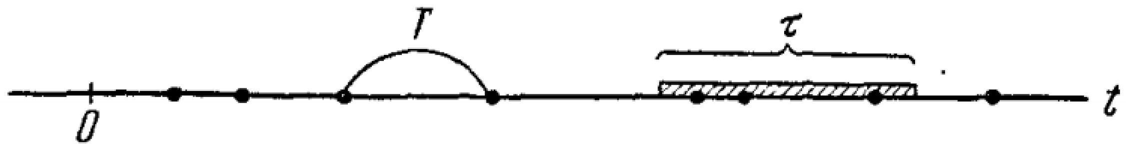


Рис. 1.8. Простейший поток

Выделим произвольный участок времени длиной τ. При условиях 1, 2 и 3 (стационарность, отсутствие последействия и ординарность) число точек, попадающих на участок τ, распределено по закону Пуассона с математическим ожиданием

$$a = \lambda\tau, \tag{1.5}$$

где λ – плотность потока (интенсивность потока заявок); $a = 1/\lambda$ определяет *средний интервал времени между двумя последовательными заявками*.

Вероятность того, что за время τ произойдет ровно m событий, равна

$$P_m(\tau) = \frac{(\lambda\tau)^m}{m!} e^{-\lambda t}. \qquad (1.6)$$

В частности, вероятность того, что участок окажется пустым (не произойдет ни одного события), будет равна

$$P_0(\tau) = e^{-\lambda t}. \qquad (1.7)$$

Пример 1. На автоматическую телефонную станцию поступает простейший поток вызовов с интенсивностью $\lambda = 1{,}2$ вызовов в 1 мин.

Найти вероятность того, что за 2 мин:

а) не придет ни одного вызова;

б) придет ровно один вызов;

в) придет хотя бы один вызов.

Решение. а) Случайная величина X – число вызовов за 2 мин – распределена по закону Пуассона с параметром $\lambda\tau = 1{,}2 \cdot 2 = 2{,}4$. Вероятность того, что вызовов не будет ($m = 0$), по формуле (1.7):

$$P_0(2) \approx e^{-2{,}4} \approx 0{,}091.$$

б) Вероятность одного вызова ($m = 1$) по формуле (1.6):

$$P_1(2) \approx 2{,}4 \quad 0{,}091 \approx 0{,}218.$$

в) Вероятность хотя бы одного вызова:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - P_0(2) \approx 1 - 0{,}091 = 0{,}909.$$

Важной характеристикой потока является закон распределения длины промежутка между соседними событиями. Рассмотрим случайную величину T – промежуток времени между произвольными двумя соседними событиями в простейшем потоке (см. рис. 1.8) и найдем ее функцию распределения:

$$F(t) = P(T < t).$$

Перейдем к вероятности противоположного события:

$$1 - F(t) = P(T \geq t),$$

то есть вероятность того, что на участке времени длиной t, начинающемся в момент $t_{\text{об}}$ – момент появления одного из событий потока, не появится ни одного из последующих событий. Так как простейший поток не обладает последействием, то наличие в начале участка (в точке t_k) какого-то события никак не влияет на вероятность появления тех или других событий в дальнейшем. Поэтому вероятность $P(T \geq t)$ можно вычислить по формуле

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

откуда

$$F(t) = 1 - e^{-\lambda t} \quad (t > 0). \tag{1.8}$$

Дифференцируя, найдем плотность распределения:

$$f(t) = \lambda e^{-\lambda t} \quad (t > 0). \tag{1.9}$$

Закон распределения с плотностью (1.9) называется показательным законом, а величина λ – его параметром. График плотности $f(t)$ представлен на рис. 1.9.

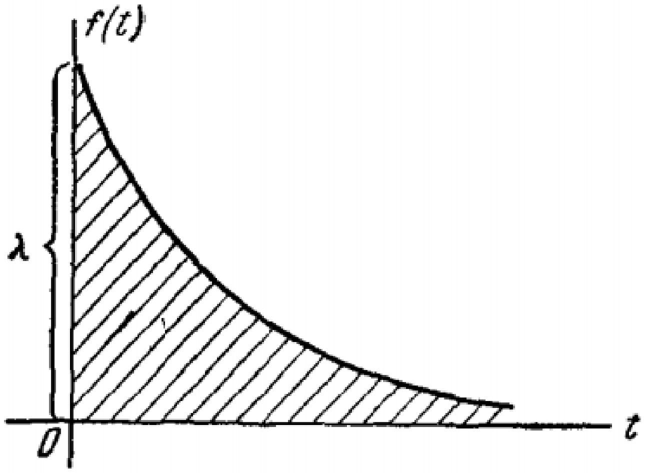


Рис. 1.9. График плотности показательного закона

Показательный закон играет большую роль в теории дискретных случайных процессов с непрерывным временем. Поэтому рассмотрим его подробнее.

Найдем математическое ожидание величины T, распределенной по показательному закону:

$$m_t = M\left[T\right] = \int_0^\infty tf\left(t\right)dt = \lambda \int_0^\infty te^{-\lambda t}dt.$$

Интегрируя по частям, получим

$$m_t = \frac{1}{\lambda}. \tag{1.10}$$

Дисперсия величины T равна

$$D_t = D(T) = \int_0^\infty t^2 f(t)\,dt - \frac{1}{\lambda^2} = \lambda \int_0^\infty t^2 e^{-\lambda t} dt - \frac{1}{\lambda^2},$$

откуда

$$D_t = \frac{1}{\lambda^2}, \tag{1.11}$$

$$\sigma_t = \frac{1}{\lambda}. \tag{1.12}$$

Докажем одно замечательное свойство показательного закона. Оно состоит в следующем: если промежуток времени, распределенный по показательному закону, уже длился некоторое время τ, то это никак не влияет на закон распределения оставшейся части промежутка: он будет таким же, как и закон распределения всего промежутка T.

Для доказательства рассмотрим случайный промежуток времени T с функцией распределения:

$$F(t) = 1 - e^{-\lambda t}, \tag{1.13}$$

и предположим, что этот промежуток уже продолжается некоторое время τ, т.е. произошло событие $T > \tau$. Найдем при этом предположении условный закон распределения оставшейся части промежутка $T_1 = T - \tau$; обозначим его $F^{(\tau)}(t)$

$$F^{(\tau)}(t) = P(T - \tau < t \mid T > \tau).$$

Докажем, что условный закон распределения $F^{(\tau)}(t)$ не зависит от τ и равен $F(t)$. Для того чтобы вычислить $F^{(\tau)}(t)$, найдем сначала вероятность произведения двух событий: $T > \tau$ и $T - \tau < t$. По теореме умножения вероятностей имеем

$$P((T > \tau)(T - \tau < t)) = P(T > \tau)P(T - \tau < t \mid T > \tau) = P(T > \tau)F^{(\tau)}(t),$$

откуда

$$F^{(\tau)}(t) = \frac{P((T > \tau)(T - \tau < t))}{P(T > \tau)};$$

событие $(T > \tau)(T - \tau < t)$ равносильно событию $(\tau < T < t + \tau)$, вероятность которого равна

$$P(\tau < T < t + \tau) = F(t + \tau) - F(t).$$

С другой стороны, $P(T > \tau) = 1 - F(\tau)$, следовательно,

$$F^{(\tau)}(t) = \frac{F(t+\tau) - F(t)}{1 - F(\tau)},$$

откуда согласно формуле (1.13) получим

$$F^{(\tau)}(t) = \frac{e^{-\lambda\tau} - e^{-\lambda(t+\tau)}}{e^{-\lambda\tau}} = 1 - e^{-\lambda\tau} = F(t),$$

что и требовалось доказать.

Таким образом, мы доказали, что если промежуток времени T распределен по показательному закону, то любые сведения о том, сколько времени уже протекал этот промежуток, не влияют на закон распределения оставшегося времени. Можно доказать, что показательный закон – единственный, обладающий таким свойством. Это свойство показательного закона представляет собой другую формулировку для «отсутствия последействия», которое является основным свойством простейшего потока.

Пример 2. В справочное бюро обращается в среднем 2 человека за 10 мин.

Найти вероятность того, что за 30 мин за справкой обратится:

а) 4 человека,

б) не менее 3 человек.

Решение. Интенсивность потока заявок равна λ = 2/10 = = 0,2 [мин$^{-1}$]. Для решения используем формулу (1.6), где полагаем $t = T$ = 30 мин.

Для пункта (а) m = 4:

$$P_4(T) = ((0,2 \cdot 30)^4 (e^{-0,2 \cdot 30}) / 4!) \approx 0,134).$$

Для пункта (б) m = 3, 4, 5, …

При решении этого пункта целесообразно использовать противоположную вероятность:

$$P_{\geq 3}(T) = 1 - P_{<3} = 1 - (P_0(T) + P_1(T) + P_2(T)) =$$
$$= 1 - (e^{-6} + 6 \cdot e^{-6} / 1! + 6^2 \cdot e^{-6} / 2!) = 1 - 0,062 = 0,938.$$

Пример 3. В приборе имеются два блока, работающих независимо друг от друга. Время безотказной работы определяется показательным законом. Среднее время безотказной работы первого блока t_1 = 2 года, второго – t_2 = 1 год.

Найти вероятность того, что за 1,5 года:

а) не откажет ни один из блоков;

б) откажет только второй блок;

в) откажут оба блока.

Решение. В качестве события выступает неисправность первого или второго блока.

Вероятность $P_m(t)$ исправности m-го блока в течение времени t определяется формулой (1.7):

$$P(T \geq t) = P_0(\tau) = e^{-\lambda t},$$

т.е.

$$p^{(1)}(T) = e^{-\lambda_1 T} = e^{-0,5 \cdot 1,5} \approx 0,472, \quad p^{(2)}(T) = e^{-\lambda_2 T} = e^{-1 \cdot 1,5} \approx 0,223,$$

где $\lambda_1 = \dfrac{1}{t_1} = 0,5$ [год$^{-1}$], $\lambda_2 = \dfrac{1}{t_2} = 1$ [год$^{-1}$].

Вероятность того, что за время T m-й блок выйдет из строя, является противоположной вероятностью $P_m(T)$:

$$\bar{p}^{(1)} = 1 - p^{(1)} = 1 - 0,472 = 0,528,$$

$$\bar{p}^{(2)} = 1 - p^{(2)} = 1 - 0,223 = 0,777.$$

Обозначим через A, B, C события, фигурирующие в пунктах (а), (б), (в) соответственно, и, учитывая, что блоки работают независимо друг от друга, найдем:

а) $p(A) = p^{(1)}(T)\, p^{(2)}(T) = 0,472 \cdot 0,223 = 0,1;$

б) $p(B) = p^{(1)}(T)\, \bar{p}^{(2)}(T) = 0,472 \cdot 0,777 = 0,367;$

в) $p(C) = \bar{p}^{(1)}(T)\, \bar{p}^{(2)}(T) = 0,528 \cdot 0,777 = 0,41.$

1.4. Нестационарный пуассоновский поток

Если поток событий нестационарен, то его основной характеристикой является мгновенная плотность $\lambda(t)$. Мгновенной плотностью потока называется предел отношения среднего числа событий, приходящегося на элементарный участок времени $(t,\ t + \Delta t)$, к длине этого участка, когда последняя стремится к нулю:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{m(t + \Delta t) - m(t)}{\Delta t} = m'(t), \tag{1.14}$$

где $m(t)$ – математическое ожидание числа событий на участке $(0, t)$.

Рассмотрим поток однородных событий, ординарный и без последействия, но не стационарный, с переменной плотностью $\lambda(t)$. Такой поток называется нестационарным пуассоновским потоком. Это первая ступень обобщения по сравнению с простейшим потоком. Для такого потока число событий, попадающих на участок длины τ, начинающийся в точке t_0, подчиняется закону Пуассона:

$$P_m(\tau, t_0) = \frac{a^m}{m!} e^{-a}, \ m = 0, 1, 2, \ldots, \tag{1.15}$$

где a – математическое ожидание числа событий на участке от t_0 до $t_0 + \tau$, равное

$$a = \int_{t_0}^{t_0+t} \lambda(t) dt. \tag{1.16}$$

Здесь величина a висит не только от длины τ участка, но и от его положения на оси Ot.

Вероятность того, что в промежутке (α, β) наступит ровно k событий данного потока, равна

$$P_k(\alpha, \beta) = e^{-[a(\beta) - a(\alpha)]} \frac{[a(\beta) - a(\alpha)]^k}{k!}.$$

Найдем для нестационарного потока закон распределения промежутка времени T между соседними событиями. Ввиду нестационарности потока этот закон будет зависеть от того, где на оси Ot расположено первое из событий. Кроме того, он будет зависеть от вида функции $\lambda(t)$. Предположим, что первое из двух соседних событий появилось в момент t_0 и найдем при этом условии закон распределения времени T между этим событием и последующим:

$$F_{t_0}(t) = P(T < t) = 1 - P(T \geq t).$$

Найдем $P(T \geq t)$ вероятность того, что на участке от t_0 до $t_0 + \tau$ не появится ни одного события:

$$P(T \geq t) = e^{-a} = \exp\left(- \int_{t_0}^{t_0+t} \lambda(t) dt \right),$$

откуда

$$F_{t_0}(t) = 1 - \exp\left(- \int\limits_{t_0}^{t_0+t} \lambda(t)dt \right). \qquad (1.17)$$

Дифференцируя, найдем плотность распределения:

$$f_{t_0}(t) = \lambda(t_0 + t)\exp\left(- \int\limits_{t_0}^{t_0+t} \lambda(t)dt \right), \quad t > 0. \qquad (1.18)$$

Этот закон распределения уже не будет показательным. Вид его зависит от параметра t_0 и вида функции $\lambda(t)$. Например, при линейном изменении

$$\lambda(t) = a + bt$$

плотность (1.18) имеет вид

$$f_{t_0}(t) = [a + b(t_0 + t)]\exp\left(-at - bt_0 t - \frac{bt^2}{2} \right). \qquad (1.19)$$

График этого закона при $a = 0,4;\ b = 2\ \ t_0 = 0,3$ представлен на рис. 1.10.

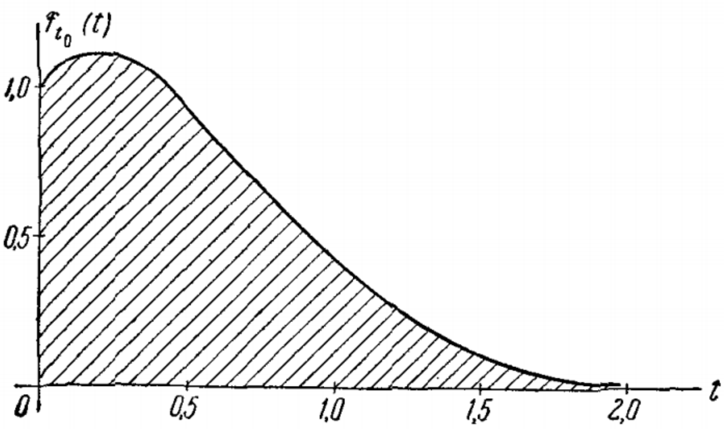


Рис. 1.10. График закона при $a = 0,4;\ b = 2$ и $t_0 = 0,3$

Несмотря на то, что структура нестационарного пуассоновского потока несколько сложнее, чем простейшего, он очень удобен в практических применениях: главное свойство простейшего потока – отсутствие последействия – в нем сохранено. А именно, если мы зафиксируем на оси Ot произвольную точку t_0, то закон распределения $f_{t_0}(t)$ времени T, отделяющего эту точку от ближайшего по времени будущего события, не зависит от того, что происходило на участке времени, предшествующем t_0 и в самой точке t_0, т.е. появлялись ли ранее другие события и когда именно).

Приведем несколько примеров.

1. Рассмотрим интервал $[t_0,\infty)$. Нас интересует момент t_i наступления i-го события. Как найти плотность вероятностей $p_i(t_i)$ величин t_i. Рассмотрим интервал $[t,t+dt]$. Вероятность того, что именно на этом интервале наступит i-е по счету событие, можно записать двумя путями:

$$P\ (t<t_i<t+dt)=p_i\left(t\right)dt=\frac{a^{i-1}\left(t_0,t\right)e^{-a\left(t_0,t\right)}\lambda\left(t\right)dt}{\left(i-1\right)!},$$

так как на интервале $[t_0, t)$ наступило $(i-1)$-е событие, а на интервале $[t,t+dt]$ ровно одно (уже i по счету). Сокращая на dt, получим

$$p_i(t)=\frac{a^{i-1}(t_0,t)e^{-a(t_0,t)}\lambda(t)}{(i-1)!}.$$

2. Рассмотрим интервал $[t_0,t_0+T]$, найдем плотность вероятностей того, что на нем наступили n событий в моменты времени $t_1,t_2,\ ...,t_n$. Выделяя бесконечно малые интервалы $[t_i,t_i+dt_i]$, будем иметь

$$p_n(t_1,t_2,\ldots,t_n)=e^{-a(t_0,t_1)}\lambda(t_1)dt_1\cdot e^{-a(t_1,t_2)}\lambda(t_2)dt_2\ldots$$

$$e^{-a(t_{n-1},t_n)}\lambda(t_n)dt_n\cdot e^{-\lambda(t_n,T)},$$

так как на интервалах $[t_0,t_1],[t_1+dt_1,t_2]\ ,\ldots$ не наступит ни одного события, а на интервалах $[t_i,t_i+dt_i]$ ровно по одному. Сокращая на $dt_1dt_2dt_n$, получим

$$p_n(t_1,t_2,\ldots,t_n)=e^{-a(t_0,t_0+T)}\lambda(t_1)\cdot\lambda(t_2)\ldots\lambda(t_n).$$

3. Пусть по моментам t_i появления событий на интервале $[t_0,T]$ мы составляем статистику $\xi=\sum\limits_{\geq 1}f(t_i)$, где $f(\bullet)$ – некоторая гладкая функция. Найдем $M\{\xi\}$ и $D\{\xi\}$.

Вспоминая выражение для $p_i(t)$, получим

$$M\{\xi\}=\sum_{i=1}^{\infty}M\{f(t_i)\}=\sum_{i=1}^{\infty}\int_{t_0}^{T}f(t)\frac{a^{i-1}(t_0,T)}{(i-1)!}e^{-a(t_0,t)}\lambda(t)dt=$$

$$=\int_{t_0}^{T}f(t)\lambda(t)e^{-a(t_0,t)}\sum_{i=1}^{\infty}\frac{a^{i-1}(t_0,T)}{(i-1)!}dt,$$

но

$$\sum_{i=1}^{\infty}\frac{a^{i-1}(t_0,T)}{(i-1)!}=e^{a(t_0,t)},$$

поэтому $M\{\xi\}=\int\limits_{t_0}^{T}f(t)\lambda(t)dt$.

Для вычисления дисперсии вычислим $M\{\xi^2\}$:

$$M\{\xi^2\}=M\left\{\sum_{i=1}^{\infty}f(t_i)\sum_{j=1}^{\infty}f(t_j)\right\}=M\left\{\sum_{i<j}f(t_i)f(t_j)\right\}+$$

$$+M\left\{\sum_{j<i}f(t_i)f(t_j)\right\}+M\left\{\sum_{i\geq 1}f(t_i)f(t_i)\right\}.$$

По аналогии с тем, что было при вычислении

$$\sum_{i=1}^{\infty}M\{f(t_i)\},$$

имеем

$$M\left\{\sum_{i=1}^{\infty}f^2(t_i)\right\}=\int\limits_{t_0}^{T}f^2(t)\lambda(t)dt,$$

далее

$$M\left\{\sum_{i<j}f(t_i)f(t_j)\right\}=\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\int\limits_{t_0}^{T}dt_i\int\limits_{t_i}^{T}dt_j\,f(t_i)f(t_j)p(t_i,t_j).$$

Легко видеть, что

$$p(t_i,t_j)dt_idt_j=\frac{a^{i-1}(t_0,t_i)}{(i-1)!}e^{-a(t_0,t_i)}\lambda(t_i)dt_i\frac{a^{j-i-1}(t_i,t_j)}{(j-i-1)!}e^{-a(t_i,t_j)}\lambda(t_j)dt_j,$$

так как на интервале $[t_0,t_i)$ произошло $(i-1)$ событий, а на интервале (t_i, t_j) – $(j-i-1)$ событие. Обозначая в интеграле (t_i) через u, а (t_j) через v, получим

$$M\left\{\sum_{i<j} f(t_i)f(t_j)\right\} = \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\int_{t_0}^{T}du\int_{t_i}^{T}dvf(u)f(v)\lambda(u)\lambda(v)\times$$

$$\times\frac{a^{i-1}(t_0,u)}{(i-1)!}\frac{a^{j-i-1}(u,v)}{(j-i-1)!}e^{-a(t_0,v)},$$

но

$$\sum_{j=i+1}^{\infty}\frac{a^{j-i-1}(u,v)}{(j-i-1)!} = e^{a(u,v)},$$

$$\sum_{i=1}^{\infty}\frac{a^{i-1}(t_0,u)}{(i-1)!} = e^{a(t_0,u)},$$

поэтому

$$M\left\{\sum_{i<j} f(t_i)f(t_j)\right\} = \int_{t_0}^{T}du\int_{u}^{T}dvf(u)f(v)\lambda(u)\lambda(v) =$$

$$=\frac{1}{2}\int_{t_0}^{T}f(u)\lambda(u)du\int_{t_0}^{T}f(v)\lambda(v)dv.$$

Легко получить, что математическое ожидание суммы при $j<i$ равно тому же выражению, поэтому

$$M\{\xi^2\} = \int_{t_0}^{T}f(u)\lambda(u)du\int_{t_0}^{T}f(v)\lambda(v)dv + \int_{t_0}^{T}f^2(u)\lambda(u)du,$$

и окончательно

$$D\{\xi\} = M\{\xi^2\} - M^2\{\xi\} = \int_{t_0}^{T}f^2(u)\lambda(u)du.$$

Пример 4. Рассмотрим простейший поток с нестационарным параметром, изменяющийся по закону $\lambda(t)=1+0,5\sin(6\pi t)$. Параметр является периодическим, его период равен 1/3. Найти вероятность отсутствия требований на отрезке [1,5].

Решение. Длина отрезка равна 4. Вычислим среднее число событий, наступающих в интервале времени τ = 4 по формуле (1.16):

$$a(1,4)=\int_{1}^{5}(1+0,5\sin(6\pi z))dz=4.$$

Тогда $P_0(1,4)=e^{-4}=0,0183.$

Суммирование и разъединение простейших потоков

При объединении нескольких независимых простейших потоков образуется также простейший поток с параметром, равным сумме параметров исходных потоков. При разъединении поступающего простейшего потока с параметром λ на n направлений так, что каждое требование исходного потока с вероятностью p_i $(\Sigma p_i=1)$ поступает на i-е направление, поток i-го направления также будет простейшим с параметром λp_i. Эти свойства простейшего потока широко используются на практике, поскольку значительно упрощают расчеты стационарного оборудования и информационных сетей.

Задание 1

1. Дан пуассоновский поток с параметром 2 мин$^{-1}$. Найти вероятность того, что длина интервала между соседними требованиями составляет от 1 до 2 мин.

2. Производится наложение (суперпозиция) двух простейших потоков с интенсивностями λ_1 и λ_2. Будет ли поток, получившийся в результате наложения, простейшим, и если да, то с какой интенсивностью?

3. Производится случайное прореживание простейшего потока событий с интенсивностью λ; каждое событие, независимо от других, с вероятностью p сохраняется в потоке, а с вероятностью $(1-p)$ выбрасывается. Каким будет поток, получающийся в результате прореживания простейшего потока?

4. Поток машин, идущих по шоссе в одном направлении, представляет собой простейший поток с интенсивностью 2 машины в 1 мин. Человек выходит на шоссе, чтобы остановить первую попавшуюся машину, идущую в данном направлении. Найти закон распределения времени T, в течение которого ему придется ждать машину; определить математическое ожидание и среднее квадратичное отклонение.

5. Поток машин, идущих по шоссе в одном направлении, представляет собой простейший поток с интенсивностью 4 маши-

ны в 1 мин. Шоссе имеет развилку в 2 направления. Вероятность движения машин в первом направлении равна 0,12, а во втором – 0,88. Определить интенсивности движения автомобилей в обоих направлениях.

6. Рассмотрим простейший поток с нестационарным параметром, изменяющимся по закону $\lambda(t) = 1 + 0,5\cos(6\pi t)$. Параметр является периодическим, его период равен 1/3. Найти вероятность отсутствия требований на отрезке [1;9].

7. Компьютерный класс связан с каналом Интернет через 10-канальный концентратор. Интенсивности передачи данных по каждому из 10 каналов равны соответственно 540, 120, 40, 170, 350, 60, 742, 153, 500 и 100 бит/с. Поток данных подчиняется пуассоновскому закону распределения. Определить интенсивность передачи данных в канале Интернет.

8. Рассмотрим простейший поток с нестационарным параметром, изменяющимся по закону $\lambda(t) = 1 + 0,5\sin(6\pi t)$. Параметр является периодическим, его период равен 1/4. Найти вероятность поступления одного, двух и трех требований.

9. Для простейшего потока с нестационарным параметром, определяемым равенством $\lambda(t) = 3 + 2^{-t}$, найти вероятность поступления двух требований на промежутке времени [3;8].

10. По железной дороге мимо наблюдателя движется в одном направлении простейший поток поездов. Известно, что вероятность отсутствия поездов в течение 10 мин равна 0,8. Требуется найти вероятность того, что за 20 мин мимо наблюдателя пройдет не более трех поездов.

11. Производится случайное прореживание простейшего потока событий с интенсивностью $\lambda = 4$; каждое событие, независимо от других с вероятностью $p = 0,6$ сохраняется в потоке, а с вероятностью $(1 - p)$ выбрасывается. Каким будет поток, получающийся в результате прореживания простейшего потока?

12. Рассмотрим простейший поток с нестационарным параметром, изменяющимся по закону $\lambda(t) = 2 + 0,5\sin(4\pi t)$. Параметр является периодическим, его период равен 1/3. Найти вероятность отсутствия требований на отрезке [1;5].

13. Дан пуассоновский поток с параметром 1 мин$^{-1}$. Найти вероятность того, что длина интервала между соседними требованиями составляет от 2 до 4 мин.

14. Поток машин, идущих по шоссе в одном направлении, представляет собой простейший поток с интенсивностью 8 машин

в 1 мин. Шоссе имеет развилку в 3 направления. Вероятность движения машин в первом направлении равна 0,12, во втором – 0,68, в третьем – 0,20. Определить интенсивность движения автомобилей во всех направлениях.

15. Поток машин, идущих по шоссе в одном направлении, представляет собой простейший поток с интенсивностью 6 машин в 1 мин. Человек выходит на шоссе, чтобы остановить первую попавшуюся машину, идущую в данном направлении. Найти закон распределения времени T, которое ему придется ждать; определить его математическое ожидание и среднее квадратичное отклонение.

16. Для простейшего потока с нестационарным параметром, определяемым равенством $\lambda(t) = 7 - 5^{-t}$, найти вероятность поступления двух требований на промежутке времени [1;10].

17. В пункт текущего отделочного ремонта вагонов поступают требования на ремонт. Поток требований можно считать простейшим с интенсивностью $\lambda = 0{,}307$. Найти вероятность того, что за час не поступит ни одного требования (вагона) на ремонт.

18. Время обслуживания для аппаратов некоторой системы массового обслуживания распределено по показательному закону $F(t){=}1{-}e^{-1{,}5t}$, где t – время [мин]. Найти вероятность того, что обслуживание продлится не более 15 мин.

19. Для простейшего потока с нестационарным параметром, определяемым равенством $\lambda(t) = 3 + 2^{-2t}$, найти вероятность поступления двух требований на промежутке времени [2;6].

20. В пункт текущего отделочного ремонта вагонов поступает требование на ремонт. Поток требований можно считать простейшим с интенсивностью $\lambda = 0{,}517$. Найти вероятность того, что за час поступит более одного требования (вагон) на ремонт.

21. Время обслуживания для аппаратов некоторой системы массового обслуживания распределено по показательному закону $F(t){=}1{-}e^{-0{,}5t}$, где t – время [мин].

Найти вероятность того, что обслуживание продлится не более 5 мин.

22. Производится случайное прореживание простейшего потока событий с интенсивностью $\lambda = 0{,}7$; каждое событие, независимо от других с вероятностью $p = 0{,}75$ сохраняется в потоке, а с вероятностью $(1 - p)$ выбрасывается. Каким будет поток, получающийся в результате прореживания простейшего потока?

23. Производится разбиение случайного простейшего потока событий с интенсивностью λ = 4,9 на 3 потока. Вероятности попадания событий в тот или иной поток соответственно равны $p_1 = 0,2, p_2 = 0,54, p_3 = 0,26$. Определить интенсивности каждого получившегося потока в результате разбиения.

24. Время обслуживания для аппаратов некоторой системы массового обслуживания распределено по показательному закону $F(t) = 4 - e^{-1,6t}$, где t – время [мин]. Найти вероятность того, что обслуживание продлится не более 8 мин.

25. В пункт текущего отделочного ремонта вагонов поступают требования на ремонт. Поток требований можно считать простейшим с интенсивностью λ = 0,617. Найти вероятность того, что за час поступит 1 требование (вагон) на ремонт.

26. Производится разбиение случайного простейшего потока событий с интенсивностью λ = 1,6 на 2 потока. Вероятности попадания событий в тот или иной поток соответственно равны $p_1 = 0,44, p_2 = 0,56$. Определить интенсивность каждого получившегося в результате разбиения потока.

27. Компьютерный класс связан с каналом Интернет через 5-канальный концентратор. Интенсивности передачи данных по каждому из 10 каналов равны соответственно 541, 110, 44, 171 и 356 бит/с. Поток данных подчиняется пуассоновскому закону распределения. Определить интенсивность передачи данных в канале Интернет.

28. Рассмотрим простейший поток с нестационарным параметром, изменяющимся по закону $\lambda(t) = 2 + 0,5\sin(4\pi t)$. Параметр является периодическим, его период равен 1/3. Найти вероятность отсутствия требований на отрезке [4;9].

29. На вокзал прибывает пуассоновский поток поездов, в среднем 2 поезда за 5 мин. Найти вероятность того, что за 15 мин прибудут 3 поезда.

30. Время обслуживания для аппаратов некоторой системы массового обслуживания распределено по показательному закону $F(t) = 1 - e^{-4.5t}$, где t – время [мин]. Найти вероятность того, что обслуживание продлится не более 20 мин.

Задание 2

Интенсивность простейшего потока заявок равна λ.

Определить:

1) *средний интервал* времени между соседними заявками в потоке (математическое ожидание равно $1/\lambda$);

2) *среднее число заявок*, поступающих в систему за время τ (математическое ожидание, $a = \lambda\tau$);

3) *вероятность* того, что за время τ в систему не поступит ни одной заявки;

4) *вероятность* того, что за время τ в систему поступит хотя бы одна заявка;

5) *вероятность* того, что за время τ в систему поступит больше одной заявки.

Вариант	1	2	3	4	5	6	7	8	9	10
λ, c^{-1}	0,25	0,25	0,25	0,5	1,0	1,0	1,5	1,5	2,0	2,0
τ, c	2,0	4,0	8,0	2,0	2,0	5,0	2,0	1,0	1,0	2,0

Задание 3

Интенсивность простейшего потока заявок равна λ.

1) Определить, поступление какого *числа заявок* за промежуток времени (τ_1, τ_2) наиболее вероятно.

2) *Сравнить* это значение со средним числом заявок, поступающих за промежуток времени (τ_1, τ_2).

3) Определить *вероятность* того, что промежуток времени между двумя соседними заявками в потоке будет находиться в интервале (τ_1, τ_2).

Вариант	1	2	3	4	5	6	7	8	9	10
λ, c^{-1}	0,25	0,25	0,5	0,5	1,0	1,0	1,5	1,5	2,0	2,0
τ_1, c	0	2,0	0	1,0	0	1,0	0	1,0	0	0,5
τ_2, c	2,0	6,0	2,0	4,0	1,5	2,0	3,0	2,0	4,0	1,0

1.5. Поток с ограниченным последействием (поток Пальма)

Мы познакомились с естественным обобщением простейшего потока – с нестационарным пуассоновским потоком. Обобщением простейшего потока в другом направлении является поток с ограниченным последействием.

Рассмотрим ординарный поток однородных событий (рис. 1.11).

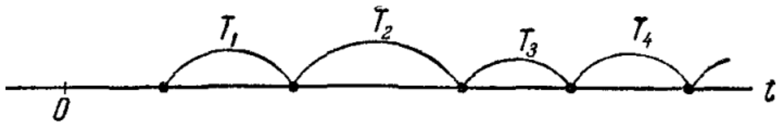


Рис. 1.11. Ординарный поток однородных событий

Этот поток называется потоком с ограниченным последействием (или потоком Пальма), если промежутки времени между последовательными событиями T_1, T_2 представляют собой независимые случайные величины.

Очевидно, простейший поток является частным случаем потока Пальма: в нем расстояния T_1, T_2 представляют собой независимые случайные величины, распределенные по показательному закону. Что касается нестационарного пуассоновского потока, то он не является потоком Пальма. Действительно, рассмотрим два соседних промежутка T_k и T_{k+1} в нестационарном пуассоновском потоке. Как мы видели в предыдущем пункте, закон распределения промежутка между событиями в нестационарном потоке зависит от того, где этот промежуток начинается, а начало промежутка T_{k+1} совпадает с концом промежутка T_k, значит, длины этих промежутков зависимы.

Рассмотрим примеры потоков Пальма.

1. Некоторая деталь технического устройства (например, радиолампа) работает непрерывно до своего отказа (выхода из строя), после чего она мгновенно заменяется новой. Срок безотказной работы детали случаен; отдельные экземпляры выходят из строя независимо друг от друга. При этих условиях поток отказов (или поток «восстановлений») представляет собой поток Пальма. Если, к тому же, срок работы детали распределен по показательному закону, то поток Пальма превращается в простейший.

2. Группа самолетов идет в боевом порядке «колонна» (рис. 1.12) с одинаковой для всех самолетов скоростью V. Каждый самолет, кроме ведущего, обязан выдерживать строй, т.е. держаться на заданном расстоянии L от впереди идущего. Это расстояние, вследствие погрешностей радиодальномера, выдерживается с ошибками. Моменты пересечения самолетами заданного рубежа образуют поток Пальма, так как случайные величины $T_1 = \dfrac{L_1}{V}$,

$T_2 = \dfrac{L_2}{V}, \ldots$ независимы. Заметим, что тот же поток не будет потоком Пальма, если каждый из самолетов стремится выдержать заданное расстояние не от соседа, а от ведущего.

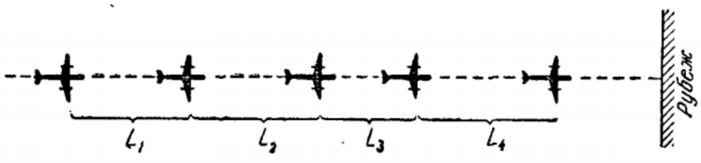


Рис. 1.12. Группа самолетов идет в боевом порядке «колонна»

Потоки Пальма часто получаются в виде выходных потоков систем массового обслуживания. Если на какую-либо систему поступает какой-то поток заявок, то он этой системой разделяется на два: поток обслуженных и поток необслуженных заявок.

Поток необслуженных заявок часто поступает на какую-либо другую систему массового обслуживания, поэтому представляет интерес изучить его свойства.

Основной в теории выходных потоков является **теорема Пальма**, которую мы сформулируем без доказательства.

Пусть на систему массового обслуживания поступает поток заявок типа Пальма, причем заявка, заставшая все каналы занятыми, получает отказ (не обслуживается). Если при этом время обслуживания имеет показательный закон распределения, то поток необслуженных заявок является также потоком типа Пальма.

В частности, если входной поток заявок будет простейшим, то поток необслуженных заявок, не будучи простейшим, будет все же иметь ограниченное последействие.

Интересным примером потоков с ограниченным последействием являются так называемые потоки Эрланга. Они образуются «просеиванием» простейшего потока.

Рассмотрим простейший поток (рис. 1.13) и выбросим из него каждую вторую точку (на рисунке выброшенные точки отмечены крестами).

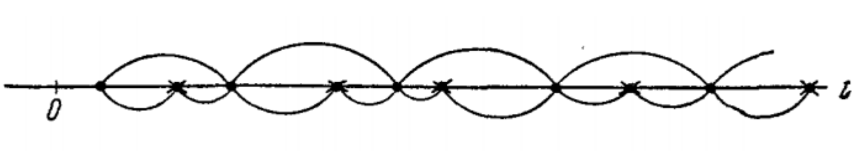


Рис. 1.13. Поток Эрланга первого порядка

Оставшиеся точки образуют поток; этот поток называется потоком Эрланга первого порядка ($Э_1$). Очевидно, этот поток есть поток Пальма: поскольку независимы промежутки между событиями в простейшем потоке, то независимы и величины $T_1, T_2, \ldots$, получающиеся суммированием таких промежутков по два.

Поток Эрланга второго порядка получится, если сохранить в простейшем потоке каждую третью точку, а две промежуточные выбросить (рис. 1.14).

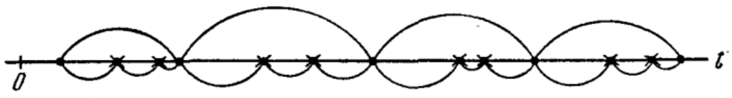


Рис. 1.14. Поток Эрланга второго порядка

Вообще, потоком Эрланга k-го порядка ($Э_k$) называется поток, получаемый из простейшего, если сохранить каждую (k + 1)-ю точку, а остальные выбросить. Очевидно, простейший поток можно рассматривать как поток Эрланга нулевого порядка ($Э_0$).

Найдем закон распределения промежутка времени T между соседними событиями в потоке Эрланга k-го порядка ($Э_k$). Рассмотрим на оси $0t$ (рис. 1.15) простейший поток с интервалами $T_1, T_2, \ldots$

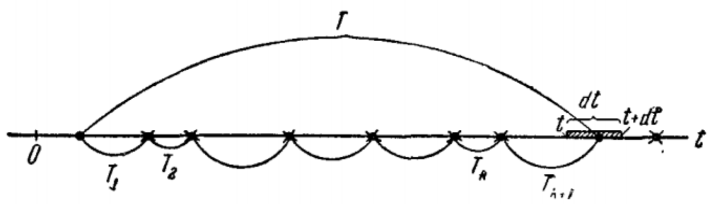


Рис. 1.15. Промежуток времени T между соседними событиями

Величина T представляет собой сумму (k + 1) независимых случайных величин

$$T = \sum_{i=1}^{k+1} T_i, \tag{1.20}$$

где $T_1, T_2, \ldots, T_{k+1}$ – независимые случайные величины, подчиненные одному и тому же показательному закону

$$f(t) = \lambda e^{-\lambda t}, \ t > 0. \tag{1.21}$$

Можно было бы найти закон распределения величины T как композицию $(k+1)$ законов (1.21). Однако проще вывести его элементарными рассуждениями.

Обозначим через $f_k(t)$ плотность распределения величины T для потока (Э$_k$); $f_k(t)dt$ есть вероятность того, что величина T примет значение между t и $t+dt$ (рис. 1.15). Это значит, что последняя точка промежутка T должна попасть на элементарный участок $(t, t+dt)$, а предыдущие k точек простейшего потока – на участок $(0,t)$. Вероятность первого события равна λdt, вероятность второго, на основании формулы (1.6), будет равна

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Перемножая эти вероятности, получим

$$f_k(t)dt = \frac{\lambda(\lambda t)^k}{k!} e^{-\lambda t} dt,$$

откуда

$$f_k(t) = \frac{\lambda(\lambda t)^k}{k!} e^{-\lambda t}, \quad t > 0. \tag{1.22}$$

Закон распределения с плотностью (1.22) называется законом Эрланга k-го порядка. Очевидно, при $k = 0$ он обращается в показательный:

$$f_0(t) = \lambda e^{-\lambda t}, \quad t > 0. \tag{1.23}$$

Найдем характеристики закона Эрланга $f_k(t)$: математическое ожидание m_k и дисперсию D_k. По теореме сложения математических ожиданий имеем

$$m_k = \sum_{i=1}^{k+1} m_0 = (k+1)m_0,$$

где $m_0 = \dfrac{1}{\lambda}$ – математическое ожидание промежутка между событиями в простейшем потоке.

Отсюда

$$m_k = \frac{k+1}{\lambda}. \tag{1.24}$$

Аналогично по теореме сложения дисперсий получаем

$$D_k = \frac{k+1}{\lambda^2}, \quad \sigma_k = \frac{\sqrt{k+1}}{\lambda}. \tag{1.25}$$

Плотность Λ_k потока $Э_k$ будет обратна величине m_k:

$$\Lambda_k = \frac{\lambda}{k+1}. \tag{1.26}$$

Таким образом, при увеличении порядка потока Эрланга увеличиваются как математическое ожидание, так и дисперсия промежутка времени между событиями, а плотность потока падает.

Выясним, как будет изменяться поток Эрланга при $k \to \infty$, если его плотность будет сохраняться постоянной. Пронормируем величину T так, чтобы ее математическое ожидание (и, следовательно, плотность потока) оставалось неизменным. Для этого изменим масштаб по оси времени и вместо T рассмотрим величину

$$\tilde{T} = \frac{T}{k+1}. \tag{1.27}$$

Назовем такой поток нормированным потоком Эрланга k-го порядка. Закон распределения промежутка $\tilde{T}$ между событиями этого потока будет равен

$$\tilde{f}_k(t) = \frac{\Lambda_k (\Lambda_k t)^k}{k!} e^{-\Lambda_k t}, \quad t > 0, \tag{1.28}$$

где $\Lambda_k = \lambda(k+1)$.

Или

$$\tilde{f}_k(t) = \frac{\lambda(k+1)}{k!} (\lambda(k+1)t)^k e^{-\lambda(k+1)t}, \ t > 0. \tag{1.30}$$

Математическое ожидание величины $\tilde{T}$, распределенной по закону (1.30), не зависит от k и равно

$$\tilde{m}_k = m_0 = \frac{1}{\lambda},$$

где λ – плотность потока, совпадающая при любом k с плотностью исходного простейшего потока.

Дисперсия величины $\tilde{T}$ равна

$$\tilde{D}_k = \frac{D_k}{(k+1)^2} = \frac{1}{\lambda^2 (k+1)} \tag{1.31}$$

и неограниченно убывает с возрастанием k.

Таким образом, мы приходим к выводу: при неограниченном увеличении k нормированный поток Эрланга приближается к регулярному потоку с постоянными интервалами, равными $\frac{1}{\lambda}$.

Это свойство потоков Эрланга удобно в практических применениях: оно дает возможность, задаваясь различными k, получить любую степень последействия: от полного отсутствия $(k=0)$ до жесткой функциональной связи между моментами появления событий ($k=\infty$). Таким образом, порядок потока Эрланга может служить как бы «мерой последействия», имеющегося в потоке. В целях упрощения часто бывает удобно заменить реальный поток заявок, имеющий последействие, нормированным потоком Эрланга с примерно теми же характеристиками промежутка между заявками: математическим ожиданием и дисперсией.

Пример 5. В результате статистической обработки промежутков между заявками в потоке получены оценки для математического ожидания и дисперсии величины *T*: $m_t=2$ мин, $D_t=0{,}8$ мин^2. Заменить этот поток нормированным потоком Эрланга с теми же характеристиками.

Решение. Имеем $\lambda=\frac{1}{m_t}=0{,}5$. Из формулы (1.31) получим

$$k+1\approx\frac{1}{D_t\lambda^2}=\frac{1}{0.2}=5,\quad k=4.$$

Поток можно приближенно заменить нормированным потоком Эрланга четвертого порядка.

Коэффициент вариации случайной величины *X* определяется по формуле

$$[X]=\frac{\sigma[X]}{M[X]}\quad (M[X]>0).$$

Коэффициент вариации нормированного распределения Эрланга меньше или равен единице: $v_{н_{Э_k}}[X]=(k+1)^{-1/2}\le 1$, но математическое ожидание *не зависит* от значения параметра k.

Найдем закон распределения величины *T* как композицию $(k+1)$ законов $f(t)=\lambda e^{-\lambda t}$. В потоке Эрланга интервал между двумя соседними требованиями равен сумме $(n+1)$ интервалов между требованиями Пуассоновского потока.

Под функцией Пальма – Хинчина $\varphi_k(t)$ понимают вероятность того, что за время t появится k требований, если в начальный момент этого интервала пришло хотя бы одно требование (в силу однородности придет только одно), причем это требование в счет k требований не включается:

$$\varphi_k(t)=\begin{cases}1, & a\le t<(k+1)a;\\ 0, & t<ka,\ t\ge(k+1)a,\end{cases}\qquad a=\frac{1}{\lambda},$$

$$\varphi_0(t)=\begin{cases}1, & t<a;\\ 0, & t\ge a;\end{cases}.$$

$P_0(t)=\begin{cases}1-t/a, & t<a;\\ 0, & t\le a\end{cases}$ – вероятность отсутствия требования.

При $k\ge 1$ $\Phi_k(t)=\varphi_{k-1}(t)-\varphi_k(t)$;

$$\varphi_{k-1}(t)=\begin{cases}1, & (k-1)a\le t<ka;\\ 0, & t<(k-1)a,\ t\ge ka,\end{cases}$$

тогда

$$\Phi_k(t)=\begin{cases}0, & t<(k-1)a,\\ 1, & (k-1)a\le t<ka,\\ -1, & ka\le t<(k+1)a,\\ 0, & t\ge(k+1)a;\end{cases}$$

$$P_k(t)=\frac{1}{a}\int_0^t\Phi_k(x)dx=\begin{cases}0, & t<(k-1)a,\\ \dfrac{t}{a}-k+1, & (k-1)a\le t<ka,\\ k+1-\dfrac{t}{a}, & ka\le t<(k+1)a,\\ 0, & t\ge(k+1)a.\end{cases}$$

$\overline{\varphi_0}(t)$ – в потоке Эрланга за время t не появилось ни одного требования, если в Пуассоновском потоке за это время появится не более n требований:

$$\overline{\varphi_0}(t)=\sum_{k=0}^{n}P_k(t)=\sum_{k=0}^{n}\frac{(\lambda t)^k}{k!}e^{-\lambda t},$$

тогда

$$\overline{\varphi_k}(t)=\sum_{i=(n+1)k}^{(n+1)(k+1)-1}\frac{(\lambda t)^i}{i!}e^{-\lambda t},\ (n+1)k\le i<(n+1)(k+1).$$

Функция распределения длин интервалов между соседними требованиями в потоке Эрланга определяется по формуле

$$\overline{F(t)} = 1 - \overline{\varphi_0}(t) = 1 - e^{-\lambda t} \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!},$$

плотность распределения:

$$\overline{f(t)} = \overline{F'(t)} = \left(-e^{-\lambda t} \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} \right)' = \frac{\lambda (\lambda t)^n}{n!} e^{-\lambda t}.$$

Вероятность того, что в потоке Эрланга за время t не появилось ни одного требования, если в Пуассоновском потоке за это время появится не более n требований, будет равна

$$\overline{P_0(t)} = e^{-\lambda t} \sum_{i=0}^{n+1} \frac{(\lambda t)^i}{i!} \left(1 - \frac{i}{n+1} \right).$$

Вероятность того, что в потоке Эрланга за время t не появилось k требований, если в Пуассоновском потоке за это время появится не более n требований, будет равна

$$\overline{P_k(t)} = e^{-\lambda t} \sum_{i=(n+1)(k-1)}^{(n+1)(k+1)} \frac{(\lambda t)^i}{i!} \left(1 - \left| \frac{i}{n+1} - k \right| \right).$$

На рис. 1.16 показаны плотности распределения Эрланга при $a = 1$ для трех значений параметра: $k=1$; $k=2$; $k=4$.

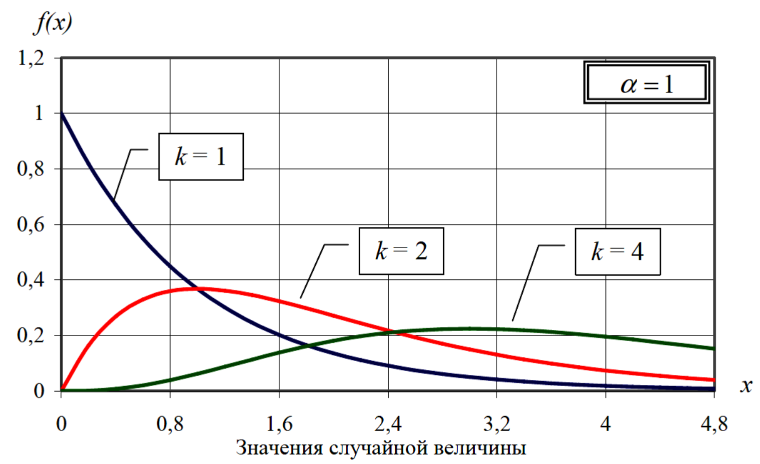


Рис. 1.16. Плотность распределения Эрланга

При $k=1$ распределение Эрланга вырождается в экспоненциальное, а при $k \to \infty$ приближается к *нормальному распределению*.

Преобразование Лапласа распределения Эрланга k-го порядка имеет вид

$$F^*(s) = \left(\frac{\alpha}{\alpha + s}\right)^k.$$

Поскольку распределение Эрланга является *двухпараметрическим*, то оно может использоваться для аппроксимации реальных распределений по двум первым моментам.

На рис. 1.17 показаны плотности распределения Эрланга при $a = 1$ для трех значений параметра: $k = 1$; $k = 2$; $k = 16$.

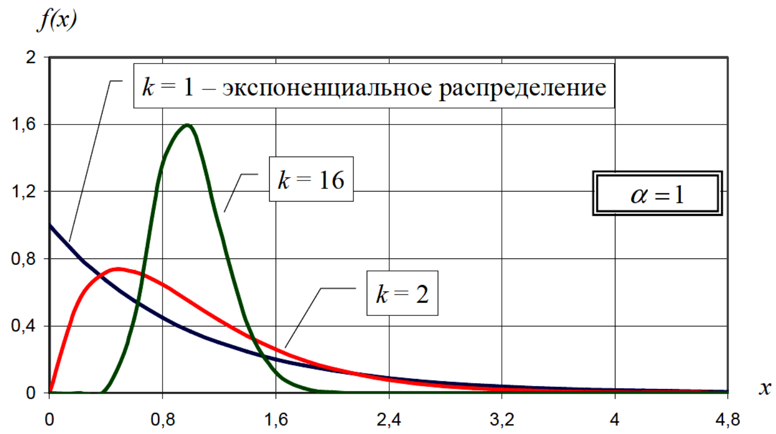


Рис. 1.17. Плотность нормированного распределения Эрланга

Преобразование Лапласа нормированного распределения Эрланга равно $F^*(s) = \dfrac{(k\alpha)^k}{k\alpha + s}$.

Задание для самостоятельной работы

1. Определить математическое ожидание, дисперсию, коэффициент вариации распределения Эрланга k-го порядка.
2. Доказать, что коэффициент вариации распределения Эрланга не превышает 1.
3. Построить графики функции и плотности распределений Эрланга пятого и восьмого порядков.
4. Построить графики функции и плотности нормированного распределения Эрланга пятого порядка и сравнить с простым распределением Эрланга.
5. Определить математическое ожидание, второй начальный момент, дисперсию, коэффициент вариации нормированного рас-

пределения Эрланга. Доказать, что коэффициент вариации нормированного распределения Эрланга не превышает 1.

Пример 6. В течение 8 ч работы оперативно-диспетчерская группа района электрических сетей получила три вызова. Определить вероятность того, что в течение девятого часа будет получен еще один вызов.

Решение. Определим интенсивность потока заявок. По формуле $\tilde{f}_k(t) = \dfrac{\Lambda_k(\Lambda_k t)^k}{k!} e^{-\Lambda_k t}$ получим:

$$\lambda = 3/8 = 0{,}37;\ f_k = 0{,}37\left(0{,}37 \cdot 8\right)4e^{-3/6} = 0{,}23.$$

Пример 7. На диспетчерский пульт поступает поток заявок, который является потоком Эрланга второго порядка. Интенсивность потока заявок – 6 заявок в 1 ч. Если диспетчер в случайный момент оставляет пульт, то при первой же очередной заявке он обязан вернуться к пульту. Найти плотность распределения времени ожидания очередной заявки и построить ее график. Вычислить вероятность того, что диспетчер сможет отсутствовать от 10 до 20 мин.

Решение. Поскольку поток Эрланга второго порядка является стационарным потоком с ограниченным последействием, то для него справедлива формула Пальма:

$$f_1(\theta) = \frac{\lambda}{k}[1 - F_k(\theta)],\quad \theta > 0, \qquad (*)$$

где $f_1(t)$ – плотность распределения вероятностей для времени ожидания первого ближайшего события; λ – интенсивность потока; k – порядок потока; $F_k(t)$ – функция распределения вероятностей для времени между двумя соседними событиями потока Эрланга k-го порядка ($Э_k$).

Известно, что функция распределения для потока $Э_k$ имеет вид

$$F_k(\theta) = 1 - \sum_{s=0}^{k-1} \frac{(\lambda\theta)^s}{s!} e^{-\lambda\theta}. \qquad (**)$$

По условиям задачи поток заявок является эрланговским порядка $k = 2$. Тогда из (*) и (**) получим

$$f_1(\theta) = \frac{\lambda}{2}\left[1 - \left(1 - \sum_{s=0}^{1} \frac{(\lambda\theta)^s}{s!} e^{-\lambda\theta}\right)\right] = \frac{\lambda}{2}\left[e^{-\lambda\theta} + \lambda\theta e^{-\lambda\theta}\right] = \frac{\lambda}{2} e^{-\lambda\theta}(1 + \lambda\theta).$$

Из последнего соотношения при λ = 6 будем иметь

$$f_1(\theta) = 3e^{-6\theta}(1+6\theta), \quad \theta \geq 0.$$

Построим график функции $f_1(\theta)$. При $\theta < 0$ имеем $f_1(\theta) = 0$. При t = 0 имеем $f_1(\theta) = 3$.

Рассмотрим предел

$$\lim_{\theta\to\infty} f_1(\theta) = 3\lim_{\theta\to\infty}\frac{1+6\theta}{(e^{6\theta})^t} = 3\lim_{\theta\to\infty}\frac{(1+6\theta)'}{((e^{6\theta})^t)''} = 3\lim_{\theta\to\infty}\frac{1}{e^{6\theta}} = 0.$$

При вычислении предела для раскрытия неопределенности типа (∞/∞) использовано правило Лопиталя. По результатам исследований строим график функции $f_1(t)$ (рис. 1.18).

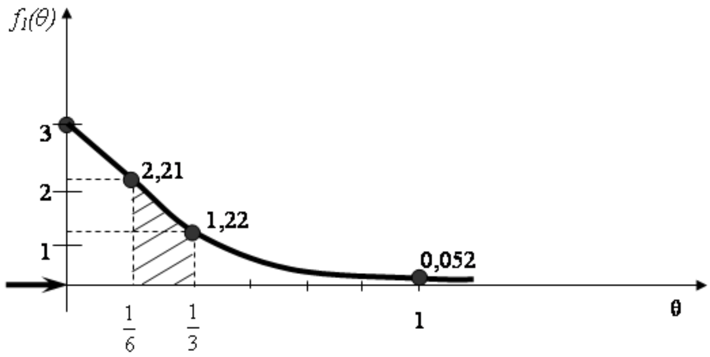


Рис. 1.18. График функции $f_1(t)$

Обратим внимание на размерности времени в тексте задачи: для интенсивности – это заявки в 1 ч, для времени – минуты. Перейдем к одним единицам времени: 10 мин = 1/6 ч, 20 мин = 1/3 ч. Для этих значений можно вычислить $f_1(t)$ и уточнить характер кривой:

$$f_1\left(\frac{1}{6}\right) = 3e^{-1}\left(1+6\cdot\frac{1}{6}\right) = \frac{6}{e} \approx 2,21,$$

$$f_1\left(\frac{1}{3}\right) = 3e^{-2}\left(1+6\cdot\frac{1}{3}\right) = \frac{9}{e^2} \approx 1,22,$$

$$f_1(1) = 3e^{-6}(1+6\cdot 1) = \frac{21}{e^6} \approx 0,05.$$

Эти ординаты указаны на графике над соответствующими точками кривой.

Из курса теории вероятностей известно, что вероятность попадания случайной величины X в отрезок $[\alpha, \beta]$ численно равна площади под кривой плотности распределения вероятностей $f(x)$. Эта площадь выражается определенным интегралом:

$$P(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f(x)dx.$$

Следовательно, искомая вероятность равна

$$P\left(\frac{1}{6} \le \theta \le \frac{1}{3}\right) = 3\int_{\frac{1}{6}}^{\frac{1}{3}} e^{-6\theta}(1+6\theta)dx.$$

Этот интеграл легко вычисляется по частям, если положить $U = 1+6\theta$ и $dV = e^{-6t}dt$, тогда $dU = 6dt$ и $V = -\frac{1}{6}e^{-6\theta}$.

Используя формулу интегрирования по частям, получим

$$P\left(\frac{1}{6} \le \theta \le \frac{1}{3}\right) = \frac{3}{2}e^{-1} - 2e^{-2} \approx 0,28.$$

Ответ: вероятность того, что диспетчер сможет отсутствовать от 10 до 20 мин, равна 0,28.

1.6. Время обслуживания

Кроме характеристик входного потока заявок, режим работы системы зависит еще от характеристик производительности самой системы: числа каналов n и быстродействия каждого канала. Одной из важнейших величин, связанных с системой, является время обслуживания одной заявки b. Эта величина может быть как неслучайной, так и случайной. Очевидно, более общим является случайное время обслуживания.

Рассмотрим случайную величину b и обозначим через $G(t)$ ее функцию распределения:

$$G(t) = P(b < t), \tag{1.32}$$

$g(t)$ – плотность распределения:

$$g(t) = G'(t). \tag{1.33}$$

Для практики особый интерес представляет случай, когда величина b имеет показательное распределение

$$g(t) = \mu e^{-\mu t} \ (t > 0), \qquad (1.34)$$

где μ – величина, обратная среднему времени обслуживания одной заявки и называемая интенсивностью обслуживания:

$$\mu = \frac{1}{b}. \qquad (1.35)$$

Особая роль показательного закона распределения величины b формулируется так: если в какой-то момент t_0 происходит обслуживание заявки, то закон распределения оставшегося времени обслуживания не зависит от того, сколько времени обслуживание уже продолжалось.

Время T_0, оставшееся до завершения обслуживания заявки, находящейся в приборе, от момента поступления некоторой заявки в систему и учитывающее, что на момент поступления в системе может и не оказаться заявок, т.е. учитывающее простои системы, называется ***временем дообслуживания***. Математическое ожидание этого времени имеет вид

$$M\ [T_0] = \lambda b^2\ (1+v_b^2)/2,$$

где λ – интенсивность *простейшего* потока заявок, поступающих в систему; v_b^2 – коэффициент вариации длительности обслуживания.

На первый взгляд допущение о том, что время обслуживания распределено по показательному закону, представляется довольно искусственным. В ряде практических задач кажется естественнее предположить его либо совсем не случайным, либо распределенным по нормальному закону. Однако существуют условия, в которых время обслуживания действительно распределяется по закону, близкому к показательному.

Это, прежде всего, все задачи, в которых обслуживание сводится к ряду «попыток», каждая из которых приводит к необходимому результату с какой-то вероятностью p.

Пусть, например, «обслуживание» состоит в обстреле какой-то цели и заканчивается в момент ее поражения. Обстрел ведется независимыми выстрелами с некоторой средней скорострельностью X выстрелов в единицу времени. Каждый выстрел поражает цель с вероятностью p. Чтобы не связывать себя необходимостью точного учета момента каждого выстрела, предположим, что они происходят в случайные моменты времени и образуют простейший поток П с плотностью X (рис. 1.19).

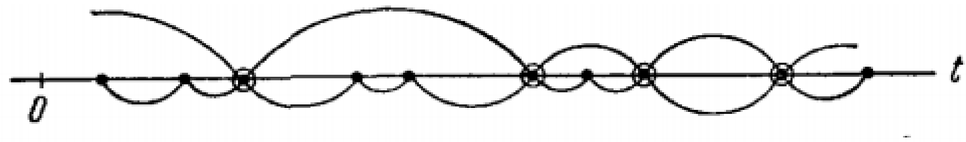


Рис. 1.19. Простейший поток П

Выделим мысленно из этого потока другой – поток «успешных», или «поражающих», выстрелов (они отмечены кружками на рис. 1.19). Выстрел будем называть «успешным», если он приводит к поражению цели (если только цель не была поражена ранее). Нетрудно убедиться, что успешные выстрелы тоже образуют простейший поток П с плотностью $\Lambda = \lambda p$ (исходный поток П – простейший, а каждый выстрел может стать поражающим, независимо от других, с вероятностью p). Вероятность того, что цель будет поражена до момента t, будет равна

$$G(t) = P(b < t) = 1 - e^{-\Lambda t},$$

откуда плотность распределения времени «обслуживания» равна

$$g(t) = \Lambda e^{-\Lambda t},$$

а это есть показательный закон с параметром $\mu = \Lambda$.

Показательным законом распределения времени обстрела до поражения цели можно приближенно пользоваться и в случае, когда выстрелы не образуют простейшего потока, а отделены, например, строго определенными промежутками времени t_1, если только вероятность поражения одним выстрелом p не очень велика. Для иллюстрации приведем на одном и том же графике (рис. 1.20) функцию распределения момента поражающего выстрела (ступенчатая линия) для случая $p = 0,4,\ t_1 = 1$ и функцию распределения показательного закона с параметром $\mu = p = 0,4$ (плавная кривая).

Как видно на рис. 1.20, непрерывное показательное распределение соответствует характеру нарастания функции распределения для дискретного случая. Естественно, если моменты выстрелов не будут строго определенными, соответствие с показательным законом будет еще лучше.

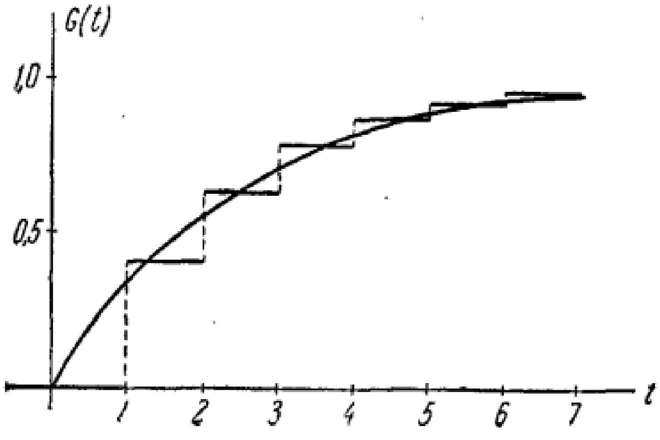


Рис. 1.20. Функция распределения момента поражающего выстрела (ступенчатая линия) и функция распределения показательного закона (плавная кривая)

Случай стрельбы – не единственный, когда обслуживание осуществляется рядом «попыток». К такому типу часто можно отнести обслуживание по устранению неисправностей технических устройств, когда поиски неисправной детали или узла осуществляются рядом тестов или проверок. К такому же типу можно отнести задачи, где «обслуживание» заключается в обнаружении какого-либо объекта радиолокатором, если объект с какой-то вероятностью может быть обнаружен при каждом цикле обзора.

Показательным законом хорошо описываются и те случаи, когда плотность распределения времени обслуживания по тем или иным причинам убывает при возрастании аргумента t. Это бывает, когда основная масса заявок обслуживается очень быстро, а значительные задержки в обслуживании наблюдаются редко. Рассмотрим, например, окно почтового отделения, где продаются марки и конверты, а также принимаются почтовые отправления и переводы. Основная масса посетителей покупает марки или конверты и обслуживается очень быстро. Реже встречаются заявки на отправление заказных писем, они обслуживаются несколько дольше. Переводы посылаются еще реже и обслуживаются еще дольше. Наконец, в самых редких случаях представители организаций отправляют сразу большое количество писем. Гистограмма распределения времени обслуживания имеет вид, представленный на рис. 1.21.

Так как плотность распределения убывает с возрастанием t, можно без особой погрешности выровнять распределение с помощью показательного закона, подобрав соответствующим образом его параметр μ.

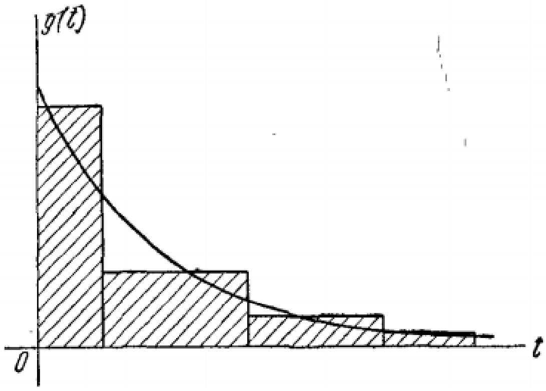


Рис. 1.21. Гистограмма распределения времени обслуживания

Разумеется, показательный закон не является универсальным законом распределения времени обслуживания. Часто время обслуживания лучше описывается, например, законом Эрланга. Однако, к счастью, пропускная способность и другие характеристики системы массового обслуживания сравнительно мало зависят от вида закона распределения времени обслуживания, а зависят, главным образом от его среднего значения $m_{t_{об}} = b$. Поэтому в теории массового обслуживания чаще всего пользуются допущением, что время обслуживания распределено по показательному закону. Эта гипотеза позволяет сильно упростить математический аппарат, применяемый для решения задач массового обслуживания, и в ряде случаев, получить простые аналитические формулы для характеристик пропускной способности системы.

1.7. Характеристики систем массового обслуживания с однородным потоком заявок

Характеристики систем со стохастическим характером функционирования являются *случайными величинами* и полностью описываются соответствующими законами распределений. На практике при моделировании часто ограничиваются определением только *средних значений* (математических ожиданий), реже – определением двух первых моментов этих характеристик.

В качестве основных характеристик СМО с однородным потоком заявок используются следующие величины (в скобках указаны экономико-математические эквиваленты):

- ***Нагрузка*** системы:

$$y = \lambda/\mu = \lambda b. \quad (1.36)$$

• ***Коэффициент загрузки*** или просто ***загрузка*** системы, определяемая как доля времени, в течение которого система (в случае одноканальной СМО – прибор) работает, т.е. выполняет обслуживание заявок; загрузка может быть рассчитана как отношение *среднего* времени T_p работы одного прибора многоканальной СМО к общему времени наблюдения T:

$$\rho = \lim_{T \to \infty} \frac{T_p}{T}; \tag{1.37}$$

время T_p для СМО с K обслуживающими приборами определяется путем усреднения времени работы по всем приборам:

$$T_p = \frac{1}{K} \sum_{i=1}^{K} T_i,$$

где T_i – время работы прибора $i = 1, \ldots, K$.

Подставляя последнее выражение в (1.37), окончательно получим:

$$\rho = \lim_{T \to \infty} \frac{1}{TK} \sum_{i=1}^{K} T_i.$$

Очевидно, что $0 \le \rho \le 1$.

• ***Коэффициент простоя системы*** (p_0):

$$\eta = 1 - \rho. \tag{1.38}$$

• ***Вероятность потери*** заявок ($P_{\text{отк}}$):

$$\pi_n = \lim_{T \to \infty} \frac{N_n(T)}{N(T)}, \tag{1.39}$$

где T – время работы системы (наблюдения за системой); $N(T)$ – число заявок, поступивших в систему за время T; $N_n(T)$ – число потерянных заявок за время T.

• ***Вероятность обслуживания*** ($P_{\text{обсл}} = Q$ – относительная пропускная способность) заявки, т.е. вероятность того, что поступившая в систему заявка будет обслужена:

$$\pi_0 = 1 - \pi_n = \lim_{T \to \infty} \frac{N_0(T)}{N(T)}, \tag{1.40}$$

где $N_0(T)$ – число обслуженных в системе заявок за время T, причем $N_n(T) + N_0(T) = N(T)$ и $\pi_0 + \pi_n = 1$.

• ***Производительность системы*** (*A* – абсолютная пропускная способность), представляющая собой *интенсивность потока обслуженных заявок*, выходящих из системы

$$\lambda' = \pi_0\lambda = (1 - \pi_n)\lambda = Q\lambda = A \qquad (1.41)$$

для *СМО с накопителем неограниченной емкости,* при условии отсутствия перегрузок, вероятность потери заявок $\pi_n = 0$ и, следовательно, производительность системы совпадает с интенсивностью поступления заявок в систему: $\lambda' = \lambda$.

• ***Интенсивность потока потерянных*** (не обслуженных) ***заявок*** из-за ограниченной емкости накопителя:

$$\lambda'' = \pi_n\lambda\ (1 - \pi_0)\lambda, \qquad (1.42)$$

очевидно, что сумма интенсивностей потоков обслуженных и потерянных заявок должна быть равна интенсивности входящего в систему потока заявок: $\lambda' + \lambda'' = \lambda$.

• ***Среднее время ожидания*** заявок в очереди: w ($T_{оч}$).

• ***Среднее время пребывания*** заявок в системе ($T_{сист}$), складывающееся из времени ожидания w и времени обслуживания b:

$$u = w + b. \qquad (1.43)$$

• ***Средняя длина очереди*** заявок ($L_{оч}$):

$$l = \lambda' w. \qquad (1.44)$$

• ***Среднее число заявок в системе*** (в очереди и на обслуживании в *приборе*) ($L_{\mathbf{сист}}$):

$$m = \lambda' u. \qquad (1.45)$$

Нагрузка и загрузка являются важнейшими характеристиками СМО, определяющими качество функционирования системы.

Нагрузка $y = \lambda b$ представляет собой интегральную оценку, объединяющую два нагрузочных параметра: частоту использования некоторого ресурса (прибора СМО), задаваемую в виде интенсивности λ поступления заявок в СМО, и время использования этого ресурса, задаваемое в виде средней длительности b обслуживания заявок в СМО. Нагрузка показывает количество работы, которую необходимо выполнить в системе. Если значение нагрузки $y < 1$, то заданная нагрузка может быть выполнена одним обслуживающим прибором, т.е. одноканальная СМО будет работать без перегрузки. Если $y > 1$, то реализация заданной нагрузки в одноканальной СМО приведет к режиму перегрузки, означающему, что с течением времени все большее число заявок будет оставаться необслуженным, и в случае накопителя неограниченной емкости очередь заявок будет расти неограниченно. Для того чтобы система работала без перегрузок, необходимо использовать многоканальную СМО, количество приборов которой должно быть больше, чем значение нагрузки: $K > y$.

В общем случае для любой СМО (с накопителем ограниченной и неограниченной емкости) загрузка системы может быть рассчитана через нагрузку следующим образом:

$$\rho = \min\left(\frac{(1-\pi_n)y}{K};1\right), \tag{1.46}$$

где K – число обслуживающих приборов в СМО; π_n – вероятность потери заявок.

Последнее выражение можно трактовать следующим образом: $\rho = \dfrac{(1-\pi_n)y}{K}$, если СМО работает без перегрузки, и $\rho = 1$, если СМО перегружена.

Покажем, что выражение (1.46) соответствует определению (1.37).

Рассмотрим достаточно большой промежуток времени $T \to \infty$, в течение которого работает СМО. За это время в систему поступит в среднем λT заявок, где λ – интенсивность поступления заявок в СМО, из которых будут обслужены системой $(1-\pi_n)\lambda T$ заявок ($\pi_n \lambda T$ заявок будут потеряны из-за ограниченной емкости накопителя). Обслуживание всех этих заявок будет длиться в течение времени $T_p = (1-\pi_n)\lambda Tb$, если СМО – одноканальная, и в течение времени $T_p = \dfrac{(1-\pi_n)\lambda Tb}{K}$, если СМО многоканальная и содержит K обслуживающих приборов. Здесь b – средняя длительность обслуживания заявки в приборе.

Подставляя выражение для T_p в (1.37), получим

$$\rho = \lim_{T\to\infty}\frac{T_p}{T} = \lim_{T\to\infty}\frac{(1-\pi_n)\lambda Tb}{TK} = \frac{(1-\pi_n)\lambda b}{K} = \frac{\lambda' b}{K}, \tag{1.47}$$

где $\lambda' = (1-\pi_n)\lambda$ – интенсивность обслуженных в СМО заявок.

Отметим, что загрузка системы, в отличие от нагрузки, определяется через интенсивность только *обслуженных* заявок, поскольку потерянные заявки не обслуживаются в приборах и, следовательно, не загружают систему.

Рассмотрим теперь СМО с накопителем неограниченной емкости и вспомним, что при возникновении перегрузок такая система не справляется с работой, что выражается в неограниченном росте очереди с течением времени.

Если $T_p < T$, то это означает, что система справляется с работой, т.е. работает без перегрузок.

Если же время $T_p = \lambda Tb / K$, которое требуется для обслуживания всех заявок, окажется больше, чем время наблюдения за системой $T_p > T$, то это означает, что система не справляется с нагрузкой, т.е. работает в режиме перегрузки. В этом случае загрузка системы $\rho = 1$ составляет 100 %, а коэффициент простоя соответственно равен нулю.

Выражение (1.46) записано с учетом указанного обстоятельства.

Получим еще одну полезную формулу для расчета вероятности потери заявок по известному значению загрузки СМО.

Из 1.41 следует, что вероятность потери заявок в СМО с накопителем ограниченной емкости может быть рассчитана как

$$\pi_n = \frac{\lambda - \lambda'}{\lambda} = 1 - \frac{\lambda'}{\lambda}.$$

В то же время из (1.47) вытекает, что интенсивность обслуженных заявок равна

$$\lambda' = \rho K / b.$$

Подставляя последнее выражение в предыдущее, получим

$$\pi_n = 1 - \frac{\rho K}{\lambda b} = 1 - \frac{\rho K}{y}, \qquad (1.48)$$

где $y = \lambda b$ – нагрузка системы.

Вероятность обслуживания поступившей в систему заявки:

$$\pi_0 = 1 - \pi_n = \frac{\rho K}{y}.$$

Выражение (1.48) оказывается полезным при расчете характеристик обслуживания заявок в марковских моделях систем и сетей массового обслуживания.

Зависимости (1.44) и (1.45), связывающие средние значения временных (w, u) и безразмерных (l, m) характеристик, известны как **формулы Литтла** и вместе с формулой (1.43) представляют собой *фундаментальные зависимости*, справедливые для широкого класса моделей массового обслуживания.

Из зависимости (1.45) можно получить зависимость, связывающую среднее число заявок в системе со средней длиной очереди заявок:

$$m = \lambda u = \lambda\left(w + b\right) = \lambda w + \lambda b = l + y,$$

откуда следует, что *нагрузка* $y = \lambda b$ *характеризует среднее число заявок, находящихся на обслуживании*.

При условии отсутствия перегрузок в одноканальной СМО загрузка совпадает с нагрузкой: $\rho = y = \lambda b$, тогда $m = l + \rho$, т.е. загрузку одноканальной СМО можно трактовать как среднее число заявок, находящихся на обслуживании в приборе. Отметим, что на обслуживании находится не одна заявка, как может показаться, а меньше единицы: $\rho < 1$. Это действительно так, если вспомнить, что речь идет о *среднем* числе находящихся на обслуживании заявок, которое может быть рассчитано следующим образом. В приборе в каждый момент времени может находиться случайное число заявок, принимающее два значения: 1, если прибор работает, то есть обслуживает заявку, и 0, если прибор простаивает. Поскольку значение загрузки лежит в интервале от 0 до 1 ($0 \le \rho \le 1$) и показывает долю времени, в течение которого прибор работает, то *загрузку можно трактовать как вероятность того, что прибор работает*, а величину $\eta = (1 - \rho)$ – как вероятность простоя прибора. Тогда математическое ожидание случайной величины, принимающей значения 1 с вероятностью ρ и 0 с вероятностью $(1 - \rho)$, будет равно: $\rho \cdot 1 + (1 - \rho) \cdot 0 = \rho$, что и требовалось показать.

Обычно исследование систем проводится в предположении о стационарности входящего потока заявок и длительности обслуживания. В этом случае *условие существования установившегося режима* для СМО с накопителем неограниченной емкости совпадает с *условием отсутствия перегрузок* в СМО и записывается в виде: $\rho < 1$.

1.8. Характеристики систем массового обслуживания с неоднородным потоком заявок

Для СМО с неоднородным потоком заявок, в которую поступают H классов заявок с интенсивностями $\lambda_1,\dots,\lambda_H$ средними длительностями обслуживания $b_1,\dots,b_H$, определяются две группы характеристик обслуживания заявок:

– характеристики по каждому классу (потоку) заявок;

– характеристики объединенного (суммарного) потока заявок.

Характеристики по каждому классу заявок $i=1,\dots,H$ идентичны характеристикам СМО с однородным потоком:

– *нагрузка*, создаваемая заявками класса i: $y_i = \dfrac{\lambda_i}{\mu_i} = \lambda_i b_i$;

– *вероятность потери* заявок: π_{n_i};

– *вероятность обслуживания* заявки: $\pi_{n_0} = 1 - \pi_{n_i}$;

– *интенсивность* потока *обслуженных* заявок (производительность по i-му классу заявок):

$$\lambda_{0_i} = \pi_{0_i}\lambda_i = (1-\lambda_{n_i})\lambda_i;$$

– *интенсивность* потока *потерянных* заявок: $\lambda_{n_i} = \pi_{n_i}\lambda_i$;

– *загрузка* системы, создаваемая заявками класса i:

$$\rho_i = \min\left\{\frac{(1-\pi_{n_i})y_i}{K}, 1\right\},$$

где π_{n_i} – вероятность потери заявок класса i из-за ограниченной емкости накопителя (π_{n_i} = 0, если емкость накопителя неограниченная); K – число обслуживающих приборов в СМО;

– *время ожидания* заявок в очереди: w_i;

– *время пребывания* заявок в системе: $u_i = w_i + b_i$;

– *длина очереди* заявок: $l_i = \lambda_i w_i = L_{\text{оч}}$;

– *число заявок в системе* (в очереди и на обслуживании): $m_i = \lambda_i u_i = L_{\text{сис}}$.

Характеристики объединенного (суммарного) потока заявок позволяют определить усредненные по всем классам заявок показатели эффективности функционирования СМО:

– *суммарная интенсивность* поступления заявок в систему (интенсивность суммарного потока):

$$\Lambda = \sum_{i=1}^{H} \lambda_i; \tag{1.49}$$

– *суммарная нагрузка Y* и *суммарная загрузка R* системы:

$$\mathrm{Y} = \sum_{i=1}^{H} \lambda; \ \mathrm{R} = \min(\sum_{i=1}^{H} \rho_i, 1), \tag{1.50}$$

причем условие отсутствия перегрузок в СМО с неоднородным потоком заявок и накопителем неограниченной емкости имеет вид

$$R < 1; \tag{1.51}$$

– *коэффициент простоя* системы: $\eta = 1 - R$;

– *среднее время ожидания W* и *среднее время пребывания U* заявок объединенного потока в системе:

$$W_i = \sum_{i=1}^{H} \xi_i w_i, \quad U_i = \sum_{i=1}^{H} \xi_i u_i, \tag{1.52}$$

где $\xi_i = \lambda_i / \Lambda$ – коэффициент, учитывающий долю заявок класса i в суммарном потоке, который может трактоваться как *вероятность того, что поступившая в систему заявка принадлежит классу i*;

– *суммарная длина очереди* и *суммарное число заявок в системе*:

$$L = \sum_{i=1}^{H} l_i, \ M = \sum_{i=1}^{H} m_i. \tag{1.53}$$

Можно доказать, что для характеристик объединенного (суммарного) потока справедливы те же фундаментальные соотношения (1.43)–(1.45), что и для однородного потока:

$$U = W + B; \ L = \Lambda W; \ M = \Lambda U,$$

где B – среднее время обслуживания любой заявки суммарного потока:

$$B = \sum_{i=1}^{H} \xi_i b_i.$$

Задание 4

Длительность обслуживания заявок в СМО распределена по экспоненциальному закону. Для заданной интенсивности обслуживания заявок μ определить *вероятность* того, что длительность обслуживания заявок будет больше величины τ.

Вариант	1	2	3	4	5	6	7	8	9	10
μ, с$^{-1}$	0,5	0,5	1,0	1,0	2,0	2,0	2,5	2,5	4,0	5,0
τ, с	3,0	2,0	2,0	0,5	1,0	2,0	0,5	1,0	0,5	1,0

Задание 5

В систему поступает простейший поток заявок с интенсивностью λ. Рассчитать:

а) *математическое ожидание, дисперсию, среднеквадратическое отклонение и коэффициент вариации* интервала времени между соседними заявками в потоке;

б) *математическое ожидание, дисперсию, среднеквадратическое отклонение и коэффициент вариации* числа заявок, поступающих в систему за время τ.

Вариант	1	2	3	4	5	6	7	8	9	10
λ, с$^{-1}$	0,2	0,4	0, 5	0,8	2,0	2,5	4,0	5,0	10	20
τ, с	5,0	4,0	2,0	1,0	1,0	2,0	4,0	5,0	0,5	1,0

Задание 6

В систему с интенсивностью поступает поток заявок, интервалы между которыми распределены по закону Эрланга k-го порядка. Рассчитать *математическое ожидание, дисперсию, среднеквадратическое отклонение и коэффициент вариации* интервалов времени между соседними заявками в потоке.

Вариант	1	2	3	4	5	6	7	8	9	10
Λ, с	0,2	0,4	0,5	0,8	2,0	2,5	4,0	5,0	10	20
k	3	2	4	9	16	2	4	9	16	2

1.9. Пуассоновский неординарный поток

Пусть события появляются в моменты $t_1, t_2, t_3, \ldots$ «пачками», так что в момент t_k появляется сразу η_k событий. Будем считать, что η_k есть последовательность независимых одинаково распределенных случайных величин с рапределением

$$P\{\eta_k = n\} = f_n.$$

Такой поток называется простым (или пуассоновским) неординарным (или групповым) потоком однородных событий.

Найдем распределение вероятностей $P_k(t)$ числа событий потока, наступивших на интервале длиной t. Пусть ν_t – число событий наступивших на этом интервале, N_t – число моментов времени, в которые наступали группы событий, т.е.

$$N_t = \max\{n : 0 < t_n < t\}.$$

Тогда

$$\nu_t = \sum_{k=1}^{N_t} \eta_k.$$

Введем производящую функцию:

$$\varphi_t(z) = M\{z^{\nu_t}\} = \sum_{n=0}^{\infty} P\{N_t = n\} M\{z^{\nu_t} | N_t = n\} =$$

$$= \sum_{n=0}^{\infty} P\{N_t = n\} M\{z^{\eta_1 + \eta_2 + \ldots + \eta_n}\} = \sum_{n=0}^{\infty} P\{N_t = n\} M\{z^{\eta_i}\},$$

так как все η_i независимы и одинаково распределены.

Введем функцию

$$f(z) = M\{z^{\eta_i}\} = \sum_{n=0}^{\infty} f_k z^k \ ,$$

имеющую смысл производящей функции числа событий в группе, тогда

$$\varphi_t(z) = \sum_{n=0}^{\infty} \frac{(\lambda t^n)}{n!} e^{-\lambda t} [f(z)]^n = \exp\{\lambda t [f(z) - 1]\}. \qquad (1.54)$$

Вероятности $P_k(t)$ можно определить из формул

$$P_k(t) = \frac{1}{k!} \frac{\partial^k \varphi_t(z)}{\partial z^k} |_{z=0}, \ k \ge 1.$$

Найдем еще среднее число событий, наступивших в интервале $(0,t)$. Имеем

$$M\{v_t\} = \sum_{k=0}^{\infty} kP_k(t) = \left.\frac{\partial\varphi_i(z)}{\partial z}\right|_{z=1} = \{\lambda t f'(z)\exp\{\lambda t[f(z)-1]\}\}\big|_{z=1} =$$
$$= \lambda t f'(1) = \lambda t\overline{k},$$

где $\overline{k} = \sum_{k=0}^{\infty} kf_k$.

Отсюда интенсивность такого потока равна

$$\frac{M\{v_t\}}{t} = \lambda\overline{k},$$

где $\overline{k}$ – среднее число событий в группе. Смысл этой формулы очевиден.

1.10. Простой поток разнотипных событий

Пусть $\{t_n\}$ – моменты, образующие пуассоновский поток событий интенсивности λ, а σ_n – независимые одинаково распределенные случайные величины, принимающие значения из множества целых чисел $\{1, 2, \ldots, k\}$.

Если $\sigma_n = \sigma(t_n) = i$, то в момент t_n происходит событие типа i. Обозначая $p_i = P\{\sigma_n = i\}$, мы получим простой поток разнотипных событий.

Обозначая через $v_i(t)$ число событий типа i на интервале $(0;t)$, определим производящую функцию

$$\varphi_i(z_1, z_2, \ldots, z_k) = M\left\{z_1^{v_1(t)}, z_2^{v_2(t)}, \ldots, z_k^{v_k(t)}\right\}.$$

Для вычисления явного вида этой функции введем величины

$$\sigma_{ni} = \begin{cases} 1, & \text{если } \sigma_n = i, \\ 0, & \text{если } \sigma_n \neq i. \end{cases}$$

Тогда $v_i(t) = \sum_n \sigma_{ni}$, и мы имеем

$$\varphi_i(z_1, z_2, \ldots, z_k) = \sum_{m=0}^{\infty} P\{N_i = m\} M\left\{\prod_{i=1}^{k} z_i^{\sum_{i=1}^{m}\sigma_{ni}}\right\} =$$
$$= \sum_{m=0}^{\infty} P\{N_i = m\}\left\{M\{z_1^{\sigma_{11}} z_2^{\sigma_{12}} \ldots z_k^{\sigma_{1k}}\}\right\}^m,$$

где учтено, что в разные моменты времени величины σ_n независимы и одинаково распределены. Так как $\sigma_{n\,i}$ принимают значения 0 или 1 и значение 1 принимает только одна из них, то

$$M\{z_1^{\sigma_{11}} z_2^{\sigma_{12}} \ldots z_k^{\sigma_{1k}}\} = p_1 z_1 + p_2 z_2 + \ldots + p_k z_k,$$

поэтому

$$\varphi_i(z_1, z_2, \ldots, z_k) = \sum_{m=0}^{\infty} \frac{(\lambda t)^m}{m!} e^{-\lambda t} \left[p_1 z_1 + p_2 z_2 + \ldots + p_k z_k \right]^m =$$

$$= \exp\left\{\lambda t[-1 + (p_1 z_1 + p_2 z_2 + \ldots + p_k z_k)]\right\} = \prod_{i=1}^{k} e^{-\lambda i p_i (1 - z_i)}. \quad (1.55)$$

Отсюда следует, что $\nu_i(t)$ есть независимые пуассоновские потоки событий с параметрами λp_i, и мы наблюдаем сумму (суперпозицию) этих потоков.

1.11. Потоки восстановления

Пусть на временной оси некоторый момент времени выбран за 0 и пусть $t_1, t_2, t_3, \ldots, t_i > 0$ – моменты наступления событий (рис. 1.22).

Рис. 1.22. Поток событий

Обозначим через $\tau_1 = t_1$, $\tau_n = t_n - t_{n-1}$ длины интервалов между моментами наступления событий. Отметим принципиальное отличие τ_1 от всех остальных τ_n: если τ_n, $n \geq 2$, отсчитываются от момента наступления предыдущего события потока, то τ_1 отсчитывается от нуля на оси времени, никакого отношения к потоку не имеющего.

Случайный поток называется потоком с ограниченным последствием, если $\{\tau_n\}$ – независимые случайные величины.

Смысл слов «ограниченное последствие» следующий: будущее поведение потока после момента времени t, $t_n < t < t_{n+1}$, зависит лишь от конечной величины $(t - t_n)$, а от остального прошлого не зависит.

Для описания потока с ограниченным последействием достаточно задать функции распределения $A_k(x) = P(\tau_k < x),\ k \geq 1$. Поток с ограниченным последействием, для которого $A_k(x) = A(x),\ k \geq 2$ (заметьте: $A_1(x)$ это не касается), называется потоком **восстановления**. В последнем случае вместо слов «наступило событие потока» часто говорят «наступило восстановление».

Пусть N_t – число событий, наступивших на интервале $(0,t)$. Функция $H(t)=M\{N_t\}$ называется **функцией восстановления**. Функция $h(t)=H'(t)$, если она существует, называется **плотностью восстановления**. По смыслу $h(t)dt$ (или $dH(t)$) есть вероятность того, что на интервале $[t, t+dt]$ наступит событие потока (интенсивность равна параметру потока).

Выведем теперь одно из основных уравнений – уравнение для $H(t)$. Имеем

$$P\{N_t = n\} = P\left\{t_n < t, t_{n+1} \geq t\right\}.$$

Но, с другой стороны, $P\{t_n < t\} = P\{t_{n+1} < t\} + P\{t_n < t, t_{n+1} > t\}$, откуда $P\left\{t_n < t, t_{n+1} > t\right\} = P\left\{t_n < t\right\} - P\left\{t_{n+1} < t\right\}$.

Поэтому для $H(t) = M\{N_t\}$ имеем

$$N(t) = M\{N_t\} = \sum_{n=1}^{\infty} nP\left\{N_t = n\right\} =$$
$$= 1[P\{t_1 < t\} - P\{t_2 < t\}] + 2[P\{t_2 < t\} - P\{t_3 < t\}] +$$
$$+ 3[P\{t_3 < t\} - P\{t_4 < t\}] + \ldots = \sum_{n=1}^{\infty} P\{t_n < t\}.$$

Так как $t_n = \tau_1 + \tau_2 + \tau_3 + \ldots \tau_n$, то

$$P\{t_1 < t\} = F_1(t) + A_1(t),$$
$$P\{t_2 < t\} = F_2(t) = P(t_1 + \tau_2 < t) = \int_0^t F_1(t-x)dA(x),$$
$$P\{t_3 < t\} = F_3(t) = P(t_2 + \tau_3 < t) = \int_0^t F_2(t-x)dA(x),$$
$$\ldots$$
$$P\{t_n < t\} = F_n(t) = P(t_{n-1} + \tau_n < t) = \int_0^t F_{n-1}(t-x)dA(x).$$

Поэтому

$$H(t)=\sum_{n=1}^{\infty}F_n(t)=A_1(t)+\sum_{n=2}^{\infty}\int_0^t F_{n-1}(t-x)dA(x)=$$

$$=A_1(t)+\int_0^t\Bigl(\sum_{n=2}^{\infty}F_{n-1}(t-x)dA(x)\Bigr)=A_1(t)+\int_0^t H(t-x)dA(x).$$

Итак, функция восстановления удовлетворяет интегральному уравнению

$$H(t)=A_1(t)+\int_0^t H(t-x)dA(x), \qquad (1.56)$$

решить которое можно с помощью преобразования Лапласа – Стилтьеса. Итак, $H^*(s)=A_1^*(s)+H^*(s)\cdot A^*(s)$, откуда

$$H^*(s)=\frac{A_1^*(s)}{1-A^*(s)}. \qquad (1.57)$$

Зная $A^*(s)$ и $A_1^*(s)$, можно найти $H^*(s)$. А по таблицам обратного преобразования Лапласа – Стилтьеса и саму функцию $H(s)$.

Пример 8. Пусть $A(x)=1-e^{-\lambda x}$, $A_1(x)=1-e^{-\mu x}$, тогда

$$A^*(s)=\int_0^{\infty}e^{-sx}d(1-e^{-\lambda x})=\lambda\int_0^{\infty}e^{-sx-\lambda x}dx=\frac{\lambda}{\lambda+s},\ \ A_1^*=\frac{\mu}{\mu+s},$$

Поэтому

$$H^*(s)=\frac{\mu/(\mu+s)}{1-\lambda/(\lambda+s)}=\frac{\mu(\lambda+s)}{s(\mu+s)}=\frac{\lambda}{s}+\frac{\mu-\lambda}{\mu+s},$$

и отсюда имеем

$$H(t)=\lambda t+\frac{\mu-\lambda}{\mu}(1-e^{-\mu t}).$$

Рассмотрим теперь некоторые следствия из уравнения (1.56) и его решения (1.57).

1. Пусть $a=\int_0^{\infty}xdA(x)$ и $\lambda=1/a$. Заметим, что $A^*(0)=A(\infty)-A(0)=1.$

Тогда

$$\lim_{t\to\infty}\frac{H(t)}{t}=\lim_{s\to 0} sH^*(s)=\lim_{s\to 0}\frac{A_1^*(s)s}{1-A^*(s)}=$$

$$=\frac{\lim\limits_{s\to 0}A_1^*(s)}{\lim\limits_{s\to 0}\dfrac{A^*(0)-A^*(s)}{s}}=\frac{1}{-A^{*\prime}(0)}=\frac{1}{a}=\lambda.$$

То есть при $t\to\infty$ функция $H(t)$ ведет себя как λt.

Такой же результат можно получить, используя преобразование Лапласа.

2. Пусть $Q(t)$ – неотрицательная интегрируемая на $(0, \infty)$ функция. Тогда по свойствам преобразования Лапласа – Стилтьеса имеем

$$\int_0^t Q(t-x)dH(x)\Leftrightarrow Q^*(s)H^*(s)=\frac{A_1^*(s)Q(s)}{1-A^*(s)}.$$

По свойствам преобразования Лапласа – Стилтьеса

$$\lim_{t\to\infty}\int_0^t Q(t-x)dH(x)=\lim_{s\to 0}Q^*(s)H^*(s)=$$

$$=\lim_{s\to 0}\frac{Q^*(s)}{s}\lim_{s\to 0}A_1^*(s)\lim_{s\to 0}\frac{s}{1-A^*(s)}=\frac{1}{a}\int_0^\infty Q(x)dx.$$

Окончательно получаем

$$\lim_{t\to\infty}\int_0^t Q(t-x)dH(x)=\lambda\int_0^\infty Q(x)dx,\ \lambda=\frac{1}{a}.$$

Эта формула называется основной теоремой теории восстановления.

1.12. Распределение величины перескока и недоскока для потоков восстановления

Рассмотрим некоторый произвольный момент времени t в потоке восстановления, который окружают моменты t_n и t_{n+1} наступления событий потока. Величина $t_{n+1}-t=\gamma(t)$ называется временем (или величиной) перескока, а величина $\gamma^* = t - t_n$ – величиной недоскока (рис. 1.23).

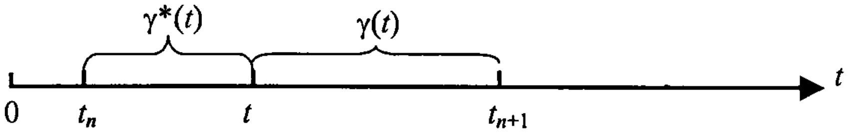


Рис. 1.23. Перескок и недоскок

Пусть $F(t,x) = P\{\gamma(t) < x\}$. Выведем уравнение для $F(t,x)$.

Возможны следующие варианты:

1. До момента t вообще не наступило ни одного события (рис. 1.24).

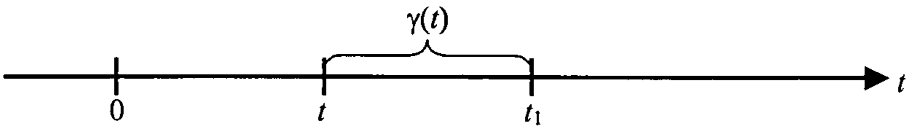


Рис. 1.24. До момента *t* не наступило ни одного события

Это будет тогда, когда $t_1 > t$. Так как при этом $\gamma(t) = t_1 - t$, то событие $\gamma(t) < x$ эквивалентно $t_1 < t + x$. Итак, в этом случае

$$P\{\gamma(t) < x\} = P\{t < t_1 < t + x\} = A_1(t + x) - A_1(t).$$

2. Пусть $t_n < t < t_{n+1}$ (рис. 1.25).

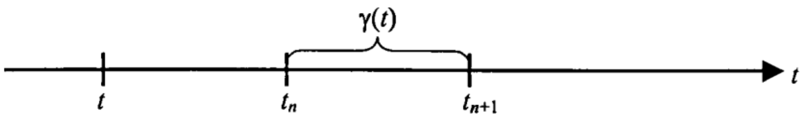


Рис. 1.25. Случай $t_n < t < t_{n+1}$

В этом случае должно быть, с одной стороны, $\tau > t - t_n$, где $\tau = t_{n+1} - t_n$ – длина интервала между моментами наступления событий в потоке восстановления. Далее, так как $\tau = (t - t_n) + \gamma(t)$, то событие $\gamma(t) < x$ эквивалентно $\tau < (t - t_n) + x$. Итак, окончательно должно выполняться условие

$$t - t_n < \tau < t - t_n + x.$$

Так как n может быть любым в интервале значений $1, \ldots, \infty$, то вероятность этой ситуации равна $(t_n \equiv s)$:

$$\sum_{n=1}^{\infty}\int_0^t [A(t-s+x)-A(t-s)]dF_n(s)=\int_0^t [A(t-s+x)-A(t-s)]dH(s).$$

Складывая эти два выражения, получим

$$F(t,x)=A_1(t+x)-A_1(t)+\int_0^t \left[A(t-s+x)-A(t-s)\right]dH(s). \quad (1.58)$$

Найдем $F(x)=\lim\limits_{t\to\infty}F(t,x)$, предполагая, что этот предел существует, тогда

$$\lim_{t\to\infty}[A_1(t+x)-A_1(t)]=A_1(\infty)-A_1(\infty)=1-1=0.$$

Далее, представляя второе слагаемое в (1.58) в виде

$$\int_0^t [A(t-s+x)-A(t-s)]dH(s)=\int_0^t [1-A(t-s)]dH(s)-$$

$$-\int_0^t [1-A(t-s+x)]dH(s)$$

(введение 1 необходимо для того, чтобы подынтегральная функция была интегрируемой на $(0,\infty)$) и используя основную теорему теории восстановления, получим

$$F(x)=\lambda\int_0^\infty [1-A(z)]dz-\lambda\int_0^\infty [1-A(z+x)]dz=[z+x=u]=$$

$$=\lambda\int_0^\infty [1-A(z)]dz-\lambda\int_x^\infty [1-A(u)]du\ .$$

Окончательно имеем следующую важную формулу:

$$F(x)=\lambda\int_0^x [1-A(z)]dz\ ,\ \lambda=\frac{1}{a}\ , \quad (1.59)$$

дающую функцию распределения величины перескока в асимптотическом случае $t\to\infty$.

Рассмотрим теперь величину недоскока $\gamma^*(t)$ и введем функцию $F^*(t,x)=P\{\gamma^*(t)<x\}$. Выведем уравнение для этой функции. Опять-таки возможны два варианта.

1. До момента t не наступило ни одного события. Тогда $\gamma^*(t)=t$.

а) Если взять $x<t$ (рис. 1.26), то получим невозможную комбинацию, так как $\gamma^*(t) = t > x$, а надо найти $P\{\gamma^*(t) = t < x\}$. Поэтому вероятность такой ситуации равна нулю.

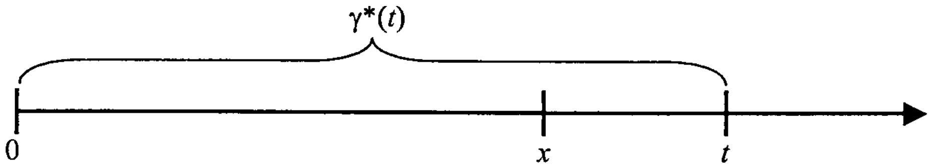


Рис. 1.26. Случай $x<t$

б) Если взять $x>t$ (рис. 1.27), то рассматриваемая ситуация наступит в том случае, когда $\tau_1 > t$. Вероятность этого равна $1-A_1(t)$.

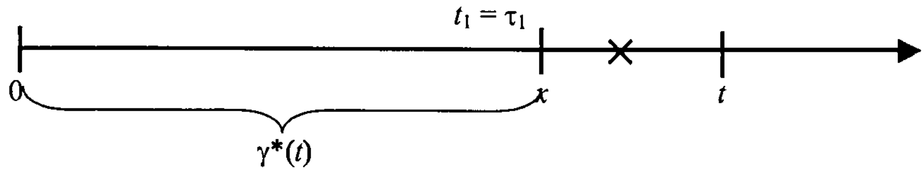


Рис. 1.27. Случай $x>t$

Итак, искомая вероятность данной ситуации равна

$$F_1(t,x) = \begin{cases} 1-A_1(t), & \text{если } t<x, \\ 0, & \text{если } t>x. \end{cases}$$

2. Пусть $t_n < t < t_{n+1}$ (рис. 1.28).

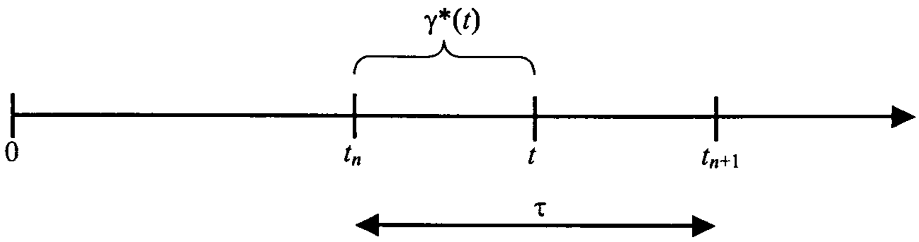


Рис. 1.28. Случай $t_n < t < t_{n+1}$

Тогда должно быть:

1) $\tau > t - t_n$;

2) $t - t_n = \gamma^*(t) < x$.

Отсюда следует, что t_n лежит в интервале $t-x < t_n < t$. Так как вероятность события $\tau > t - t_n$ равна $1-A(t-t_n)$, то вероятность этой ситуации равна $(t_n \equiv s)$:

$$\sum_{n=1}^{\infty}\int_{t-x}^{t}[1-A(t-s)]dF_n(s)=\int_{t-x}^{t}\left[1-A(t-s)\right]dH(s).$$

Окончательно получаем

$$F^*\left(t,x\right)=F_1(t,x)+\int_0^t[1-A(t-s)]dH(s)-\int_0^{t-x}[1-A(t-s)]dH(s)\ . \quad (1.60)$$

Найдем $F^*(x)=\lim\limits_{t\to\infty}F^*(t,x)$, предполагая, что предел существует. Легко видеть, что $\lim\limits_{t\to\infty}F_1(t,x)=0$. Далее по основной теореме теории восстановления получаем

$$\lim_{t\to\infty}\int_0^t[1-A(t-s)]dH(s)=\lambda\int_0^{\infty}[1-A(z)]dz.$$

$$\lim_{t\to\infty}\int_0^{t-x}[1-A(t-s)]dH(s)=\lim_{(t-x)\to\infty}\int_0^{t-x}[1-A(t-x)+x-s]dH(s)=$$

$$\lim_{T\to\infty}\int_0^{T}[1-A(T+x-s)]dH(s)=\lambda\int_0^{\infty}[1-A(u+x)]du=\lambda\int_x^{\infty}[1-A(z)]dz$$

Окончательно имеем

$$F^*(x)=\lambda\int_0^{\infty}[1-A(z)]dz-\lambda\int_x^{\infty}[1-A(z)]dz=\lambda\int_0^{x}[1-A(z)]dz\ ,$$

то есть $F^*(x)=F(x)$, определяемое формулой (1.59).

Заметим, что распределение величины перескока и недоскока можно получить и другим способом, используя тот факт, что процессы $\gamma(t)$ и $\gamma^*(t)$ являются марковскими.

Очевидны следующие выкладки.

Для процесса $\gamma(t)$, характеризующего величину перескока, обозначим $P\{\gamma(t)<x\}=F(t,x)$, тогда

$$F(t+\Delta t,x)=F(t,x+\Delta t)-F(t,\Delta t)+F(t,\Delta t)A(x)+o(\Delta t).$$

Разложим функцию F по приращению аргументов в ряд Тейлора с точностью до *o*(Δt)*:*

$$F(t,x)+\Delta t\frac{\partial F(t,x)}{\partial t}=F(t,x)+\Delta t\frac{\partial F(t,x)}{\partial x}-$$

$$-\Delta t\frac{\partial F(t,0)}{\partial x}+\Delta t\frac{\partial F(t,0)}{\partial x}A(x)+o(\Delta t).$$

Приводим подобные и сокращаем на Δt:

$$\frac{\partial F(t,x)}{\partial t}=\frac{\partial F(t,x)}{\partial x}-\frac{\partial F(t,0)}{\partial x}+\frac{\partial F(t,0)}{\partial x}A(x)=$$

$$=\frac{\partial F(t,x)}{\partial x}-\frac{\partial F(t,0)}{\partial x}(1-A(x)).$$

В стационарном режиме $F(t,x)\equiv F(x)$, тогда $F'(x)=F'(0)(1-A(x))$, следовательно,

$$F(x)=F'(0)\int_0^x(1-A(s))ds.$$

Полагая $x\to\infty$, получим

$$1=F'(0)\int_0^\infty(1-A(s))ds=F'(0)a,$$

откуда

$$F'(0)=\frac{1}{a}=\lambda.$$

Таким образом, функция распределения $F(x)$ величины перескока имеет вид

$$F(x)=\lambda\int_0^x(1-A(s))ds.$$

Для процесса $\gamma^*(t)$, характеризующего величину недоскока, обозначим

$$P\{\gamma^*(t)<x\}=G\left(t,x\right),\quad P\{x\le\gamma^*(t)<x+dx\}=g(t,x)dx,$$

тогда

$$g(t+\Delta t,x+\Delta t)=g(x,t)P\{\tau>x+\Delta t|\tau>x\}=g(x,t)\frac{P\{\tau>x+\Delta t\}}{P\{\tau>x\}}=.$$

$$=g(x,t)\frac{1-A(x+\Delta t)}{1-A(x)}=g(x,t)\frac{1-A(x)+\Delta tA'(x)}{1-A(x)}=g(x,t)\left(1-\Delta t\frac{A'(x)}{1-A(x)}\right).$$

и, следовательно,

$$\frac{\partial g(t,x)}{\partial t}+\frac{\partial g(t,x)}{\partial x}=-\frac{A'(x)}{1-A(x)}g(t,x).$$

В стационарном режиме $g(t,x) \equiv g(x)$, поэтому $g(x)$ удовлетворяет уравнению

$$g'(x) = -\frac{A'(x)}{1-A(x)} g(x),$$

решение которого имеет вид $g(x) = C(1-A(x))$.

Интегрируя это равенство по x в интервале $0 \le x < \infty$, получим

$$1 = C\int_0^\infty (1-A(s))ds = Ca,$$

откуда следует, что $C = \frac{1}{a} = \lambda$.

Таким образом,

$$g(x) = \lambda(1 - A(x)),$$

а для функции распределения $G(x)$ величины недоскока можно записать как

$$G(x) = \lambda\int_0^x (1-A(s))ds.$$

1.13. Парадокс остаточного времени

Пусть поток трамваев есть поток событий восстановления с функцией распределения интервалов между трамваями $A(x)$. В случайный момент времени t человек приходит на остановку. Сколько в среднем времени ему ждать трамвая?

1. Первое решение. Так как в среднем длина интервала времени между трамваями равна $a = \int_0^\infty x dA(x) dx$ и человек приходит совершенно произвольно, то ему ждать в среднем $a/2$.

2. Второе решение. Время ожидания трамвая – это время перескока, распределенное с функцией распределения $\frac{1}{a}\int_0^x (1-A(z))dz$, т.е. с плотностью вероятностей $\frac{1-A(x)}{a}$. Поэтому ждать ему в среднем

$$\frac{1}{a}\int_0^\infty x(1-A(x))dx = \frac{1}{2a}\int_0^\infty \left(1-A(x)\right)d(x^2) =$$

$$= \frac{1}{2a}x^2[1-A(x)]\Big|_0^\infty + \frac{1}{2a}\int_0^\infty x^2\,dA(x) = \tag{1.61}$$

$$= \frac{M\{x^2\}}{2a} = \frac{a^2+\sigma_a^2}{2a} = \frac{a}{2} = \frac{\sigma_a^2}{2a} > \frac{a}{2}.$$

Как видно, два решения дали два разных результата, причем второе решение дает большее среднее время ожидания, чем первое. Какой же результат верен? Верен второй результат. Разгадка в том, что промежутки времени между приходами трамваев разные. И, приходя в произвольный момент времени, человек с большей вероятностью попадает на более длинный промежуток времени между трамваями, чем на более короткий.

1.14. Основное свойство рекуррентных потоков

Определение. Стационарный поток восстановления называется рекуррентным.

Теорема. Если поток восстановления является стационарным, т.е. является рекуррентным и $a = \int_0^\infty x dA(x) dx < +\infty,$ то

$$A_1(x) = 1/a\int_0^x ((1-Az(z))dz. \tag{1.62}$$

Доказательство. Во-первых, докажем, что если поток стационарен, то $H\,(t)$ – линейная по t функция.

Так как поток стационарен, то величины $N_{t+s} - N_s$ $N_t - N_0 = N_t$ распределены одинаково, тогда

$$H(t+s) = M\{N_{t+s}\} = M\{N_{t+s} - N_s + N_s - N_0\} =$$
$$= M\{N_{t+s} - N_s\} + M\{N_s - N_0\} = M\{N_t\} + M\{N_s\} = H(t) + H(s).$$

Следовательно, $H(t)$ является линейной функцией вида

$$H\left(t\right) = kt.$$

а так как $\lim\limits_{t\to\infty}\frac{H(t)}{t} = \lambda,$ то $k = \lambda = \frac{1}{a}.$

Во-вторых, возьмем уравнение для функции восстановления

$$H(t) = A_1(t) + \int_0^t A(t-s)dH(s)$$

и подставим в него $H(t) = kt$, где $k = \frac{1}{a}$, тогда

$$\frac{1}{a}t = A_1(t) + \frac{1}{a}\int_0^t A(t-s)dH(s) = A_1(t) + \frac{1}{a}\int_0^t A(z)dz,$$

откуда

$$A_1(t) = \frac{1}{a}t - \frac{1}{a}\int_0^t A(t-s)ds = \frac{1}{a}\int_0^t [1-A(z)]dz,$$

или, если заменить $t = x$:

$$A_1(x) = \frac{1}{a}\int_0^x [1-A(z)]dz.$$

Таким образом, рекуррентный поток определяется единственной функцией распределения $A(x)$.

Следствие. Если $A_1(x) \neq \frac{1}{a}\int_0^x [1-A(z)]dz$, то процесс восстановления не может быть стационарным, т.е. не является рекуррентным.

Частный случай. Найдем, при каких условиях выполняется соотношение $A_1(x) = A(x)$. Тогда

$$A(x) = \frac{1}{a}\int_0^x [1-A(z)]dz = \lambda\int_0^x [1-A(z)]dz.$$

Дифференцируя по x, получим

$$A'(x) = \lambda[1-A(x)] = -[1-A(x)]' = -\frac{d[1-A(x)]}{dx}.$$

Разделяя переменные

$$\frac{d\left(1-A(x)\right)}{1-A(x)} = -\lambda dx$$

и интегрируя, получим

$$\ln(1-A(x)) = \lambda x,$$

откуда $A(x) = 1 - e^{-\lambda x}$, и поток является пуассоновским потоком с постоянной интенсивностью λ.

1.15. Сумма независимых рекуррентных потоков

Рассмотрим сумму двух независимых рекуррентных потоков, определяемых функциями $A_1(x)$ и $A_2(x)$ (рис. 1.29).

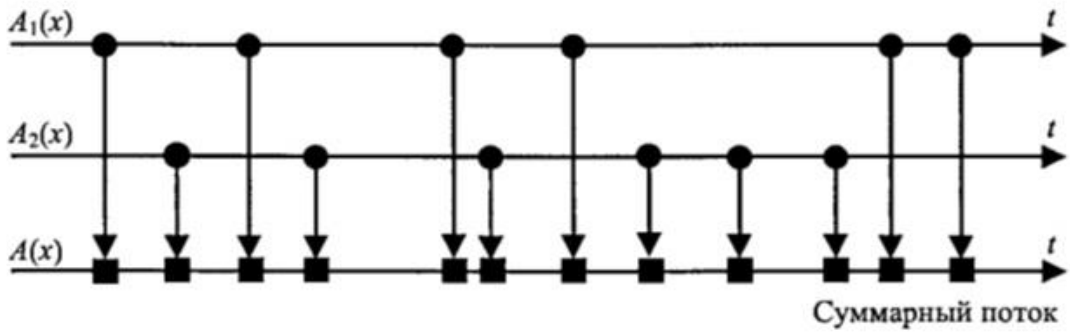


Рис. 1.29. Схема образования суммарного потока

Пусть τ – величина перескока суммарного потока. Тогда очевидно, что $\tau = \min(\tau_1, \tau_2)$, где τ_1, τ_2 – независимые величины перескоков для первого и второго суммируемых рекуррентных потоков. Поэтому

$$P(\tau > x) = P(\min(\tau_1, \tau_2) > x) = P(\tau_1 > x)P(\tau_2 > x) =$$

$$= (1 - \lambda_1 \int_0^x (1 - A_1(s))ds)(1 - \lambda_2 \int_0^x (1 - A_2(s))ds).$$

Следовательно,

$$1 - \lambda \int_0^x (1 - A(s))ds = \left(1 - \lambda_1 \int_0^x (1 - A_1(s))ds\right)\left(1 - \lambda_2 \int_0^x (1 - A_2(s))ds\right),$$

тогда

$$\lambda \int_0^x (1 - A(s))ds = \lambda_1 \int_0^x (1 - A_1(s))ds + \lambda_2 \int_0^x (1 - A_2(s))ds -$$

$$- \lambda_1 \lambda_2 \int_0^x (1 - A_1(s))ds \int_0^x (1 - A_2(s))ds.$$

Таким образом,

$$\lambda(1 - A(x)) = \lambda_1(1 - A_1(x)) + \lambda_2(1 - A_2(x)) -$$

$$-\lambda_1 \lambda_2 [(1 - A_1(x)) \int_0^x (1 - A_2(s)) + (1 - A_2(x))ds \int_0^x (1 - A_1(s))ds].$$

Так как $\lambda = \lambda_1 + \lambda_2$, то функция $A(x)$ для суммы рекуррентных потоков будет иметь вид

$$A(x) = 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2}(1 - A_1(x)) - \frac{\lambda_2}{\lambda_1 + \lambda_2}(1 - A_2(x)) +$$

$$+ \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}[(1 - A_1(x))\int_0^x (1 - A_2(s)) + (1 - A_2(x))ds\int_0^x (1 - A_1(s))ds].$$

В частности, если $A_k(x) = 1 - e^{-\lambda_k x}$, $k = 1, 2$, т.е. суммируются простейшие потоки, то

$$A(x) = 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-\lambda_1 x} - \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-\lambda_2 x} +$$

$$+ \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}\left[e^{-\lambda_1 x} \frac{1}{\lambda_2}(1 - e^{-\lambda_2 x}) + e^{-\lambda_2 x} \frac{1}{\lambda_1}(1 - e^{-\lambda_1 x}) \right] =$$

$$= 1 - \frac{1}{\lambda_1 + \lambda_2}\left[\lambda_1 e^{-\lambda_1 x} + \lambda_2 e^{-\lambda_2 x} - \lambda_1 e^{-\lambda_1 x}(1 - e^{-\lambda_2 x}) - \lambda_2 e^{-\lambda_2 x}(1 - e^{-\lambda_1 x}) \right] =$$

$$= 1 - e^{-(\lambda_1 + \lambda_2)x}.$$

Следовательно, суммой независимых простейших потоков является простейший поток с параметром, равным сумме параметров исходных потоков.

1.16. Биномиальная схема деления рекуррентного потока

Биномиальная схема деления рекуррентного потока реализуется следующим образом. Каждое событие исходного потока с заданной и постоянной вероятностью p назначается в первый, а с вероятностью $(1 - p)$ во второй разделенные потоки соответственно (рис. 1.30).

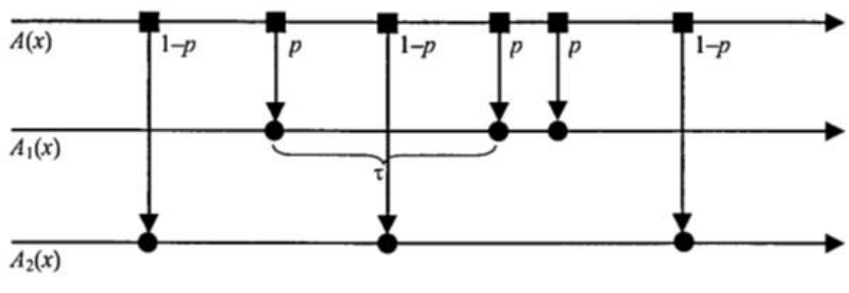


Рис. 1.30. Схема деления потока

Исходный поток определяется функцией распределения $A(x)$ длин интервалов между моментами наступления событий. Обозначим через $A_1(x)$ и $A_2(x)$ функции распределения длин интервалов между моментами наступления событий в разделенных потоках. Найдем характеристическую функцию $M\{e^{-\alpha\tau}\}$ длин интервалов для первого из разделенных потоков.

Пусть τ – длина интервала в первом из разделенных потоков, а ν – случайное число интервалов исходного потока, реализующий интервал длины τ. Очевидно:

ν	Вероятность
1	p
2	$p(1-p)$
k	$(1-p)^{k-1}p$

Получаем

$$M\{e^{-\alpha\tau}\} = \sum_{k=1}^{\infty} M\{e^{-\alpha\tau} \big| \nu = k\} P\{\nu = k\} = \sum_{k=1}^{\infty} \varphi^k(\alpha)(1-p)^{k-1} p =$$

$$= p\varphi(a) \cdot \frac{1}{1-(1-p)\varphi(a)}.$$

Обращая эту функцию, можно найти функцию распределения $A_1(x)$ для первого потока. Аналогично определяется и $A_2(x)$.

Найдем среднее значение длины интервалов в первом из разделенных потоков. Используя свойства характеристических функций и дифференцируя по α в нуле это равенство, запишем

$$-M\tau = -a_1 = p \left. \frac{\varphi'(\alpha)(1-(1-p)\varphi(\alpha)) + \varphi(\alpha)(1-p)\varphi'(\alpha)}{[1-(1-p)\varphi(\alpha)]^2} \right|_{\alpha=0} =$$

$$= p \frac{-ap-(1-p)a}{p^2} = -\frac{a}{p},$$

то есть средняя длина интервала для выделенного потока имеет вид $a_1 = \dfrac{a}{p}$. Тогда $\lambda_1 = \dfrac{1}{a_1} = \dfrac{p}{a} = \lambda p \Rightarrow \lambda_1 = \lambda p.$

Для простейшего потока с параметром λ имеем

$$\varphi(\alpha) = \int_0^{\infty} e^{-\alpha x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda+\alpha}.$$

Следовательно,

$$M\{e^{-\alpha\tau}\} = p\frac{\dfrac{\lambda}{\lambda+\alpha}}{1-(1-p)\dfrac{\lambda}{\lambda+\alpha}} = p\frac{\lambda}{\lambda+\alpha-(1-p)\lambda} = \frac{\lambda p}{\alpha+\lambda p}.$$

При биномиальной схеме деления простейшего потока каждый поток – простейший с параметром λp или $(1-p)\lambda$.

1.17. Преобразования Лапласа и Лапласа – Стилтьеса. Производящая функция

В теории массового обслуживания интенсивно используется аппарат преобразований Лапласа и Лапласа – Стилтьеса и производящих функций. Приведем основные сведения об этих функциях и преобразованиях.

Определение 1. Преобразованием Лапласа – Стилтьеса распределения $B(t)$ будем называть функцию $\beta(s)$, определяемую следующим образом:

$$\beta(s) = \int_0^\infty e^{-st} dB(t),$$

а преобразованием Лапласа – функцию $\varphi(s)$, определяемую как

$$\varphi(s) = \int_0^\infty e^{-st} B(t) dt.$$

Если s есть чисто мнимая переменная, преобразование Лапласа – Стилтьеса совпадает с характеристической функцией распределения $B(t)$. Областью определения функций $\beta(s), \varphi(s)$ обычно считается правая полуплоскость комплексной плоскости. Однако без существенного ограничения общности, в рамках данной главы и книги можно рассматривать s как действительное положительное число.

Отметим некоторые из свойств преобразования Лапласа – Стилтьеса.

Свойство 1. Если оба преобразования $\beta(s)$ и $\varphi(s)$ существуют (т.е. соответствующие несобственные интегралы сходятся), то они связаны между собой следующим образом:

$$\beta(s) = s\varphi(s).$$

Свойство 2. Если две независимые случайные величины имеют преобразования Лапласа – Стилтьеса, а $\beta_1(s)$ и $\beta_2(s)$ – их функции распределения, то преобразованием Лапласа – Стилтьеса функции распределения суммы этих величин является $\beta_1(s)\,\beta_2(s)$.

Свойство 3. Преобразованием Лапласа – Стилтьеса производной $B'(t)$ функции $B(t)$ является $s\beta(s) - sB(+0)$.

Свойство 4. $\lim\limits_{s\to 0}\beta(s) = \lim\limits_{t\to\infty} B(t)$.

Свойство 5. Пусть b_k есть k-й начальный момент распределения:

$$b_k = \int_0^\infty t^k dB(t), \quad k \ge 1.$$

Он вычисляется через преобразование Лапласа – Стилтьеса следующим образом:

$$b_k = (-1)^k \left.\frac{d^k\beta(s)}{ds^k}\right|_{s=0}. \tag{1.63}$$

Свойство 6. Преобразованию Лапласа – Стилтьеса $\beta(s)$ может быть придан вероятностный смысл следующим образом. Считаем, что $B(t)$ есть функция распределения длины некоторого интервала времени и в этом интервале времени поступает простейший поток «катастроф» с параметром $s > 0$. Тогда легко видеть, что $\beta(s)$ есть вероятность того, что за интервал не наступит ни одна «катастрофа».

Свойство 7. Преобразование Лапласа – Стилтьеса $\beta(s)$, рассматриваемое как функция действительной переменной $s > 0$, является вполне монотонной функцией, т.е. оно имеет производные $\beta^{(n)}(s)$ всех порядков и $(-1)^n\beta^{(n)}(s) \ge 0$, $s > 0$.

Определение 2. Производящей функцией распределения вероятностей q_k, $k \ge 0$, дискретной случайной величины ξ называется функция

$$Q(z) = M(z^\xi) = \sum_{k=0}^{\infty} q_k z^k, \quad |z| < 1.$$

Перечислим основные свойства этой функции.

Свойство 1. $|Q(z)| \le 1$, $Q(0) = q_0$, $Q(1) = 1$.

Свойство 2. Для того чтобы случайная величина ξ имела m-й начальный момент $M(\xi^m)$, необходимо и достаточно, чтобы существовала конечная левосторонняя производная $Q^{(m)}(1)$ производящей функции $Q(z)$ в точке $z=1$; начальные моменты легко подсчитываются через факториальные моменты:

$$M(\xi(\xi-1)\ldots(\xi-m+1)) = Q^{(m)}(1).$$

В частности, $M(\xi) = Q'(1)$.

Свойство 3. В принципе производящая функция $Q(z)$ позволяет вычислить (произвести) вероятности q_i по следующей формуле:

$$q_i = \frac{1}{i!}\frac{d^i Q(z)}{dz^i}\Big|_{z=0}, \quad i \geq 0. \tag{1.64}$$

Свойство 4. Производящей функции $Q(z)$ можно придать вероятностный смысл следующим образом. Интерпретируем случайную величину ξ как число запросов, пришедших за некоторый промежуток времени. Каждый приходящий запрос с вероятностью z, $0 \leq z \leq 1$, окрашиваем в красный цвет, а с дополнительной вероятностью – в синий. Тогда из формулы полной вероятности следует, что $Q(z)$ есть вероятность того, за этот промежуток времени пришли только запросы красного цвета.

Таким образом, зная производящую функцию $Q(z)$ распределения вероятностей q_k, $k \geq 0$, мы легко можем вычислить моменты этого распределения и, в принципе, можем вычислить сами вероятности q_k, $k \geq 0$. Если непосредственный подсчет по формуле (1.64) затруднителен, можно воспользоваться методом обращения производящей функции путем разложения ее на простые дроби или численными методами. При решении практических задач можно пытаться аппроксимировать это распределение путем сглаживания по заданному числу совпадающих моментов распределения.

ГЛАВА 2

ЦЕПИ МАРКОВА

2.1. Определения

Цепь Маркова – марковский процесс с дискретным временем, заданный в измеримом пространстве.

Марковские случайные процессы названы по имени выдающегося русского математика А. А. Маркова (1856–1922), впервые начавшего изучение вероятностной связи случайных величин и создавшего теорию, которую можно назвать «динамикой вероятностей».

В дальнейшем основы этой теории явились исходной базой общей теории случайных процессов, а также таких важных прикладных наук, как теория диффузионных процессов, теория надежности, теория массового обслуживания и т.д. В настоящее время теория марковских процессов и ее приложения широко применяются в самых различных областях.

Благодаря сравнительной простоте и наглядности математического аппарата, высокой достоверности и точности получаемых решений особое внимание марковские процессы приобрели у специалистов, занимающихся исследованием операций и теорией принятия оптимальных решений.

Пусть имеется физическая система S, которая может находиться в n состояниях $E_1, E_2, \ldots, E_n$. Переходы из состояния в состояние возможны только в моменты времени $t_1, t_2, \ldots, t_k$; назовем эти моменты времени шагами. Будем рассматривать случайный процесс в системе S как функцию целочисленного аргумента $1, 2, \ldots, k$, где аргументом является номер шага.

Пример. $E_1 \to E_2 \to E_3 \to E_2$.

Условимся, что $E_i^{(k)}$ – событие, состоящее в том, что после k шагов система находится в состоянии E_i. При любом k события $E_1^{(k)}, E_2^{(k)}, \ldots, E_n^{(k)}$ образуют *полную группу событий* и являются *несовместными.* Процесс в системе можно представить как цепочку событий.

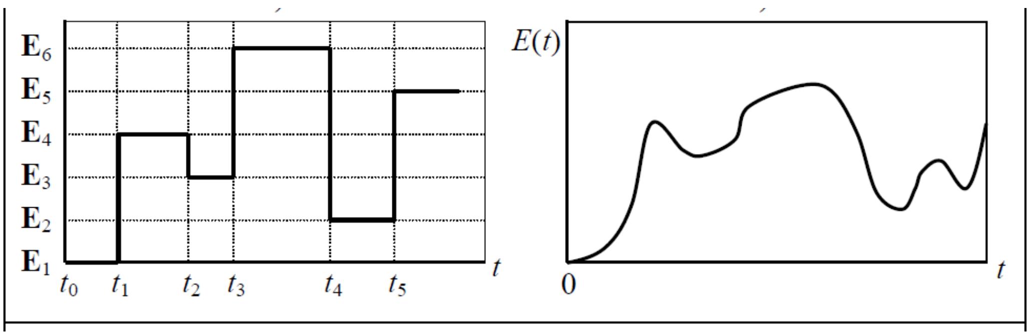
Пример: $E_1^{(0)}, E_2^{(1)}, E_3^{(2)}, E_5^{(3)}, \ldots$

Такая последовательность называется ***марковской цепью***, если для каждого шага вероятность перехода из любого состояния E_i

в любое состояние E_j не зависит от того, когда и как система пришла в состояние E_i.

Цепь Маркова называется ***однородной***, если вероятности переходов не зависят от момента времени t_k, и ***неоднородной***, если вероятности переходов являются функциями t_k, т.е. $p_{ij} = p_{ij}(k)$.

Для случайных процессов с дискретным временем (рис. 2.1,*а*) изменения состояний происходят только в определенные моменты времени $t_1, t_2, \ldots, t_k, \ldots$ Переходы между состояниями описываются ***вероятностями переходов***. Если непосредственный переход из одного состояния в другое невозможен, то вероятность, соответствующая данному переходу, равна нулю. Обозначим через p_{ij} условную вероятность того, что в момент времени t_{k+1} случайный процесс перейдет в состояние E_j при условии, что в момент t_k процесс находился в состоянии E_i. Если переход из состояния E_i в E_j зависит только от этих двух состояний, т.е. условная вероятность p_{ij} не изменяется при дополнительной информации о поведении процесса до момента t_k, получим цепь Маркова.



а) *б*)

Рис. 2.1. Процессы с дискретным (*а*) и непрерывным (*б*) временем

Вероятности переходов задаются в виде квадратной ***матрицы вероятностей переходов*** $\mathbf{P} = [p_{ij} \mid i, j = 1, \ldots, n]$, элементы которой удовлетворяют условиям:

$$0 \le p_{ij} \le 1; \ \sum_{j=1}^{n} p_{ij} = 1, \quad i = 1, \ldots, n, \ j = 1, \ldots, n.$$

Матрица, элементы которой удовлетворяют указанным условиям, называется ***стохастической***.

Зная матрицу переходных вероятностей и начальное состояние системы, можно найти вероятности состояний $P_1(k), P_2(k), \ldots, P_n(k)$ после любого k-го шага. Пусть в начальный момент времени система находится в состоянии E_m. Тогда для $t = 0$ $P_1(0)=0,\ P_2(0)=0,\ ...,P_m(0)=1,\ldots,P_n(0)=0.$ Найдем вероятности после первого шага. Из состояния E_m система перейдет в состояния E_1, E_2 и т.д. с вероятностями $P_{m1},\ P_{m2}\ldots,\ P_{mm}\ldots,\ P_{mn}$. Тогда после первого шага вероятности будут равны

$$P_1(1)=P_{m1};\ P_2(1)=P_{m2},...,\ P_n(1)=P_{mn}. \qquad (*)$$

Найдем вероятности состояния после второго шага: $P_1(2),\ P_2(2),\ ...,\ P_n(2)$. Будем вычислять эти вероятности по формуле полной вероятности с гипотезами:

$$P(A)=\sum_{i=1}^{n}P(H_i)P_{H_i}(A).$$

Гипотезами будут следующие утверждения:

- после первого шага система была в состоянии $E_1 - H_1$;
- после второго шага система была в состоянии $E_2 - H_2$; …;
- после n-го шага система была в состоянии $E_n - H_n$.

Вероятности гипотез известны из выражения (*). Условные вероятности перехода в состояние A при каждой гипотезе тоже известны и записаны в матрицы переходных состояний. Тогда по формуле полной вероятности получим:

$$P_1(2)=P_1(1)P_{11}+P_2(1)P_{21}+\ldots+P_n(1)P_{n1};$$

$$P_n(2)=P_1(1)P_{1n}+P_2(1)P_{2n}+\ldots+P_n(1)P_{nn}.$$

Вероятность любого состояния после второго шага:

$$P_i(2)=P_j(1)P_{j1}+P_j(1)P_{j1}+\ldots+P_j(1)P_{j1}, i=\overline{1,n}.$$

В последней формуле суммируются все переходные вероятности P_{ij}, но учитываются только отличные от нуля. Вероятность любого состояния после k-го шага:

$$P_i\left(k\right)=\sum_{j=1}^{n}P_j\left(k-1\right)P_{ji},\ i=\overline{1,n}\,.$$

Таким образом, вероятность состояния после k-го шага определяется по рекуррентной формуле через вероятности (k –1)-го шага.

Для случайных процессов с непрерывным временем (см. рис. 2.1,*б*) время между переходами из одного состояния в другое случайно. Это означает, что вероятность перехода из одного состояния в другое не может быть задана, поскольку вероятность такого перехода точно в произвольный момент времени t равна нулю. Для описания переходов между состояниями случайного процесса с непрерывным временем вместо вероятностей переходов вводится параметр, называемый *интенсивностью перехода*.

Случайный процесс называется *марковским* или *случайным процессом без последействия*, если для любого момента времени t_0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент t_0 и не зависят от того, когда и как система пришла в это состояние.

Интенсивность перехода g_{ij} из состояния E_i в состояние E_j определяется как предел отношения вероятности перехода $P_{ij}(\Delta\tau)$ системы за промежуток времени Dt из E_i в E_j к длине этого промежутка:

$$g_{ij} = \lim_{\Delta\tau \to 0} \frac{P_{ij}(\Delta\tau)}{\Delta\tau} \quad (i, j = \overline{1,n}, i \neq j). \tag{2.1}$$

Отсюда следует, что вероятность перехода за бесконечно малый промежуток времени $\Delta\tau$ равна: $g_{ij}\ \Delta\tau(i \neq j)$. Вероятность двух и более переходов за время $\Delta\tau$ имеет порядок $(\Delta\tau)^2$ и выше и предполагается бесконечно малой величиной.

Если интенсивности переходов постоянны и не зависят от времени t, т.е. от того, в какой момент начинается промежуток Dt, то марковский процесс называется ***однородным***. Если интенсивности g_{ij} представляют собой функции времени t, процесс называется ***неоднородным***.

В дальнейшем будем рассматривать только однородные марковские процессы.

Интенсивности переходов задаются в виде квадратной матрицы $\mathbf{G} = [g_{ij}\ |i, j = 1, \ldots, n]$, называемой ***матрицей интенсивностей переходов***, диагональные элементы которой определяются из условия:

$$\sum_{j=1}^{n} g_{ij} = 0, \ i = 1, \ldots, n,$$

откуда

$$g_{ij} = -\sum_{\substack{j=1 \\ j \neq i}}^{n} g_{ij} = 0, \quad i, \ j = 1, \ldots, n. \tag{2.2}$$

Матрица, в которой сумма элементов в каждой строке равна нулю, называется ***дифференциальной***.

Вероятность перехода из одного состояния в другое за бесконечно малый интервал времени:

$$P_{ij}(\Delta\tau) = 1 - (1 - \lambda_{ij}\Delta\tau) = \lambda_{ij}\,\Delta\tau.$$

Отсюда следует, что *интенсивность перехода представляет собой параметр экспоненциального распределения*:

$$g_{ij} = \lim_{\Delta\tau\to 0}\frac{P_{ij}(\Delta\tau)}{\Delta\tau} = \lambda_{ij}\ \left(i, j = \overline{1,n}, i \neq j\right),$$

а диагональные элементы определяются из условия: $\sum_{j=1}^{n} g_{ij} = 0,\ i = \overline{1,n}.$ Вектор состояний $P(t) = \{p_1(t), ..., p_n(t)\}$ является *основной* характеристикой Марковского случайного процесса.

Изучение случайных процессов заключается в определении вероятностей состояний $p_1(t), ..., p_n(t)$, которые могут быть представлены *стохастическим* вектором:

$$P(t) = \{p_1(t), ..., p_n(t)\},\ 0 \le p_{\text{i}} \le 1,\ \text{причем}\ \sum_{i=1}^{n} p_i = 1.$$

Для *Марковского процесса с дискретным временем*, обладающего эргодическим свойством, стационарные вероятности состояний определяются из системы линейных алгебраических уравнений: $p_j = \sum_{i=1}^{n} p_i g_{ij} = 0,\ \ j = \overline{1,n},$ которая совместно с нормировочным условием $\sum_{i=1}^{n} p_i = 1$ образует систему, обладающую единственным решением.

Для *Марковского процесса с непрерывным временем*, обладающего эргодическим свойством, стационарные вероятности состояний определяются из системы линейных алгебраических уравнений:

$$\sum_{i=1}^{n} p_i g_{ij} = 0,\ \ j = \overline{1,n}.$$

2.2. Модель «гибели и размножения»

В качестве примера марковского процесса с непрерывным временем рассмотрим модель «гибели и размножения», которая часто встречается в разнообразных практических задачах. Своим названием эта модель обязана биологической задаче об изменении численности популяции и распространении эпидемий, которая формулируется следующим образом.

Рассмотрим развитие некоторой популяции, особи которой могут рождаться и умирать. Положим, что при наличии i особей в популяции рождение новых особей происходит с интенсивностью λ_i, а с интенсивностью μ_i особи умирают. Пусть в любой момент времени может происходить рождение или гибель только одной особи, и интервалы времени между двумя моментами рождения и гибели распределены по экспоненциальному закону с параметрами λ_i и μ_i соответственно. Тогда процесс «гибели и размножения» может быть представлен марковским случайным процессом с непрерывным временем (рис. 2.2,*а*), в котором состояние E_i соответствует наличию i особей в популяции (i = 0, 1, …), причем число состояний может быть конечным или бесконечным. Отметим, что состояние E_0 соответствует вырождению популяции.

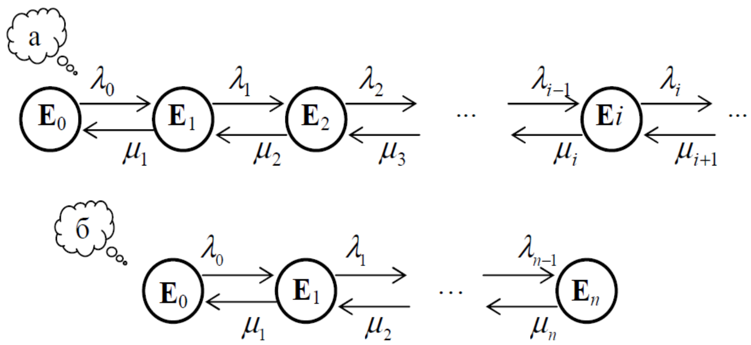


Рис. 2.2. Графы состояний процесса гибели и размножения с бесконечным (*а*) и конечным (*б*) числом состояний

Таким образом, марковский процесс называется «**процессом гибели и размножения**», если его граф переходов имеет вид цепочки состояний, в которой каждое состояние (кроме крайних) связано с двумя соседними состояниями, а крайние состояния E_0 и E_n (в случае конечного числа состояний) или только нулевое состояние E_0 (в случае бесконечного числа состояний) – только с одним соседним состоянием.

Графу переходов процесса гибели и размножения с конечным числом состояний (рис. 2.2,*б*) соответствует матрица интенсивностей переходов:

G =

$\mathbf{E}_i$	0	1	2	…	n – 1	n
0	$-\lambda_0$		0	...	0	0
1	μ_1	$-(\lambda_1+\mu_1)$	λ_1	...	0	0
2	0	μ_2	$-(\lambda_2+\mu_2)$	...	0	0
...	…	...	...	...	...	...
$n-1$	0	0	0	...	$-(\lambda_{n-1}+\mu_{n-1})$	λ_{n-1}
n	0	0	0	...	$+\mu_n$	$-\mu_n$

Диагональные элементы матрицы определяются из условия (2.2) – сумма элементов каждой строки должна быть равна нулю. *Система линейных алгебраических уравнений* для определения стационарных вероятностей может быть составлена по графу переходов или по матрице интенсивностей переходов.

Сформулируем **правила составления уравнений для стационарных вероятностей состояний** марковского процесса с непрерывным временем по графу переходов и по матрице интенсивностей переходов.

Правило 1 (по графу переходов). В левой части каждого уравнения записывается вероятность рассматриваемого состояния, умноженная на сумму интенсивностей переходов из данного состояния во все другие состояния. Правая часть уравнения представляет собой сумму членов, число которых равно числу входящих в данное состояние дуг, и каждый такой член представляет собой произведение интенсивности перехода, соответствующей данной дуге, на вероятность состояния, из которого исходит эта дуга.

Для нашего примера применение правила 1 дает следующую систему линейных алгебраических уравнений, где последнее уравнение представляет собой нормировочное условие:

$$\begin{cases} \lambda_0 p_0 = \mu_1 p_1; \\ (\lambda_1+\mu_1)p_1 = \lambda_0 p_0 + \mu_2 p_2; \\ (\lambda_k+\mu_k)p_k = \lambda_{k-1}p_{k-1} + \mu_{k+1}p_{k+1}; \\ \qquad\qquad\qquad \ldots \\ \mu_n p_n = \lambda_{n-1}p_{n-1}; \\ p_0 + p_1 + \ldots + p_n = 1. \end{cases}$$

Правило 2 (по матрице интенсивностей переходов). Для каждого столбца матрицы интенсивностей переходов составляется соответствующее уравнение как сумма произведений интенсивностей переходов на стационарную вероятность состояния с номером соответствующей строки, приравненная нулю.

Применение правила 2 для нашего примера дает следующую систему линейных алгебраических уравнений:

$$\begin{cases} -\lambda_0 p_0 + \mu_1 p_1 = 0; \\ -(\lambda_1 + \mu_1) p_1 + \lambda_0 p_0 + \mu_2 p_2 = 0; \\ -(\lambda_k + \mu_k) p_k + \lambda_{k-1} p_{k-1} + \mu_{k+1} p_{k+1} = 0; \\ \qquad\qquad \dots \\ -\mu_n p_n + \lambda_{n-1} p_{n-1} = 0; \\ p_0 + p_1 + \dots + p_n = 1. \end{cases}$$

Эргодические свойства цепей Маркова с непрерывным временем полностью определяются эргодическими свойствами вложенных цепей Маркова с дискретным временем.

Определение 1. Состояние i называется **несущественным**, если существует такое состояние j, в которое система может перейти за конечное число шагов n, но не может вернуться в i-е ни за какое число шагов, т.е.

$$p_{ij}(n) > 0, \;\; p_{ji}(m) = 0 \qquad \forall m.$$

Все остальные состояния существенные.

Определение 2. Состояния i и j называются **сообщающимися** $(i \leftrightarrow j)$, если существуют n и k такие, что

$$p_{ij}(n) > 0, \;\; p_{ji}(k) > 0.$$

Определение 3. Все существенные состояния можно разбить на классы, которые состоят из сообщающихся состояний, и ни из одного состояния данного класса нельзя перейти в состояние другого класса. Такие классы называют **неразложимыми**, или **замкнутыми**.

Рассмотрим неразложимый класс S. Тогда для $\forall i \in S \; \exists n$, что $p_{ii}(n) > 0$. Пусть M_i – множества числа шагов n, для которых $P_{ii}(n) > 0$.

Определение 4. Наибольший общий делитель d_i этих чисел называется периодом состояния i.

Теорема солидарности 1. Все состояния одного неразложимого класса имеют одинаковый период d.

Определение 5. Если $d=1$, то класс называется непериодическим, или эргодическим.

Введем вероятность

$$f_i(n)=P\left(\xi(n)=i \middle| \xi(0)=i,\ \xi(1)\neq i,\ \xi(2)\neq i,\ldots,\xi(n-1)\neq i\right)$$

того, что система впервые возвращается в i состояние на n-м шаге.

Тогда вероятность

$$f_i^*=\sum\nolimits_{n=1}^{\infty}f_i(n)$$

можно рассматривать как вероятность того, что система, выйдя из i-го состояния, хотя бы один раз вернется в него.

Определение 6. Состояние i называется возвратным, если $f_i^*=1$, и невозвратным, если $f_i^*<1$.

Теорема солидарности 2. Если имеется два сообщающихся состояния и одно из них возвратно, то второе возвратно также.

Для возвратного состояния i вероятности $f_i(n)$ образуют вероятностное распределение времени возвращения в i-е состояние.

Определение 7. Если i – возвратное состояние и $\sum\nolimits_{n=1}^{\infty}nf_i(n)=\infty$, то состояние i называется **нулевым**, если $\sum\nolimits_{n=1}^{\infty}nf_i(n)<\infty$, то возвратное состояние называется **ненулевым**, или **положительным**.

В любой момент времени t система может находиться в одном из n возможных состояний, т.е. для любого момента времени t выполняется условие

$$\sum_{i=1}^{n}p_i(t)=1,$$

которое называется нормировочным.

Основная эргодическая теорема для марковских цепей. Рассмотрим возвратную, неприводимую, непериодическую марковскую цепь, тогда имеет место равенство

$$\lim_{n\to\infty}\mathrm{p_{ij}}(n)=\frac{1}{\sum\nolimits_{n=1}^{\infty}nf_i(n)}\begin{cases}=0,\text{ если } i \text{ – нулевое состояние},\\>0,\text{ если } i \text{ – положительное состояние}.\end{cases}$$

При этих же условиях

$$\lim_{\mathrm{n}\to\infty}p_{ij}(n)=\lim_{\mathrm{n}\to\infty}p_{ii}(n).$$

Теорема альтернативы. Пусть для марковской цепи со счетным числом состояний и матрицей переходных вероятностей p_{ij} существуют пределы

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j \ \ \forall j,$$

тогда $\sum_i \pi_i P_{ij} = \pi_j$, и либо все $\pi_i = 0$, либо $\sum_i \pi_i = 1$.

Если $\pi_i = 0$, то стационарное распределение не существует; если $\sum_i \pi_i = 1$, то π_i называются финальными вероятностями, а их набор образует эргодическое распределение, которое совпадает с единственным стационарным распределением.

Таким образом, для того чтобы существовало стационарное распределение вероятностей, необходимо и достаточно, чтобы цепь Маркова была

1) неразложима;
2) непериодична;
3) возвратна;
4) положительна.

Для проверки условий возвратности и положительности сформулируем две конструктивные теоремы.

Теорема Фостера (эргодическая теорема Фостера). Для того чтобы неприводимая, непериодическая цепь Маркова была строго эргодической, необходимо и достаточно, чтобы система уравнений

$$x(j) = \sum_{i\in X} \pi_i P_{ij}, \ \ j \in X,$$

имела нетривиальное решение $x(i),\ i \in X$, такое, что $\sum_{i\in X} |x(i)| < \infty$. При этом существует единственное стационарное распределение, которое совпадает с эргодическим.

Теорема Мустафы (эргодическая теорема Мустафы). Для того чтобы неприводимая, непериодическая цепь Маркова была эргодической, достаточно существование $\varepsilon > 0$, натурального числа i_0 набора неотрицательных чисел $x_0, x_1, x_2, \ldots$, таких, что

$$\sum_{j\in X} P_{ij} x_j \le x_i - \varepsilon \text{ для } i > i_0,$$

$$\sum_{j\in X} P_{ij} x_j < \infty \text{ для } i \le i_0.$$

При этом существует единственное эргодическое распределение, которое совпадает со стационарным.

Теорему Мустафы можно сформулировать и по-другому.

Теорема Мустафы. Для того чтобы неприводимая, непериодическая цепь Маркова была эргодической, достаточно существование $\varepsilon > 0$, натурального числа i_0 и неотрицательной функции $f(i)$ такой, что

$$M\{f(\xi_{n+1}) | \xi_n = i\} \le f(i) - \varepsilon \text{ для } i > i_0,$$

$$M\{f(\xi_{n+1}) | \xi_n = i\} \le \infty \text{ для } i \le i_0.$$

При этом существует единственное эргодическое распределение, которое совпадает со стационарным.

Теорема 2.1. Пусть система S имеет множество возможных состояний $\{S_k\}_{k=1}^{n}$, а процесс изменения состояний этой системы представляет собой случайный процесс, причем для всех пар возможных состояний S_i и S_j определены плотности вероятностей переходов $\lambda_{ij}(t)$ и $\lambda_{ji}(t)$.

Тогда вероятности состояний системы $p_k(t)$ удовлетворяют ***системе уравнений Колмогорова***:

$$p'_k(t) = \sum_{\substack{i=1 \\ i \ne k}}^{n} \lambda_{ik}(t) p_i(t) - \left(\sum_{\substack{i=1 \\ i \ne k}}^{n} \lambda_{ki}(t) \right) p_k(t),\ k = \overline{1,n}\ t \in T. \quad (2.3)$$

Так как процесс изменения состояний системы S представляет собой марковский процесс и $\{S_k\}_{k=1}^{n}$ полная группа событий, то по формуле полной вероятности имеем

$$P\left[s_k^{t+\Delta t} \right] = \sum_{i=1}^{n} P\left[s_k^{t+\Delta t} \middle| s_i^t \right] P\left[s_i^t \right] =$$

$$= \sum_{\substack{i=1 \\ i \ne k}}^{n} P\left[s_k^{t+\Delta t} \middle| s_i^t \right] P\left[s_i^t \right] + P\left[s_k^{t+\Delta t} \middle| s_k^t \right] P\left[s_k^t \right]. \quad (2.4)$$

С точностью $o(\Delta t)$ можно написать

$$\sum_{i=1}^{n} P\left[s_k^{t+\Delta t} \middle| s_i^t \right] = \lambda_{ik}(t)\Delta t, \quad i \ne k.$$

Если $i = k$, то, переходя к противоположному событию, находим с точностью $o(\Delta t)$

$$P\left[s_k^{t+\Delta t} \middle| s_k^t \right] = 1 - \sum_{\substack{i=1 \\ i \ne k}}^{n} \lambda_{ki}(t)\Delta t$$

Таким образом, с точностью $o(\Delta t)$ равенство (2.4) может быть преобразовано к следующему:

$$\frac{p_k(t+\Delta t)-p_k(t)}{\Delta t}=\sum_{\substack{i=1\\ i\neq k}}^{n}\lambda_{ik}(t)p_i(t)-\left(\sum_{\substack{i=1\\ i\neq k}}^{n}\lambda_{ik}(t)p_i(t)\right)p_k(t).$$

И для завершения доказательства достаточно перейти к пределу при $\Delta t\to +0.$

Система уравнений Колмогорова (2.3) не является линейно независимой, т.е. она является избыточной.

Если $I=(1...1)\in M_{1n}(R),$ а компонентами вектора

$$\lambda_k^0(t)=(\lambda_{k1}(t)...\lambda_{k,k-1}(t),\ \lambda_{k,k+1}...\lambda_{kn}(t))^{\tau}$$

являются плотности вероятностей переходов системы S из состояния S_k во все иные возможные состояния, то

$$\lambda_{kk}(t)\overset{\Delta}{=}-I\lambda_k^0(t)=-\sum_{\substack{i=1\\ i\neq k}}^{n}\lambda_{ik}(t) \tag{2.5}$$

есть суммарная плотность вероятности перехода системы из состояния S_k, взятая со знаком «минус». При этом, если ввести матрицу

$$\lambda(t)\overset{\Delta}{=}(\lambda_{ij}(t))\in M_n(R), \tag{2.6}$$

диагональные элементы которой определены согласно (2.5), то (2.3) переходит в ***матричное уравнение Колмогорова***:

$$p'(t)=\lambda(t)p(t),\ t\in T=[a;b], \tag{2.7}$$

где $p(t)$ – вектор вероятностей состояния системы S в момент времени t.

Если определен вектор вероятностей p_a начальных состояний системы S в момент $t=a$ и определена матрица $\lambda(t)$, то с учетом (2.7) приходим к задаче Коши для матричного уравнения Колмогорова:

$$\begin{cases} p'(t)=\lambda(t)p(t),\ t>a;\\ p(a)=p_a. \end{cases} \tag{2.8}$$

Если $\lambda(t)$ непрерывна при $t > a$, то задача Коши (2.8) имеет единственное решение и, следовательно, вектор вероятностей состояний исходной системы S определен однозначно в любой момент времени $t > a$.

Если при этом матрицы $\lambda(t)$ и $\int_a^t \lambda(\tau)d\tau$ коммутируют при каждом фиксированном $t \in T = [a,b]$, решение задачи Коши (2.8) с использованием матричной экспоненты можно записать в явном виде

$$p(t) = \exp\left(\int_a^t \lambda(\tau)d\tau\right) p_a, \ t \in T. \tag{2.9}$$

Согласно (2.6) k-м столбцом матрицы $\lambda(t)$ является вектор $\lambda_k^0(t)$, у которого k-я компонента заменена (по (2.5)) на $\lambda_{kk}(t) = -I\lambda_k^0(t)$, а все остальные – соответствующие плотности вероятностей переходов из состояния E_k. Эту информацию проставляют на размеченном графе состояний системы на стрелках, «выходящих из состояния E_k» (при их отсутствии соответствующая компонента равна нулю). При этом

$$I\lambda(t) \equiv 0. \tag{2.10}$$

Пример 10. Размеченный граф состояний системы S, процесс изменения состояний которой есть однородный марковский процесс с дискретными состояниями, изображен на рис. 2.3.

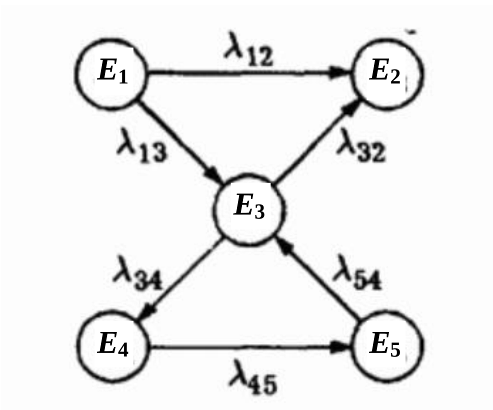


Рис. 2.3. Размеченный граф состояний системы

Сформулируем задачу Коши для системы уравнений Колмогорова, если $T = [0, \infty]$, и в начальный момент времени $t = 0$ система находится в состоянии E_1.

Согласно заданному размеченному графу состояний, имеем

$$\lambda(t) \equiv \lambda = \begin{pmatrix} -\lambda_{12}-\lambda_{13} & 0 & 0 & 0 & 0 \\ \lambda_{12} & 0 & \lambda_{32} & 0 & 0 \\ \lambda_{13} & 0 & -\lambda_{32}-\lambda_{34} & 0 & \lambda_{53} \\ 0 & 0 & \lambda_{34} & -\lambda_{45} & 0 \\ 0 & 0 & 0 & \lambda_{45} & -\lambda_{53} \end{pmatrix}$$

$$p(0) = (10000)^T.$$

Таким образом, вектор $p(t)$ вероятностей состояний изучаемой системы S является решением следующей задачи Коши:

$$\begin{cases} p'_1(t) = -(\lambda_{12}+\lambda_{13})\,p_1(t); \\ p'_2(t) = \lambda_{12}p_1(t) + \lambda_{32}p_3(t); \\ p'_3(t) = \lambda_{13}p_1(t) - (\lambda_{32}+\lambda_{34})\,p_3(t) + \lambda_{53}p_5(t); \\ p'_4(t) = \lambda_{34}p_3(t) - \lambda_{45}p_4(t); \\ p'_5(t) = \lambda_{45}p_4(t) - \lambda_{53}p_5(t); \\ p_1(0) = 1, \quad p_k(0) = 0, \quad k = \overline{2,5}. \end{cases}$$

Если решение задачи Коши (2.8) представимо в виде (2.9), то оно удовлетворяет теореме Колмогорова.

Задача 1. Построить граф состояний следующего случайного процесса: устройство S состоит из двух узлов, каждый из которых в случайный момент времени может выйти из строя, после чего мгновенно начинается ремонт узла, продолжающийся заранее неизвестное случайное время.

Решение. Возможные состояния системы: E_0 – оба узла исправны; E_1 – первый узел ремонтируется, второй исправен; E_2 – второй узел ремонтируется, первый исправен; E_3 – оба узла ремонтируются. Граф системы приведен на рис. 2.4.

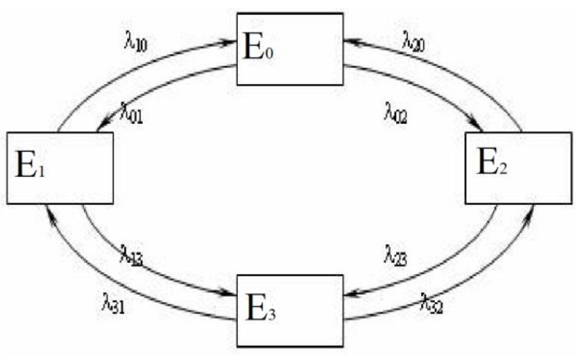


Рис. 2.4. Граф системы задачи 1

Стрелка, направленная, например, из E_0 в E_1 означает переход системы в момент отказа первого узла, из E_1 в E_0 – переход в момент окончания ремонта этого узла.

На графе отсутствуют стрелки из E_0 в E_3 и из E_1 в E_2. Это объясняется тем, что выходы узлов из строя предполагаются независимыми друг от друга и, например, вероятностью одновременного выхода из строя двух узлов (переход из E_0 в E_3) или одновременного окончания ремонтов двух узлов (переход из E_3 в E_0) можно пренебречь.

Задача 2. Задана матрица $P_1 = \begin{Vmatrix} 0,4 & 0,6 \\ 0,3 & 0,7 \end{Vmatrix}$ вероятностей перехода дискретной цепи Маркова из i-го состояния в j-е за один шаг (i, j = 1, 2). Распределение вероятностей по состояниям в начальный момент $t = 0$ определяется вектором $\vec{q}\,(0{,}1;\ 0{,}9)$.

Найти:

1) матрицу P_2 перехода цепи из состояния i в состояние j за два шага;

2) распределение вероятностей по состояниям в момент $t = 2$;

3) вероятность того, что в момент $t = 1$ состоянием цепи будет E_2;

4) стационарное распределение.

Решение. Для дискретной цепи Маркова в случае ее однородности справедливо соотношение

$$P_n = P_1^n, \tag{2.11}$$

где P_1 – матрица переходных вероятностей за один шаг; P_n – матрица переходных вероятностей за n шагов.

1. Найдем матрицу P_2 перехода за два шага:

$$P_2 = P_1^2 = \begin{Vmatrix} 0,4 & 0,6 \\ 0,3 & 0,7 \end{Vmatrix} \cdot \begin{Vmatrix} 0,4 & 0,6 \\ 0,3 & 0,7 \end{Vmatrix} = \begin{Vmatrix} 0,34 & 0,66 \\ 0,33 & 0,67 \end{Vmatrix}.$$

Пусть распределение вероятностей по состояниям на S-м шаге определяется вектором

$$\vec{p}(s) = \left(p_1(s), p_2(s), \ldots, p_k(s) \right),\ 0 \le p_j(s) \le 1,\ \sum_{j=1}^{k} p_j = 1.$$

Зная матрицу P_n перехода за n шагов, можно определить распределение вероятностей по состояниям на $(s+n)$-м шаге:

$$\vec{p}(s+n) = \vec{p}(s) \cdot P_n. \tag{2.12}$$

2. Найдем распределение вероятностей по состояниям системы в момент t = 2. Положим в (2.12) s = 0 и n = 2, тогда $\vec{p}(0) = \vec{q} = (0,1; 0,9)$. Получим

$$\vec{p}(2) = \vec{q} \cdot P_2 = (0,1; 0,9) \begin{Vmatrix} 0,34 & 0,66 \\ 0,33 & 0,67 \end{Vmatrix} = (0,331;\ 0,669).$$

3. Найдем распределение вероятностей по состояниям системы в момент t = 1.

Положим в (2.12) s = 0 и n = 1, тогда

$$\vec{p}(1) = \vec{q} \cdot P_1 = \left(0,1; 0,9\right) \begin{Vmatrix} 0,4 & 0,6 \\ 0,3 & 0,7 \end{Vmatrix} = \left(0,31;\ 0,69\right),$$

откуда видно, что вероятность того, что в момент t = 1 состоянием цепи будет E_2, равна $p_2(1)$ = 0,69.

Распределение вероятностей по состояниям называется стационарным, если оно не меняется от шага к шагу, т.е. $\vec{p}(s) = \vec{p} = (p_1, p_2, p_3, \ldots, p_k),\ p_j = \mathrm{const}, j = 1, \ldots, k.$

Тогда из соотношения (2.12) при n = 1 получим $\vec{p}(s) = \vec{p},$ $\vec{p}(s+1) = \vec{p};\ P_N = P_1$, или

$$\vec{p}(s) = \vec{p} \cdot P_1,\ 0 \le p_j \le 1,\ j = 1, \ldots, k,\ \ \sum_{j=1}^{k} p_j = 1. \qquad (2.13)$$

4. Найдем стационарное распределение. Так как k = 2, имеем $\vec{p}(t) = (p_1; p_2)$. Напишем систему линейных уравнений (2.13) в координатной форме:

$$\begin{cases} p_1 = 0,4p_1 + 0,3p_2; \\ p_2 = 0,6p_1 + 0,7p_2; \end{cases} p_1 + p_2 = 1.$$

Последнее условие нормировочное. В системе (2.13) всегда одно уравнение является линейной комбинацией других. Следовательно, его можно вычеркнуть. Решим совместно первое уравнение системы и нормировочное. Имеем $0,6p_1 = 0,3p_2$, т.е. $p_2 = 2p_1$. Тогда $p_1 + 2p_1 = 1$ или $p_1 = 1/3$, т.е. $p_2 = 2/3$. Следовательно, $\vec{p}(t) = \left(\frac{1}{3}, \frac{2}{3}\right)$.

Ответ:

1) матрица перехода за два шага для данной цепи Маркова имеет вид

$$P_2 = \begin{Vmatrix} 0{,}34 & 0{,}66 \\ 0{,}33 & 0{,}67 \end{Vmatrix};$$

2) распределение вероятностей по состояниям в момент $t = 2$ равно $\vec{p}(t) = (0.331; 0{,}669)$;

3) вероятность того, что в момент $t = 1$ состоянием цепи будет E_2, равна $p_2(t) = 0{,}69$;

4) стационарное распределение имеет вид $\vec{p}(t) = \left(\frac{1}{3}, \frac{2}{3}\right)$.

Задача 3. Для заданной матрицы переходных вероятностей **P** найти вероятности перехода за 2 шага и стационарные вероятности, если они существуют.

Задача 4. Задана матрица $\mathbf{P_1} = \begin{pmatrix} 0{,}8 & 0{,}2 \\ 0{,}2 & 0{,}8 \end{pmatrix}$ и вектор $\overline{q}(t) = (0{,}1; 0{,}9)$ вероятностей перехода дискретной цепи Маркова из состояния i ($i = 1, 2$) в состояние j ($j = 1, 2$) за один шаг. Распределение вероятностей по состояниям в момент $t = 0$ определяется вектором $\overline{q}$.

Найти:

1) матрицу $\mathbf{P_2}$ перехода из состояния i в состояние j за два шага;

2) распределение вероятностей по состояниям в момент $t = 2$;

3) вероятность того, что в момент $t = 1$ состоянием цепи будет $i = 2$;

4) стационарное распределение.

Задача 5. Задана матрица $\Lambda = \begin{pmatrix} -2 & 1 & 1 \\ 2 & -3 & 1 \\ 4 & 0 & 4 \end{pmatrix}$ интенсивностей переходов непрерывной цепи Маркова. Составить размеченный граф состояний, соответствующий матрице Λ; составить систему дифференциальных уравнений Колмогорова для вероятностей состояний; найти предельное распределение вероятностей.

Решение. Однородная цепь Маркова с конечным числом состояний E_1, E_2, …, E_k характеризуется матрицей интенсивностей переходов $\Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \ddots & \vdots \\ \lambda_{k1} & \cdots & \lambda_{kk} \end{pmatrix}$, где $\lambda_{ij} = \lim\limits_{\Delta t \to 0} \dfrac{p_{ij}(\Delta t)}{\Delta t}$ – интенсивность перехода цепи Маркова из состояния E_i в состояние E_j; $p_{ij}(\Delta t)$ – вероятность перехода $E_i \to E_j$ за интервал времени Δt.

Переходы системы из состояния в состояние удобно задавать с помощью размеченного графа состояний, на котором отмечаются дуги, соответствующие интенсивностям $\lambda_{ij} > 0$. Составим размеченный граф состояний для заданной матрицы интенсивностей переходов (рис. 2.5).

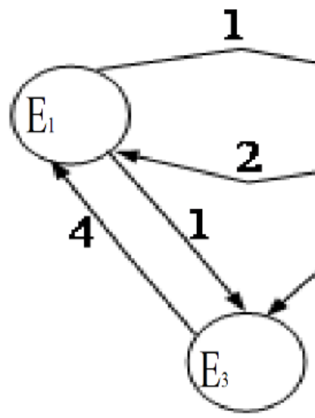


Рис. 2.5. Размеченный граф состояний задачи 5

Пусть $\vec{p}(t) = (p_1(t), p_2(t), p_3(t), \ldots, p_k(t))$ – вектор вероятностей $p_j(t)$, $j = 1, 2, \ldots, k$, нахождения системы в состоянии E_j в момент t.

Очевидно, что $0 \le p_j(t) \le 1$ и $\sum_{j=1}^{k} p_j(t) = 1$. Тогда по правилу дифференцирования векторной функции скалярного аргумента получим $\vec{p}'(t) = (p'_1(t), p'_2(t), p'_3(t), \ldots, p'_k(t))$. Вероятности $p_j(t)$ удовлетворяют системе дифференциальных уравнений Колмогорова (СДУК), которая в матричной форме имеет вид

$$\vec{p}'(t) = \vec{p}(t)\Lambda. \tag{2.14}$$

Если в начальный момент система находилась в состоянии E_j, то СДУК следует решать при начальных условиях:

$$p_i(0) = 1, p_j(0) = 0, j \ne i, j = 1, 2, \ldots, k. \tag{2.15}$$

Совокупность СДУК (2.14) и начальных условий (2.15) однозначно описывает однородную цепь Маркова с непрерывным временем и конечным числом состояний.

Составим СДУК для заданной цепи Маркова. Поскольку $k = 3$, то $j = 1, 2, 3$.

Из соотношения (2.14) получим

$$(p'_1(t), p'_2(t), p'_3(t)) = (p_1(t), p_2(t), p_3(t))\begin{pmatrix} -2 & 1 & 1 \\ 2 & -3 & 1 \\ 4 & 0 & -4 \end{pmatrix}.$$

Отсюда будем иметь

$$\begin{cases} p'_1(t) = -2p_1(t) + 2p_2(t) + 4p_3(t); \\ p'_2(t) = p_1(t) - 3p_2(t); \\ p'_3(t) = p_1(t) + p_2(t) - 4p_3(t); \\ p_1(t) + p_2(t) + p_3(t) = 1. \end{cases}$$

Последнее условие называется нормировочным. Распределение $\vec{p}(t)$ вероятностей по состояниям называется *стационарным*, если оно не меняется с течением времени, т.е.

$$\vec{p}(t) = \vec{p} = (p_1, p_2, p_3, \ldots, p_k), \text{ где } p_j = \text{const}, j = 1, 2, \ldots, k.$$

Отсюда $\vec{p}'(t) = \vec{p}' = 0$. Тогда из СДУК (2.13) получаем систему для нахождения стационарного распределения

$$\vec{p}(t)\Lambda = 0, \text{ где } \sum_{j=1}^{k} p_j(t) = 1, \ 0 \le p_j \le 1. \qquad (2.16)$$

Для данной задачи из СДУК будем иметь

$$\begin{cases} -2p_1(t) + 2p_2(t) + 4p_3(t) = 0; \\ p_1(t) - 3p_2(t) = 0; \\ p_1(t) + p_2(t) - 4p_3(t) = 0; \\ p_1(t) + p_2(t) + p_3(t) = 1. \end{cases} \Rightarrow \begin{cases} p_1 = 3p_2, \\ p_2 = p_3. \end{cases}$$

Из нормировочного условия получим $3p_2 + p_2 + p_2 = 1$ или $p_2 = 1/5$, $p_1 = 3/5$, $p_3 = 1/5$. Следовательно, предельное распределение имеет вид $\vec{p} = \left(\frac{3}{5}, \frac{1}{5}, \frac{1}{5}\right)$.

Заметим, что этот результат можно получить непосредственно по размеченному графу состояний, если воспользоваться правилом: для стационарного распределения сумма произведений $\lambda_{ji}p_i$, $j \ne i$, для стрелок, выходящих из i-го состояния, равна сумме произведений $\lambda_{ji}p_i$, $j \ne i$, для стрелок, входящих в i-е состояние. Действительно,

$$\begin{cases} E_1: 2p_2 + 4p_3 = p_1 + p_1, & p_1 + p_2 + p_3 = 1, \\ E_2: p_1 = 2p_2 + p_2, & 0 \le p_j \le 1, \\ E_3: p_1 + p_2 = 4p_3, & \forall j = 1,\ 2,\ 3. \end{cases}$$

Очевидно, что полученная система эквивалентна той, которая составлена по СДУК. Следовательно, она имеет то же решение.

Ответ: стационарное распределение имеет вид $\vec{p} = \left(\frac{3}{5}, \frac{1}{5}, \frac{1}{5}\right)$

Задача 6. В заданной матрице **L** = $\begin{pmatrix} 0010 \\ 4030 \\ 0334 \\ 0110 \end{pmatrix}$ элемент λ_{ij} есть интенсивность случайного пуассоновского процесса переходов из состояния i в состояние j (размерность – количество переходов в единицу времени).

Необходимо:

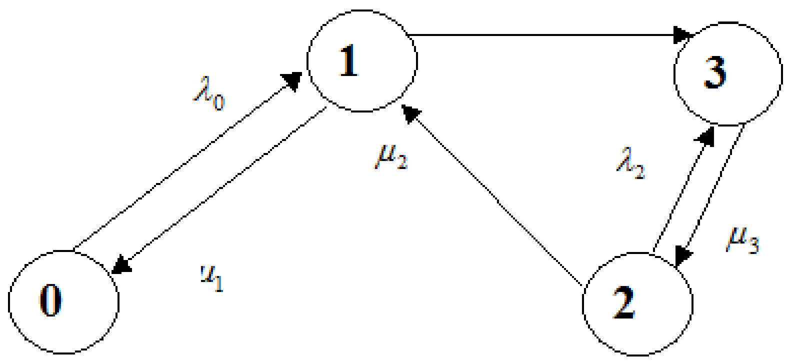
1) построить граф переходов между состояниями, ребра которого помечены соответствующими интенсивностями переходов;

2) написать систему уравнений для определения предельных вероятностей различных состояний;

3) решить эту систему уравнений, найти предельную вероятность каждого состояния.

Ответ: предельная вероятность каждого состояния $p_1 = 10/23$, $p_2 = 5/46$, $p_3 = 7/46$, $p_4 = 7/23$.

Задача 7. Найти стационарные вероятности и математическое ожидание для марковского процесса N, заданного графом переходов состояний:

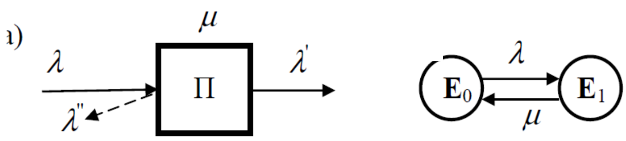


$\lambda_0 = 1,\ \lambda_1 = 1\ ,\ \lambda_2 = 1,\ \mu_1 = 1,\ \mu_2 = 2,\ \mu_3 = 3.$

2.3. Одноканальная система массового обслуживания без накопителя ($M/M/1/0$)

Рассмотрим простейшую одноканальную СМО с отказами, в которую поступает случайный поток заявок, задерживаемых в приборе на случайное время (рис. 2.6,*а*). Поскольку перед обслуживающим прибором нет накопителя, то заявка, поступившая в систему и заставшая прибор занятым, получает отказ в обслуживании и теряется. Таким образом, в системе, кроме входящего потока заявок с интенсивностью λ, образуются еще два потока: выходящий поток обслуженных в приборе заявок с интенсивностью λ' и поток необслуженных заявок (получивших отказ в обслуживании) с интенсивностью λ''.

Очевидно, что $\lambda' + \lambda'' = \lambda$.



а) *б*)

Рис. 2.6. СМО с отказами (*а*) и ее граф переходов (*б*)

1. Описание системы.

1.1. Система содержит один обслуживающий прибор (П), т.е. является *одноканальной*.

1.2. В систему поступает один класс заявок, т.е. поток заявок *однородный.*

1.3. В приборе происходит задержка (обслуживание) поступающих в систему заявок на некоторое случайное время.

1.4. Перед прибором не предусмотрены места для ожидания заявок, т.е. в системе отсутствует накопитель.

2. Предположения и допущения.

2.1. Поступающие в систему заявки образуют *простейший* поток с интенсивностью λ.

2.2. Длительность обслуживания заявок в приборе распределена по *экспоненциальному* закону с интенсивностью $\mu = 1/b$, где b – средняя длительность обслуживания.

2.3. Дисциплина буферизации – *с отказами*: заявка, поступившая в систему и заставшая прибор занятым обслуживанием другой заявки, теряется.

2.4. Дисциплина обслуживания – *в естественном порядке*: заявка, поступившая в систему и заставшая прибор свободным, принимается на обслуживание.

Очевидно, что в СМО с отказами всегда будет существовать установившийся режим, поскольку даже при больших значениях нагрузки (y >> 1) число заявок в системе не может вырасти до бесконечности. Это обусловлено тем, что с ростом нагрузки увеличивается доля заявок, получающих отказ в обслуживании.

3. Кодирование состояний случайного процесса.

В качестве параметра, описывающего состояние случайного процесса, будем рассматривать количество заявок k, находящихся в СМО.

Очевидно, что система в любой момент времени может находиться в одном из двух состояний:

E_0: $k = 0$ – в системе нет заявок (прибор простаивает);

E_1: $k = 1$ – в системе (на обслуживании в приборе) находится 1 заявка (прибор работает).

4. Размеченный граф переходов случайного процесса (см. рис. 2.6,*б*).

В процессе функционирования рассматриваемой системы в один и тот же момент времени может наступить только одно из двух возможных событий, которые приводят к изменению состояния случайного процесса, протекающего в системе.

1. *Поступление заявки в систему* с интенсивностью λ, при этом:

– если случайный процесс находится в состоянии E_0 (прибор простаивает), то произойдет переход в состояние E_1 (начнется обслуживание поступившей заявки), причем *интенсивность перехода совпадает с интенсивностью поступления* заявок в систему λ;

– если же случайный процесс находится в состоянии E_1 (прибор работает), то состояние E_1 случайного процесса не изменится, что будет соответствовать отказу в обслуживании поступившей заявке.

Таким образом, переход из состояния E_0 в состояние E_1 происходит с интенсивностью λ.

2. *Завершение обслуживания заявки*, находящейся в приборе. Это событие может наступить только в том случае, если в приборе на обслуживании находится заявка, т.е. случайный процесс находится в состоянии E_1. При этом происходит переход в состояние E_0, причем *интенсивность перехода совпадает с интенсивностью обслуживания* заявки в приборе μ.

5. Диаграммы функционирования системы.

Рассмотрим диаграммы функционирования системы и покажем, что случайный процесс, протекающий в системе, при сформулированных выше предположениях является марковским.

На рис. 2.7 показаны диаграммы следующих процессов:

а) **поступления** в СМО заявок, интервалы между которыми в случае простейшего потока распределены по экспоненциальному закону;

б) **перехода** из состояния E_0 в состояние E_1 и обратно, в которых может находиться система, при этом время нахождения случайного процесса в состоянии E_1 равно длительности обслуживания заявки в приборе, которая представляет собой случайную величину, распределенную по экспоненциальному закону;

в) **выхода** из системы обслуженных заявок в моменты времени t'_1, t'_3, t'_4, t'_7;

г) **выхода** из системы необслуженных заявок, получивших отказ из-за занятости прибора в моменты времени $t''_2, t''_5, t''_6, t''_8$;

д) **формирования** интервалов времени между соседними переходами случайного процесса.

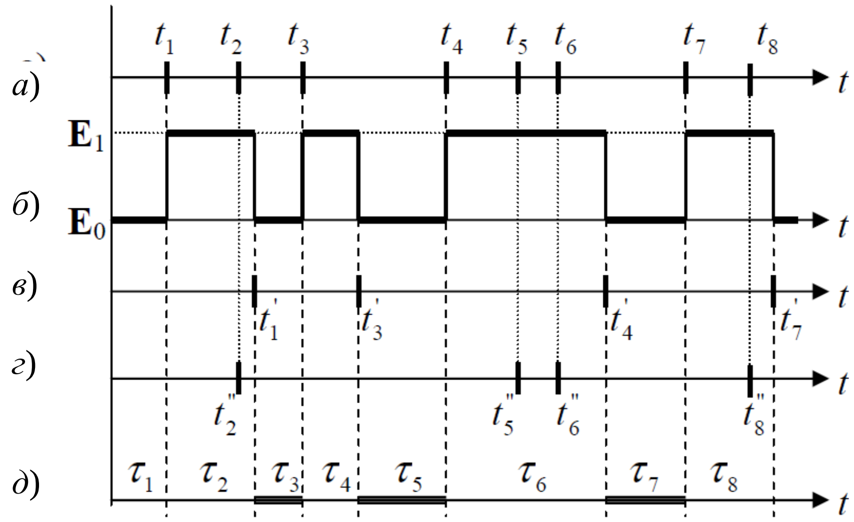


Рис. 2.7. Диаграммы процессов в СМО с отказами

Дискретный случайный процесс с непрерывным временем будет марковским, если интервалы между соседними переходами распределены по экспоненциальному закону.

В нашем случае интервал τ_1 представляет собой интервал между поступающими заявками, который для простейшего потока имеет экспоненциальное распределение. Интервалы τ_2, τ_4, τ_6, τ_8, как видно из диаграммы, представляют собой время нахождения случайного процесса в состоянии E_1, равное длительности обслуживания заявки в приборе, которая распределена по экспоненциальному закону. Таким образом, интервалы τ_1, τ_2, τ_4, τ_6, τ_8

распределены по экспоненциальному закону и, следовательно, удовлетворяют сформулированному выше условию.

Рассмотрим теперь выделенные интервалы τ_3, τ_5, τ_7. Каждый из этих интервалов представляет собой промежуток времени от момента завершения обслуживания некоторой заявки до момента поступления новой заявки, принимаемой на обслуживание в приборе. По *свойству экспоненциального распределения*: в случае экспоненциального распределения интервалов времени между двумя событиями *интервал времени от любого случайного момента до момента наступления очередного события* имеет *такое же экспоненциальное распределение с тем же параметром*. В соответствии с этим свойством интервалы τ_3, τ_5, τ_7 имеют *экспоненциальное распределение* с параметром λ и, следовательно, также удовлетворяют сформулированному выше условию для марковского процесса.

Таким образом, случайный процесс, протекающий в системе с простейшим потоком заявок и экспоненциальным обслуживанием, является марковским.

6. Матрица интенсивностей переходов.

Графу переходов соответствует **матрица интенсивностей переходов**:

$$\mathbf{G} = \begin{array}{c|cc} \mathbf{E}_i & 0 & 1 \\ \hline \mathit{0} & -\lambda & \lambda \\ 1 & \mu & -\mu \end{array}$$

Действительно, переход из состояния E_0 в состояние E_1 соответствует поступлению заявки в систему с интенсивностью λ, а переход из состояния E_1 в состояние E_0 соответствует завершению обслуживания заявки в приборе с интенсивностью μ.

Диагональные элементы матрицы определяются из условия: сумма элементов каждой строки должна быть равна нулю.

7. Система уравнений.

Система уравнений для определения стационарных вероятностей, составленная по графу переходов с применением *правила 1*, имеет вид

$$\lambda p_0 = \mu p_1;$$

$$\mu p_1 = \lambda p_0;$$

$$p_0 + p_1 = 1,$$

где последнее уравнение представляет собой нормировочное условие.

Учитывая, что первое и второе уравнения одинаковы (или, как говорят математики, линейно зависимы), и удаляя одно из них, окончательно получим:

$$\lambda p_0 = \mu p_1;$$

$$p_0 + p_1 = 1.$$

Решая эту систему, получим следующие значения стационарных вероятностей состояний марковского процесса:

$$p_0 = \frac{\mu}{\mu + \lambda} = \frac{1}{1 + y}, \quad p_1 = \frac{\lambda}{\mu + \lambda} = \frac{y}{1 + y},$$

где $y = \lambda/\mu$ – нагрузка системы.

8. Расчет характеристик СМО.

Для расчета характеристик СМО можно воспользоваться следующими математическими зависимостями:

1) нагрузка: $y = \lambda/\mu = \lambda b$ (по определению);

2) загрузка определяется как *вероятность работы прибора*: $\rho = p_1$, и не совпадает с нагрузкой даже в случае $y < 1$, что характерно для систем с отказами и потерями заявок, причем всегда $\rho < y$;

3) коэффициент простоя системы определяется как вероятность отсутствия заявок в системе или, по определению, через загрузку системы: $\eta = p_0 - \rho$;

4) среднее число заявок в системе: $m = p_1 = \rho$, определяемое как математическое ожидание случайной величины: в системе может находиться либо ноль заявок с вероятностью p_0, либо одна заявка с вероятностью p_1, тогда среднее число заявок равно $m = 0 \times p_0 + 1 \times p_1 = p_1$;

5) вероятность потери заявок в результате отказа в обслуживании из-за занятости прибора в соответствии с формулами Литтла совпадает с вероятностью того, что система занята обслуживанием заявок:

$$\pi = \pi_n = 1 - \frac{\rho}{y} K = 1 - \frac{p_1}{y} = \frac{y}{1 + y} = p_1,$$

где учтено, что для рассматриваемой СМО без накопителя $p_1 = \dfrac{y}{1 + y}$;

6) производительность системы (абсолютная пропускная способность): $\lambda' = (1 - \pi)\lambda$, определяемая как интенсивность потока обслуженных заявок на выходе системы;

7) интенсивность потока не обслуженных заявок, т.е. получивших отказ: $\lambda'' = \pi\lambda$;

8) среднее время пребывания заявок в системе: $u = m/\lambda' = b$ (определяется по формуле Литтла и, как следовало ожидать, равно средней длительности обслуживания заявок; отметим, что в формуле Литтла используется интенсивность λ' потока *обслуженных* заявок, а не входящего потока λ).

В литературе часто используются: $Q = p_0$ – относительная пропускная способность, $\lambda' = \lambda Q$ – абсолютная пропускная способность.

9. Анализ свойств системы.

Анализ полученных зависимостей (рис. 2.8) показывает, что с ростом нагрузки коэффициент простоя системы, равный p_0, уменьшается, а загрузка системы, определяемая как вероятность p_1 того, что прибор работает (а также среднее число заявок в системе и вероятность отказа), увеличивается, причем их сумма всегда равна единице. При $y \to \infty$ коэффициент простоя $\eta \to 0$, в то время как загрузка $\rho \to 1$.

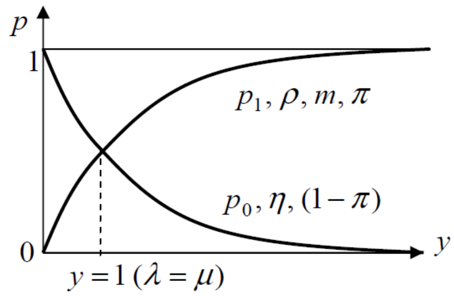


Рис. 2.8. Характеристики СМО

Заметим также, что нагрузка системы определяется через стационарные вероятности как отношение вероятности работы системы к вероятности простоя: $y = p_1/p_0$ (что легко может быть получено из выражений для p_0 и p_1) или, что то же самое, через загрузку и коэффициент простоя: $y = \rho/\eta$.

Задача. В парикмахерской работает один мастер. Время обслуживания распределено по показательному закону со средним 12 мин. Клиент, пришедший в парикмахерскую, когда мастер занят, не ожидает обслуживания, а покидает парикмахерскую. Поток клиентов – простейший с интенсивностью 8 клиентов/ч. Найти показатели эффективности работы данной парикмахерской.

Решение. Имеем $\lambda = 8$ (клиентов/ч), среднее время обслуживания b = 12 мин = 0,2 ч. Следовательно, интенсивность потока обслуживания $\mu = 1/b = 5$ 1/ч. По формулам для p_0 и p_1 получаем, что относительная пропускная способность p_0 = 0,38, т.е. в среднем 38 % поступающих клиентов будут обслужены. Соответственно вероятность отказа в обслуживании составит p_1 = 0,62 (т.е. 62 %). Абсолютная пропускная способность СМО $\lambda' = \lambda p_0 = 8 \cdot 0{,}38 =$ $= 3{,}04$, т.е. в среднем в 1 ч будут обслужены 3 клиента.

Задачи для самостоятельного решения.

1. Рассматривается круглосуточная работа пункта проведения профилактического осмотра автомашин с одним каналом (одной группой проведения осмотра). На осмотр и выявление дефектов каждой машины затрачивается в среднем 0,5 ч. На осмотр поступает в среднем 36 машин в 1 сут. Потоки заявок и обслуживания – простейшие. Если машина, прибывшая в пункт осмотра, не застает ни одного канала свободным, она покидает пункт осмотра необслуженной. Определить предельные вероятности состояний и характеристики обслуживания профилактического пункта осмотра и сделать вывод об эффективности его работы.

Ответ: p_0 = 0,571; Q = 0,571; $p_1 = \pi$ = 0,429; $\lambda' = A = \lambda p_0$ = 0,857.

2. Одноканальная СМО с отказами представляет собой одну телефонную линию, на вход которой поступает простейший поток вызовов с интенсивностью 0,4 вызовов/мин. Средняя продолжительность разговора 3 мин; время разговора имеет показательное распределение. Найти предельные вероятности состояний и характеристики обслуживания СМО. Сравнить пропускную способность СМО с номинальной, которая была бы, если разговор в точности длился 3 мин, а заявки шли одна за другой регулярно, без перерывов.

Ответ: p_0 = 0,455; Q = 0,455; $p_1 = \pi$ = 0,545; λ' = 0,182, номинальная пропускная способность $\lambda'_{\text{ном}}$ = 1/3 = 0,333 почти вдвое больше.

2.4. Многоканальная система массового обслуживания без накопителя, или с отказом, $(M/M/N/0)$

Рассмотрим многоканальную СМО с отказами, в которую поступает случайный поток заявок, задерживаемых в приборе на случайное время (рис. 2.9). Заявка, поступившая в систему и заставшая прибор занятым, получает отказ в обслуживании и теряется. Таким образом, в системе, кроме входящего потока заявок с интенсивностью λ, образуются еще два потока: выходящий поток обслуженных в приборе заявок с интенсивностью λ' и поток необслуженных заявок (получивших отказ в обслуживании) с интенсивностью λ''. Очевидно, что $\lambda' + \lambda'' = \lambda$.

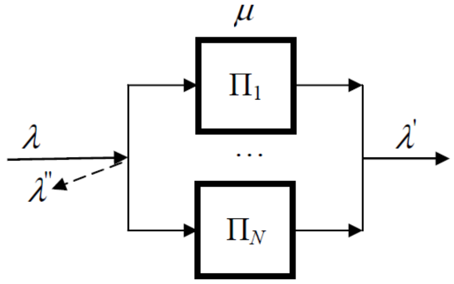


Рис. 2.9. Многоканальная СМО с отказами

1. Описание системы

1.1. Система содержит N обслуживающих приборов $П_1$, …, $П_N$, т.е. является *многоканальной*.

1.2. В систему поступает один класс заявок (поток *однородный*).

1.3. Все приборы идентичны, т.е. любая заявка может быть обслужена любым прибором за одно и то же случайное время.

1.4. В системе отсутствует накопитель.

2. Предположения и допущения.

2.1. Поступающие в систему заявки образуют *простейший* поток с интенсивностью λ.

2.2. Длительность обслуживания заявок в любом приборе распределена по экспоненциальному закону с интенсивностью $\mu = \frac{1}{b}$, где b – средняя длительность обслуживания заявок в приборе.

2.3. Дисциплина буферизации – *с отказами*: заявка, поступившая в систему и заставшая все приборы занятыми обслуживанием других заявок, теряется.

2.4. Дисциплина обслуживания – *в естественном порядке*: заявка, поступившая в систему, принимается на обслуживание, если есть хотя бы один свободный прибор. Если заявка застала свободными несколько приборов, то она направляется в один из них случайным образом.

3. Кодирование состояний случайного процесса.

В качестве параметра, описывающего состояние случайного процесса, как и ранее, будем рассматривать количество заявок k, находящихся в СМО. При этом система в любой момент времени может находиться в одном из $(N+1)$ состояний:

E_0: $k = 0$ – в системе нет заявок (система простаивает);

E_1: $k = 1$ – в системе находится 1 заявка (один прибор работает, остальные простаивают);

E_2: $k = 2$ – в системе находится 2 заявки (два прибора работают, остальные простаивают);

…

E_N: $k = N$ – в системе находится N заявок (все приборы работают).

3. Размеченный граф переходов случайного процесса представлен на рис. 2.10.

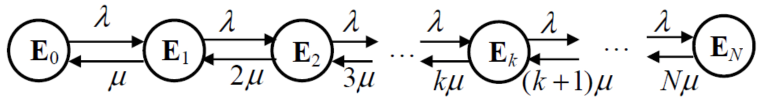


Рис. 2.10. Граф переходов марковского процесса

В один и тот же момент времени в системе может произойти только одно из двух событий, которые приводят к изменению состояния случайного процесса.

1. *Поступление заявки в систему* с интенсивностью λ. При этом:

– если случайный процесс находится в состоянии E_k, причем $k < N$, то произойдет переход в состояние E_{k+1} (начнется обслуживание поступившей заявки в одном из свободных приборов), причем интенсивность перехода равна интенсивности поступления λ;

– если же случайный процесс находится в состоянии E_N (все приборы заняты обслуживанием заявок), то состояние E_N случайного процесса не изменится, что будет соответствовать отказу в обслуживании поступившей заявке.

Таким образом, переход из состояний E_k в состояние E_{k+1} (при $k < N$) происходит с интенсивностью λ.

2. *Завершение обслуживания заявки* в одном из приборов с интенсивностью μ.

Это событие может наступить только в том случае, если в системе на обслуживании находится хотя бы одна заявка, т.е. случайный процесс находится в состояниях E_1, E_2, …, E_N. При этом случайный процесс переходит соответственно в состояния E_0, E_1, …, E_{N-1}, причем интенсивности перехода различны. Действительно, если в системе обслуживается только одна заявка (состояние E_1), то интенсивность перехода в состояние E_0 равна μ. Если же в системе обслуживается две заявки (состояние E_2), т.е. работают два прибора, то переход случайного процесса в состояние E_1 возможен либо в результате завершения обслуживания заявки в первом приборе с интенсивностью μ, либо в результате завершения обслуживания заявки во втором приборе с такой же интенсивностью μ, причем вероятность завершения обслуживания заявок в обоих приборах в один и тот же момент времени равна нулю. Таким образом, интенсивность перехода из состояния E_2 в состояние E_1 будет равна 2μ (как сумма интенсивностей двух рассмотренных способов).

В общем случае, если в многоканальной системе на обслуживании находится k =1,2,..., N заявок (случайный процесс находится в состоянии E_k), то интенсивность перехода в состояние E_{k-1} будет равна kμ.

По аналогии с предыдущим примером здесь и в последующих примерах можно показать, что случайный процесс, протекающий в системе, при сформулированных предположениях является марковским.

5. Матрица интенсивностей переходов.

Графу переходов соответствует матрица интенсивностей переходов:

	E$_i$	0	1	2	…	N–1	N
	0	–λ	λ	0	…	0	0
	1	μ	–(λ + μ)	λ	…	0	0
G=	2	0	2μ	–(λ + 2μ)	…	0	0
	…	…	…	…	…	…	…
	N–1	0	0	0	…	–(λ + (N – 1)μ)	λ
	N	0	0	0	…	Nμ	–Nμ

6. Система уравнений.

Система уравнений для определения стационарных вероятностей имеет вид

$$\lambda p_0 = \mu p_1$$

$$(\lambda + \mu) p_1 = \lambda p_0 + 2\mu p_2;$$

$$\dots$$

$$(\lambda + k\mu) p_k = \lambda p_{k-1} + (k+1)\mu p_{k+1};$$

$$\dots$$

$$N\mu p_N = \lambda p_{N-1};$$

$$p_0 + p_1 + p_2 + \dots + p_N = 1.$$

Используя метод математической индукции, можно показать:

$$p_k = \frac{y^k}{k!} p_0, \quad k = 0, 1, \ldots, N,$$

где $y = \lambda b$ – нагрузка системы.

Подставляя полученное выражение в последнее уравнение системы линейных алгебраических уравнений, найдем вероятность простоя системы:

$$p_0 = \frac{1}{\sum_{i=0}^{N} \frac{y^i}{i!}}.$$

Тогда стационарные вероятности состояний марковского случайного процесса, протекающего в многоканальной СМО с отказами, имеют вид

$$p_k = \frac{\frac{y^k}{k!}}{\sum_{i=0}^{N} \frac{y^i}{i!}}, \quad k = 0, 1, \ldots, N.$$

Получили формулы Эрланга. Из последнего выражения при $N = 1$, как частный случай, вытекает результат, полученный в предыдущем примере для одноканальной СМО с отказами.

Задание на самостоятельную работу: проверить полученные выражения, используя метод математической индукции.

7. Расчет характеристик СМО.

Для расчета характеристик СМО можно воспользоваться следующими математическими зависимостями:

1) нагрузка: $y = \lambda/\mu = b$ (по определению);

2) загрузка $\rho = \frac{1}{N} \sum_{k=0}^{N} k p_k$, учитывающая долю ($k/N$) работающих приборов; действительно, система загружена полностью, когда работают все приборы, если же из 10 приборов работает один, то система загружена на 10 %, если работают 5 приборов, то система загружена на 50 %;

3) коэффициент простоя системы: $\eta = \frac{1}{N} \sum_{k=0}^{N} (N-k) p_k = 1 - \rho$;

4) среднее число заявок в системе, равное среднему числу работающих приборов:

$$m = \sum_{k=1}^{N} k p_k = N\rho;$$

5) среднее число простаивающих приборов: $\hat{N} = N - m$;

6) вероятность отказа в обслуживании, определяемая как вероятность того, что все приборы заняты обслуживанием заявок:

$$\pi = p_N = \frac{\dfrac{y^N}{N!}}{\sum_{i=0}^{N} \dfrac{y^i}{i!}} = \frac{y^n}{n!} p_0;$$

7) производительность системы, определяемая как интенсивность потока обслуженных заявок: $\lambda' = \lambda(1 - \pi) = A$;

8) интенсивность потока необслуженных заявок, т.е. получивших отказ: $\lambda'' = \lambda\pi$ или $Q = P_{\text{обсл}} = 1 - P_{\text{отк}}$;

9) среднее время пребывания заявок в системе

$$u = T_{\text{сист}} = \frac{m}{\lambda'} = b.$$

8. Анализ свойств системы.

Анализ свойств многоканальной СМО без накопителя показывает, что с увеличением нагрузки уменьшается вероятность простоя системы и увеличивается загрузка системы, а вместе с ней и число работающих приборов и вероятность отказа.

Распишем еще раз многоканальную СМО с отказами.

Рассмотрим классическую задачу Эрланга. Имеется n каналов, на которые поступает поток заявок с интенсивностью λ. Поток обслуживания каждого канала имеет интенсивность μ. Найти предельные вероятности состояний системы и показатели ее эффективности. В качестве показателей эффективности многоканальной СМО с отказами будем рассматривать: λ' – абсолютную пропускную способность СМО; Q – относительную пропускную способность; π – вероятность отказа; $\bar{k}_{\text{зан}}$ – среднее число занятых каналов. Размеченный граф состояний представлен на рис. 2.10 с теми же состояниями $E_0 \ldots E_N$:

$$\pi = \frac{y^n}{n!} p_0;$$

$$Q = 1 - \pi = 1 - \frac{y^n}{n!} p_0 = \frac{\overline{k}_{\text{зан}}}{y};$$

$$\lambda' = \lambda Q = \lambda \left(1 - \frac{y^n}{n!} p_0 \right);$$

$$\overline{k}_{\text{зан}} = \sum_{k=0}^{n} k p_k = \frac{\lambda'}{\mu} y = \left(1 - \frac{y^n}{n!} p_0 \right) = \frac{A}{\mu};$$

$k_з$ – коэффициент занятости каналов:

$$k_з = \frac{n_з}{n} = \frac{y P_{\text{обсл}}}{n}.$$

Пример 11. Рассматривается работа автозаправочной станции (АЗС) с тремя заправочными колонками. Если заняты все три колонки, то машина не встает в очередь, а покидает АЗС. Среднее время заправки автомобиля 3 мин. Интенсивность потока автомобилей – 0,25 ед./мин. Найти предельные вероятности состояний и показатели эффективности работы АЗС.

Решение. В принятых обозначениях условия примера можно записать в виде λ = 0,25 (ед./мин), $\bar{b}$ = 3 мин. Тогда интенсивность потока обслуживания $\mu = \frac{1}{\bar{b}} = \frac{1}{3} =$ (1/мин). Интенсивность нагрузки канала $y = \lambda/\mu = 0{,}75$, т.е. в среднем за время заправки одного автомобиля, равного 3 мин, поступит 0,75 новых заявок. Найдем предельные вероятности состояний СМО:

$$p_0 = \left(1 + 0{,}75 + \frac{0{,}75^2}{2!} + \frac{0{,}75^3}{3!} \right)^{-1} = 0{,}476;$$

$$p_1 = 0{,}75 \cdot 0{,}476 = 0{,}357;$$

$$p_2 = \frac{0{,}75^2}{2!} 0{,}476 = 0{,}134;$$

$$p_3 = \frac{0{,}75^3}{3!} 0{,}476 = 0{,}033.$$

Таким образом, в предельном стационарном режиме 47,6 % времени АЗС не работает из-за отсутствия заявок; 35,7 % времени занята только одна колонка; 13,4 % времени занято две колонки, а 3,3 % времени занято три колонки. Вероятность отказа в обслуживании $\pi = p_3 = 0{,}033$, т.е. в среднем 3,3 % поступающих заявок (автомобилей) не будут обслужены. Относительная пропускная способность (вероятность обслуживания): $Q = 1 - \pi = 1 - 0,033 = 0,967$, т.е. в среднем 96,7 % поступающих заявок (автомобилей) будут обслужены.

Абсолютная пропускная способность СМО, т.е. среднее число заявок (автомобилей), обслуживаемых в единицу времени: $\lambda' = \lambda Q = \lambda p_0 = 0,25 \cdot 0,967 = 0,242$.

Среднее число занятых каналов: $\bar{k}_{\text{зан}} = \lambda' / \mu =$ $= 0,242/(1/3) = 0,725$; т.е. в трехканальной СМО среднее число занятых каналов меньше одного (Эффективно ли это?).

Пример 12. В коммерческом банке в отделе кредитования физических лиц три телефонные линии. В среднем в отдел поступает 75 звонков в 1 ч. Среднее время предварительных переговоров составляет 2 мин. Определить основные характеристики системы, оцените ее работу.

Решение. В задаче $n = 3$, $\lambda = 75$ звонков в 1 ч, $\mu = 1/2 = 0,5$ звонков/мин или $60 \cdot 0,5 = 30$ звонков в 1 ч, $y = 75/30 = 2,5$. Находим вероятность того, что обслуживанием не занят ни один канал:

$$p_0 = \left(1 + 2,5 + \frac{2,5^2}{2!} + \frac{2,5^3}{3!}\right)^{-1} = 0,11 \approx 11\ \%;$$

$$P_{\text{отк}} = \frac{y^n}{n} p_0 = \frac{2,5^3}{3!} \cdot 0,11 = 0,282 \approx 28\ \%;$$

$$Q = P_{\text{обсл}} = 1 - P_{\text{отк}} = 1 - 1,282 = 0,718;$$

$$\bar{k}_{\text{зан}} = \frac{A}{\mu} = y \cdot P_{\text{обсл}} = 2,5 \cdot 0,718 = 1,795 \approx 2 \text{ канала}.$$

Вывод. Согласно представленным расчетам в отделе необходима еще одна телефонная линия, так как входящий поток обслуживается только на 72 %, в отделе в постоянном режиме заняты две телефонные линии – это неудовлетворительно характеризует работу системы в целом.

Задачи для самостоятельного решения

1. Рассматривается круглосуточная работа пункта проведения профилактического осмотра автомашин с четырьмя каналами (четыре группы проведения осмотра). На осмотр и выявление дефектов каждой машины затрачивается в среднем 0,5 ч. На осмотр поступает в среднем 36 машин в 1 сут. Потоки заявок и обслужи-

вания – простейшие. Если машина, прибывшая в пункт осмотра, не застает ни одного канала свободным, она покидает пункт осмотра необслуженной. Определить предельные вероятности состояний и характеристики обслуживания профилактического пункта осмотра. Найти минимальное число каналов, при котором относительная пропускная способность пункта осмотра будет не менее 0,9.

Ответ: $p_0 = 0,473$; $p_1 = 0,355$; $p_2 = 0,133$; $p_3 = 0,033$; $p_4 = \pi = 0,006$. $Q = 0,994$, $\lambda' = 1,491$, $n = 3$.

2. Имеется двухканальная простейшая СМО с отказами. На ее вход поступает поток заявок с интенсивностью 4 заявки/ч. Среднее время обслуживания одной заявки 0,8 ч. Каждая обслуженная заявка приносит доход 4 ден. ед. Содержание каждого канала обходится 2 ден. ед./ч. Выясните, выгодно или невыгодно в экономическом отношении увеличить число каналов до трех.

Ответ: при $n = 2$, $p_0 = 0,107$; $p_2 = 0,550$; $Q = 0,450$; $\lambda' = 1,8$ заявки/ч, доход $D = 7,20$ ден. ед./ч; при $n = 3$, $p'_0 = 0,0677$, $p'_3 = 0,371$, $Q' = 0,627$, $\lambda' = 2,52$ заявки/ч, доход $D' = 10,08$ ден. ед./ч. Увеличение числа каналов с двух до трех выгодно, так как увеличение дохода на 2,88 ден. ед./ч превосходит увеличение затрат – 2 ден. ед./ч).

2.5. Одноканальная система массового обслуживания с накопителем ограниченной емкости ($M/M/1/r$)

1. Описание системы.

1.1. Система (рис. 2.11) содержит один обслуживающий прибор (П), т.е. является *одноканальной*.

1.2. Поток поступающих в систему заявок *однородный.*

1.3. Длительность обслуживания заявок в приборе – величина *случайная*.

1.4. Перед прибором имеется r мест для заявок, ожидающих обслуживания и образующих очередь, т.е. в системе имеется накопитель *ограниченной* емкости: $r < \infty$.

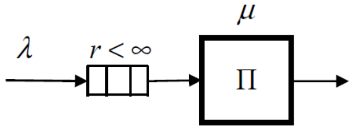


Рис. 2.11. СМО с накопителем ограниченной емкости

2. Предположения и допущения.

2.1. Поступающие в систему заявки образуют *простейший* поток с интенсивностью λ.

2.2. Длительность обслуживания заявок в приборе распределена по *экспоненциальному* закону с интенсивностью $\mu = \frac{1}{b}$, где b – средняя длительность обслуживания заявок в приборе.

2.3. Дисциплина буферизации – *с потерями*: заявка, поступившая в систему и заставшая накопитель заполненным, теряется.

2.4. Дисциплина обслуживания – *в порядке поступления* по правилу «первым пришел – первым обслужен» (FIFO).

В СМО с накопителем ограниченной емкости всегда существует установившийся режим, поскольку длина очереди не будет расти до бесконечности даже при больших значениях нагрузки.

3. Кодирование состояний марковского процесса.

В качестве параметра, описывающего состояние марковского процесса, будем рассматривать количество заявок k, находящихся в СМО (в приборе и в накопителе). Тогда марковский процесс в любой момент времени может находиться в одном из следующих $(r + 2)$ -х состояний:

E_0: $k = 0$ – в системе нет ни одной заявки;

E_1: $k = 1$ – в системе находится 1 заявка на обслуживании в приборе;

E_2: $k = 2$ – в системе находятся 2 заявки: одна – на обслуживании в приборе и вторая ожидает в накопителе;

…

E_{r+1}: $k = r +1$ – в системе находятся $(r +1)$ заявок: одна – на обслуживании в приборе и r – в накопителе.

4. Размеченный граф переходов случайного процесса представлен на рис. 2.12.

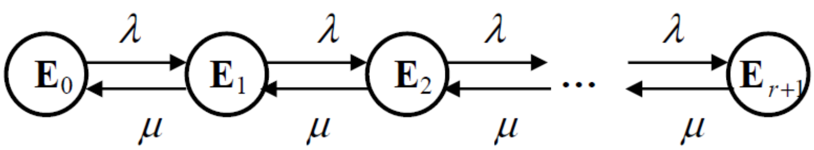


Рис. 2.12. Граф переходов случайного процесса

В один и тот же момент времени в системе может произойти только одно событие:

– поступление заявки с интенсивностью λ, что соответствует увеличению на единицу числа заявок в системе и переходу случайного процесса в состояние с номером на единицу больше;

– завершение обслуживания заявки в приборе с интенсивностью μ, что соответствует уменьшению числа заявок в системе и переходу случайного процесса в состояние с номером, на единицу меньше.

Задание на самостоятельную работу: по графу переходов рис. 2.12 построить матрицу интенсивностей переходов.

5. Система уравнений.

Составим по графу переходов систему уравнений для определения стационарных вероятностей:

$$\begin{cases} \lambda p_0 = \mu p_1 \\ \ldots \\ (\lambda+\mu)\,p_1 = \lambda p_0 + 2\mu p_2; \\ (\lambda+\mu)\,p_2 = \lambda p_1 + \mu p_3; \\ \ldots \\ \mu p_{r+1} = \lambda p_r; \\ \ldots \\ p_0 + p_1 + p_2 + \ldots + p_{r+1} = 1. \end{cases}$$

Используя метод математической индукции, можно показать, что $p_k = y^k p_0$, где $y = \lambda b$ – нагрузка системы.

Подставляя полученное выражение в последнее уравнение системы линейных алгебраических уравнений, найдем вероятность простоя системы в зависимости от нагрузки:

$$p_0 = \frac{1}{\sum_{k=0}^{r+1} y^k} = \begin{cases} \dfrac{1-y}{1-y^{r+2}}, & y \neq 1, \\ \dfrac{1}{r+2}, & y = 1. \end{cases}$$

Тогда стационарные вероятности состояний p_k (k = 0, 1, …, r + 1) равны

$$p_k = \begin{cases} \dfrac{(1-y)\,y^k}{1-y^{r+2}}, & y \neq 1, \\ \dfrac{y^k}{r+2}, & y = 1. \end{cases}$$

Задание на самостоятельную работу: вывести представленные математические зависимости.

5. Расчет характеристик СМО.

Характеристики СМО при найденных значениях стационарных вероятностей состояний случайного процесса могут быть рассчитаны по следующим формулам:

1) нагрузка: $y = \lambda/\mu = \lambda b$;

2) загрузка: $\rho = \sum_{k=1}^{r+1} p_k = 1 - p_0$;

3) коэффициент простоя системы: $\eta = p_0 = 1 - \rho$;

4) среднее число заявок в очереди: $l = L_{\text{оч}} = \sum_{k=2}^{r+1} (k-1) p_k$;

5) среднее число заявок в системе: $m = \mathrm{L}_{\text{сист}} = \sum_{k=1}^{r+1} k p_k$;

6) вероятность потери заявок: $\pi = p_{r+1} = y^{r+1} p_0$.

Задание на самостоятельную работу: доказать, что вероятность потери заявок равна вероятности того, что система заполнена, т.е. в накопителе нет свободных мест для вновь поступающих заявок.

7) производительность системы (интенсивность потока обслуженных заявок): $\lambda' = \lambda(1 - \pi)$;

8) интенсивность потока потерянных заявок: $\lambda'' = \lambda\pi$;

9) среднее время ожидания заявок: $w = T_{\text{оч}} l / \lambda'$;

10) среднее время пребывания заявок: $u = T_{\text{сист}} = m/\lambda' = w + b$.

Распишем еще раз одноканальную СМО с ограниченной длиной очереди.

В качестве показателей эффективности одноканальной СМО с ограниченной длиной очереди будем рассматривать:

λ' – абсолютную пропускную способность СМО (А);

Q – относительную пропускную способность;

π – вероятность отказа (p_{r+1}); $\pi = \dfrac{1-y}{1-y^{r+2}}$;

m – среднее число находящихся в системе заявок ($L_{\text{сист}}$);

u – среднее время пребывания заявки в системе ($T_{\text{сист}}$);

l – средняя длина очереди ($L_{\text{оч}}$);

w – среднее время ожидания в очереди ($T_{\text{оч}}$).

Вероятности состояний определяются уравнениями ($y = \lambda/\mu$):

$$Q = 1 - \pi, \tag{2.16}$$

$$A = \lambda' = \lambda Q, \tag{2.17}$$

$$m=\sum_{n=0}^{r+1} np_n, \tag{2.18}$$

$$u=\frac{m}{\lambda}, \tag{2.19}$$

$$l=\begin{cases}\dfrac{1-y^r\left(r-rp+1\right)}{\left(1-y\right)^2}y^2 p_0, \text{ если } y\neq 1;\\ \dfrac{r\left(r+1\right)}{2\left(r+2\right)}, \qquad\qquad \text{если } y=1,\end{cases} \tag{2.20}$$

$$w=\frac{l}{\lambda}. \tag{2.21}$$

Пример 13. Автозаправочная станция представляет собой СМО с одним каналом обслуживания (одной колонкой). Площадка при станции допускает пребывание в очереди на заправку не более пяти машин одновременно (m = 5). Если в очереди уже находятся пять машин, очередная машина, прибывшая к станции, в очередь не становится. Поток машин, прибывающих для заправки, имеет интенсивность 2 машины в 1 мин. Интенсивность потока обслуживания составляет $\mu = 2$.

Определите характеристики СМО и сделайте вывод об эффективности ее работы.

Решение:

$$\lambda/\mu = y = 1,\ p_0=\frac{1}{r+2}=\frac{1}{7},$$

$$p_{r+1}=y^{r+1}p_0 \Rightarrow p_1 \Rightarrow p_2=...p_6=\frac{1}{7},$$

$$\pi = p_6=\frac{1}{7},\ Q=1-\pi=1-\frac{1}{7}=\frac{6}{7}.$$

Среднее число находящихся в системе заявок:

$$m=0\cdot\frac{1}{7}+1\cdot\frac{1}{7}+2\cdot\frac{1}{7}+3\cdot\frac{1}{7}+4\cdot\frac{1}{7}+5\cdot\frac{1}{7}+6\cdot\frac{1}{7}=3.$$

Среднее время пребывания машины в системе u = 3/2 = 1,5.

Средняя длина очереди $l=\dfrac{5\times 6}{2\times 7}=2{,}1$ [машин].

Среднее время ожидания в очереди: u = 2,1/2 = 1,05 [мин].

Вывод. Каждому седьмому клиенту отказывают в обслуживании, следовательно, эффективность СМО низкая.

Пример 14. В небольшом магазине самообслуживания установлено, что поток покупателей является простейшим с интенсивностью 1 покупатель в 1 мин. В этом магазине установлен один кассовый аппарат, позволяющий добиться такой производительности труда, при которой среднее время обслуживания одного клиента составляет примерно 1,25 мин. Определить характеристики СМО при условии, что очередь ограничена контролером при входе в зал самообслуживания тремя покупателями.

Решение. Найдем интенсивность потока обслуживания: $\mu = 1/b = 1/1{,}25 = 0{,}8$. Нагрузка $y = \lambda/\mu = 1/1{,}25 = 0{,}8$. Найдем предельные вероятности:

$$p_0 = \frac{1-1{,}25}{1-1{,}25^{3+2}} \approx 0{,}122,\ p_1 = 1{,}25 \cdot 0{,}122 \approx 0{,}152,$$

$$p_2 = 1{,}25^2 \cdot 0{,}122 = 0{,}191, \quad p_3 = 1{,}25^3 \cdot 0{,}122 \approx 0{,}238,$$

$$p_4 = 1{,}25^4 \cdot 0{,}122 = 0{,}297.$$

Вероятность отказа $\pi = p_4 \approx 0{,}297$.

Относительная пропускная пособность СМО: $Q = 1 - 0{,}297 = 0{,}703$.

Абсолютная пропускная способность СМО: $\lambda' = 1 \cdot 0{,}703 = 0{,}703$ покупателей в минуту.

Среднее число покупателей у кассы: $m = 0 \cdot 0{,}122 + 1 \cdot 0{,}152 + 2 \cdot 0{,}191 + 3 \cdot 0{,}238 + 4 \cdot 0{,}297 = 2{,}436$.

Среднее время пребывания у кассы: $u = 2{,}436/1 = 2{,}436$ мин.

Среднее число покупателей в очереди:

$$l = 1{,}25^2 \frac{1-1{,}25^3\left(3-3\cdot 1{,}25+1\right)}{(1-1{,}25)^2} \cdot 0{,}122 \approx 1{,}56.$$

Среднее время ожидания покупателей в очереди:

$$w = \frac{1{,}56}{1} = 1{,}56 \text{ мин.}$$

Вывод. Вероятность простоя кассира мала, среднее время обслуживания покупаля не большое, вероятность отказа примерно 0,297. Таким образом, можно сказать, что система работает эффективно.

Задачи для самостоятельного решения

1. На автомойке оборудован один блок для обслуживания и есть место для очереди. Автомобили прибывают по пуассоновскому распределению с интенсивностью 5 машин/ч. Среднее время обслуживания одной машины – 10 мин. Найти все средние характеристики СМО.

Ответ: $m = 5$, $l = 4{,}2$, $w = 50$ мин., $u = 1$ ч.

2. Автосервис (пост диагностики) представляет собой одноканальную СМО. Число стоянок для автомобилей, ожидающих обслуживания, ограничено и равно 3. Если все стоянки заняты, т.е. в очереди уже находятся три автомобиля, то очередной автомобиль, прибывший в автосервис на диагностику, в очередь на обслуживание не становится. Поток автомобилей, прибывающих на диагностику, распределен по показательному закону Пуассона и имеет интенсивность 0,85 (автомобиля в 1 ч). Время диагностики распределено по показательному закону и в среднем равно 1,05 ч. Определить вероятностные характеристики работы СМО и сделать вывод об эффективности ее работы.

Ответ: $Q = 0{,}842$; $\lambda' = 0{,}716$; $m = 1{,}77$; $l = 1{,}02$, $w = 1{,}423$ ч, $u = 2{,}473$ ч.

2.6. Одноканальная система массового обслуживания с накопителем неограниченной емкости ($M/M/1$)

1. Описание системы (рис. 2.13).

1.1. Система – *одноканальная* – с одним обслуживающим прибором.

1.2. Поток заявок – *однородный.*

1.3. В приборе происходит задержка поступающих в систему заявок на некоторое *случайное* время.

1.4. В системе имеется накопитель *неограниченной* емкости: $r = \infty$, т.е. любая заявка, поступившая в систему, найдет место для ожидания в очереди и не будет потеряна.

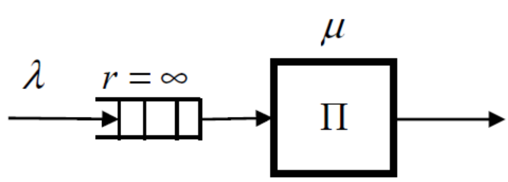


Рис. 2.13. СМО с накопителем неограниченной емкости

2. Предположения и допущения.

2.1. Поступающие в систему заявки образуют *простейший* поток с интенсивностью λ.

2.2. Длительность обслуживания заявок в приборе распределена по *экспоненциальному* закону с интенсивностью $\mu = 1/b$, где b – средняя длительность обслуживания заявок в приборе.

2.3. Дисциплина буферизации отсутствует, поскольку накопитель имеет неограниченную емкость.

2.4. Дисциплина обслуживания – *в порядке поступления* по правилу «первым пришел – первым обслужен» (FIFO).

2.5. Нагрузка системы совпадает с загрузкой, причем выполняется условие: $y = \rho < 1$, т.е. система работает в установившемся режиме без перегрузок. При $y > 1$, в отличие от предыдущих моделей, в СМО устанавливается режим перегрузок.

3. Кодирование состояний марковского процесса.

В качестве параметра, описывающего состояние марковского процесса, как и в предыдущем примере, будем рассматривать количество заявок k, находящихся в СМО (в приборе и в накопителе). Поскольку в системе в произвольный момент времени может находиться любое сколь угодно большое число заявок, то количество состояний марковского процесса равно бесконечности:

E_0: $k = 0$ – в системе нет ни одной заявки;

E_1: $k = 1$ – в системе находится 1 заявка (на обслуживании в приборе);

E_2: $k = 2$ – в системе находятся 2 заявки (одна – на обслуживании в приборе и вторая ожидает в накопителе);

…

E_k: k – в системе находятся k заявок (одна – на обслуживании в приборе и $(k-1)$ – в накопителе).

…

4. Размеченный граф переходов случайного процесса (рис. 2.14).

В один и тот же момент времени может происходить только одно событие: поступление заявки в систему с интенсивностью λ или завершение обслуживания заявки с интенсивностью μ. Размеченный граф переходов содержит бесконечное число состояний.

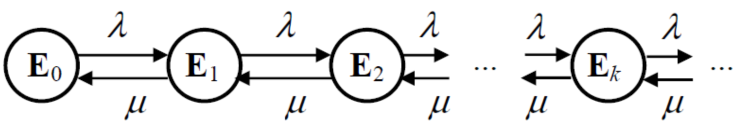


Рис. 2.14. Граф переходов марковского процесса

5. Система уравнений.

Не выписывая матрицу интенсивностей переходов, составим по графу переходов систему уравнений для определения стационарных вероятностей:

$$\begin{cases} \lambda p_0 = \mu p_1, \\ \ldots \\ (\lambda + \mu) p_1 = \lambda p_0 + \mu p_2; \\ (\lambda + \mu) p_2 = \lambda p_1 + \mu p_3; \\ \ldots \\ \left(\lambda + \mu\right) p_k = \lambda p_{k-1} + \mu p_{k+1}; \\ \ldots \\ p_0 + p_1 + p_2 + \ldots + p_k = 1. \end{cases}$$

Несмотря на то, что система содержит *бесконечное число уравнений* и, соответственно, *бесконечное число неизвестных*, нетрудно методом математической индукции получить аналитическое решение в явном виде для расчета вероятностей состояний одноканальной экспоненциальной СМО с однородным потоком заявок и накопителем неограниченной емкости при условии, что нагрузка системы $y < 1$:

$$p_k = y^k (1 - y) = \rho^k (1 - \rho),\ k = 1, 2, \ldots,$$

где $y = \lambda/\mu$ – загрузка системы, совпадающая с нагрузкой, причем $y < 1$, что гарантирует существование установившегося режима в системе.

Таким образом, вероятность нахождения марковского процесса в состоянии E_k или, что то же самое, вероятность того, что в произвольный момент времени в системе находится k заявок, распределена по геометрическому закону с параметром, равным загрузке (нагрузке) системы.

6. Расчет характеристик СМО.

Для расчета характеристик СМО можно воспользоваться следующими математическими зависимостями:

1) нагрузка: $y = \lambda/\mu = \lambda b$;

2) загрузка: $\rho = 1 - p_0 = \lambda b$ и совпадает с нагрузкой;

3) коэффициент простоя системы: $\eta = p_0 = 1 - \rho$;

4) среднее число заявок в очереди: $l = \sum_{k=1}^{\infty} (k-1) p_k = \dfrac{y^2}{1-y}$;

5) среднее число заявок в системе: $m = \sum_{k=0}^{\infty} k p_k = \dfrac{y}{1-y}$;

6) вероятность потери заявок: $\pi = 0$;

7) производительность системы при отсутствии потерь совпадает с интенсивностью поступления заявок в систему: $\lambda' = \lambda$;

8) интенсивность потерянных заявок: $\lambda'' = 0$;

9) среднее время ожидания заявок: $w = \dfrac{l}{\lambda} = \dfrac{yb}{1-y}$;

10) среднее время пребывания заявок: $u = w + b$ или $u = \dfrac{m}{\lambda} = \dfrac{b}{1-y}$;

11) относительная пропускная способность: $Q = 1 = P_{\text{обсл}}$.

7. Анализ свойств одноканальной СМО.

1. Среднее время ожидания заявок в очереди минимально при постоянной (детерминированной) длительности обслуживания заявок, когда коэффициент вариации длительности обслуживания $V_b = 0$, и увеличивается с ростом коэффициента вариации (дисперсии) длительности обслуживания. Заметим, что зависимость среднего времени ожидания от коэффициента вариации V_b носит нелинейный характер. Так, например, при экспоненциально распределенной длительности обслуживания, когда $V_b = 1$, среднее время ожидания заявок увеличивается в 2 раза, а при $V_b = 2$ – в 5 раз, по сравнению с детерминированным обслуживанием.

2. Среднее время ожидания заявок существенно зависит от нагрузки y (загрузки ρ) системы (рис. 2.15). При $y \geq 1$ ($\rho \to 1$) время ожидания заявок возрастает неограниченно: $w \to \infty$, т.е. заявки могут ожидать обслуживания сколь угодно долго. Отметим, что увеличение нагрузки может быть обусловлено двумя факторами: увеличением интенсивности поступления заявок в систему или увеличением длительности обслуживания заявок (например, в результате уменьшения скорости работы обслуживающего прибора).

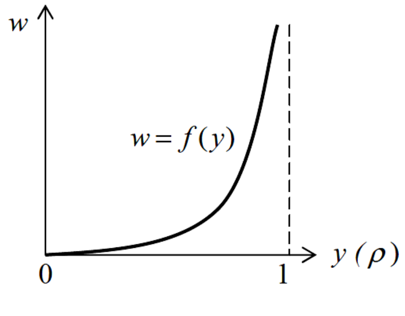


Рис. 2.15. Зависимость среднего времени ожидания от нагрузки

3. Можно показать, что для бесприоритетных дисциплин обслуживания в обратном порядке (ООП) и обслуживания в случайном порядке (ОСП) *средние времена ожидания заявок будут такими же*, как и при обслуживании в порядке поступления, но *дисперсии времени ожидания будут больше*. Это обусловлено, в частности для дисциплины ООП, тем, что часть заявок, поступивших последними, будут ожидать незначительное время, в то время как заявки, попавшие в начало очереди, могут ожидать обслуживания достаточно долго, т.е. увеличивается разброс значений времени ожидания относительно среднего значения.

Пример 15. В парикмахерской работает один мастер. Интенсивность потока клиентов составляет 4 клиента в 1 ч. Интенсивность обслуживания – 5 клиентов в 1 ч. Предполагается, что очередь может быть неограниченной длины. Определить показатели эффективности работы парикмахерской и вероятность того, что ожидают своей очереди не более двух клиентов.

Решение: $y = \lambda/\mu = 4/5 = 0{,}8 < 1$.

Следовательно, предельные вероятности существуют. Предельная вероятность того, что парикмахер простаивает, определяется соотношением $p_0 = 0{,}8^0\,(1 - 0{,}8) = 0{,}2$, а вероятность того, что он занят $P_{\text{зан}} = 1 - p_0 = 0{,}8$.

Вероятность того, что в очереди находятся не более трех клиентов: $P(k \le 3) = p_1 + p_2 + p_3$, где

$$p_1 = y^1(1 - y) = 0{,}8(1 - 0{,}8) = 0{,}16;$$
$$p_2 = y^2(1 - y) = 0{,}8^2(1 - 0{,}8) = 0{,}128;$$
$$p_3 = y^3(1 - y) = 0{,}8^3(1 - 0{,}8) = 0{,}1024.$$

Получаем: $P(k \le 3) = p_1 + p_2 + p_3 = 0{,}16 + 0{,}128 + 0{,}1024 =$
$= 0{,}3904$.

Среднее число заявок и среднее время пребывания в системе определяется по формулам:

$$m = \frac{y}{1-y} = \frac{0{,}8}{1-0{,}8} = 4,\;\; u = \frac{m}{\lambda} = \frac{4}{4} = 1 \text{ ч.}$$

Среднее число клиентов, ожидающих в очереди, и среднее время пребывания в очереди:

$$l = \frac{y^2}{1-y} = \frac{0{,}8^2}{1-0{,}8} = 3{,}2;\quad w = \frac{3{,}2}{4} = 0{,}8 \text{ ч.}$$

Задачи для самоконтроля

1. В одноканальную СМО поступают заявки с интенсивностью 0,85 заявок в 1 ч. Время обслуживания распределено по показательному закону и в среднем равно 1,05 ч. Очередь может расти практически неограниченно. Поток заявок простейший. Найти показатели эффективности работы СМО.

Ответ: $p_0 = 0{,}107$, $l = 7{,}453$, $w = 8{,}766$, $m = 8{,}346$, $u = 9{,}817$.

2. В порту имеется один причал для разгрузки судов. Интенсивность потока судов равна 0,4 судов в 1 сут. Среднее время разгрузки одного судна составляет 2 сут. Предполагается, что очередь может быть неограниченной длины. Найти показатели эффективности работы причала, а также вероятность того, что ожидают разгрузки не более чем 2 судна.

Ответ: $p_0 = 0{,}2$, $p_{1,2,3} = 0{,}3904$, $l = 3{,}2$, $w = 8$ сут, $m = 4$, $u = 10$ сут.

2.7. Многоканальная система массового обслуживания с накопителем ограниченной емкости

2.7.1. Двухканальная система массового обслуживания с накопителем ограниченной емкости (M/M/2/1)

Рассмотрим многоканальную СМО с отказами, в которую поступает однородный поток заявок. Все приборы системы являются идентичными, и поступившая в систему заявка занимает любой свободный прибор. Заявка, поступившая в систему и заставшая все приборы занятыми, заносится в накопитель ограниченной емкости, если он не заполнен до предела. В противном случае заявка получает отказ и покидает систему необслуженной. Таким образом, в системе, кроме входящего потока заявок с интенсивностью λ, образуются еще два потока заявок: поток обслуженных заявок с интенсивностью λ' и поток необслуженных заявок (получивших отказ в обслуживании) с интенсивностью λ''. Очевидно, что $\lambda' + \lambda'' = \lambda$.

1. Описание системы (рис. 2.16).

1.1. Система *двухканальная* – содержит два обслуживающих прибора.

1.2. В систему поступает *однородный* поток заявок.

1.3. Приборы – *идентичные*, т.е. время обслуживания заявок в приборах одинаково.

1.4. Перед прибором имеется накопитель *единичной емкости*: $r = 1$.

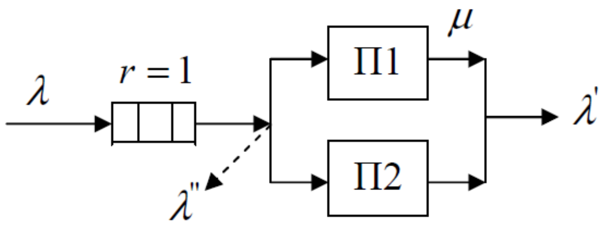


Рис. 2.16. Двухканальная СМО

2. Предположения и допущения.

2.1. Поступающие в систему заявки образуют *простейший* поток с интенсивностью λ.

2.2. Длительность обслуживания заявок в приборе распределена по *экспоненциальному* закону с интенсивностью $\mu = 1/b$, где b – средняя длительность обслуживания заявок в приборе.

2.3. Дисциплина буферизации – *с потерями*: заявка, поступившая в систему и заставшая накопитель заполненным, теряется.

2.4. Дисциплина обслуживания – *в естественном порядке*: заявка, поступившая в систему и заставшая прибор свободным, принимается на обслуживание.

3. Кодирование состояний случайного процесса.

В качестве параметра, описывающего состояние случайного процесса, будем рассматривать количество заявок k, находящихся в СМО.

Система в любой момент времени может находиться в одном из следующих состояний:

E_0: $k = 0$ – в системе нет заявок (оба прибора простаивают);

E_1: $k = 1$ – в системе (на обслуживании в одном из приборов) находится 1 заявка;

E_2: $k = 2$ – в системе (на обслуживании в обоих приборах) находятся две заявки;

E_3: $k = 3$ – в системе находятся три заявки: две – на обслуживании в приборах и одна – в накопителе.

4. Размеченный граф переходов случайного процесса (рис. 2.17).

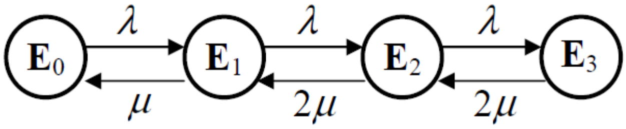


Рис. 2.17. Размеченный граф переходов

В один и тот же момент времени может произойти одно из двух событий, которые приводят к изменению состояния случайного процесса, протекающего в системе:

– поступление заявки в систему с интенсивностью λ, приводящее к переходу в соседнее состояние с бо́льшим номером, причем если случайный процесс находится в состоянии E_3, то его состояние не изменится, что соответствует отказу в обслуживании поступившей заявке;

– завершение обслуживания заявки в одном из приборов с интенсивностью μ, при этом случайный процесс переходит в соседнее состояние с меньшим номером с интенсивностью μ, если работает один прибор, и с интенсивностью 2μ, если работают оба прибора.

5. Расчет характеристик СМО.

Не выписывая матрицу интенсивностей переходов и систему уравнений для определения стационарных вероятностей, запишем формулы для расчета характеристик СМО:

1) нагрузка: $y = \lambda/\mu = \lambda b$;

2) загрузка: $\rho = (p_1 + 2p_2 + 2p_3)/2$;

3) среднее число работающих приборов: $\overline{k}_{\text{зан}} = (p_1+2p_2+2p_3) = 2\rho$;

4) коэффициент простоя системы: $\eta = 1 - \rho$;

5) среднее число заявок в очереди: $l = \dfrac{y^3}{2 \cdot 2!}$;

6) среднее число заявок в системе: $m = p_1 + 2p_2 + 3p_3 = l + \overline{k}_{\text{зан}}$;

7) вероятность потери заявок: $\pi = p_3$.

Задание на самостоятельную работу: доказать, что вероятность потери заявок равна вероятности того, что система заполнена, т.е. в накопителе нет свободных мест для вновь поступающих заявок.

8) производительность системы (интенсивность потока обслуженных заявок): $\lambda' = \lambda(1 - \pi)$;

9) интенсивность потока потерянных заявок: $\lambda'' = \lambda\pi$;

10) среднее время ожидания заявок: $w = l/\lambda'$;

11) среднее время пребывания заявок: $u = m/\lambda' = w + b$.

2.7.2. Многоканальная система массового обслуживания с ограниченной длиной очереди

Рассмотрим n-канальную систему массового обслуживания с ожиданием, в которую поступает простейший поток заявок с интенсивностью λ; интенсивность обслуживания μ (т.е. в среднем непрерывно занятый канал будет выдавать $y = \lambda/\mu$ обслуженных заявок в единицу времени). Длительность обслуживания – случайная величина, подчиненная показательному закону распределения. Поток обслуживания является простейшим пуассоновским потоком событий. Заявка, поступившая в момент, когда все n каналов заняты, становится в очередь и ожидает обслуживания. Предположим, что количество мест в очереди ограничено числом M, т.е. если заявка пришла в момент, когда в очереди уже стоят M заявок, она покидает систему необслуженной. В качестве показателей эффективности одноканальной СМО с ограниченной длиной очереди будем рассматривать:

λ' – абсолютную пропускную способность СМО (А);

Q – относительную пропускную способность;

π – вероятность отказа;

$P_{оч}$ – вероятность образования очереди;

$\bar{k}_{зан}$ – среднее число занятых каналов;

m – среднее число находящихся в системе заявок;

u – среднее время пребывания заявки в системе;

l – среднюю длину очереди;

w – среднее время ожидания в очереди.

Размеченный граф состояний представлен на рис. 2.18.

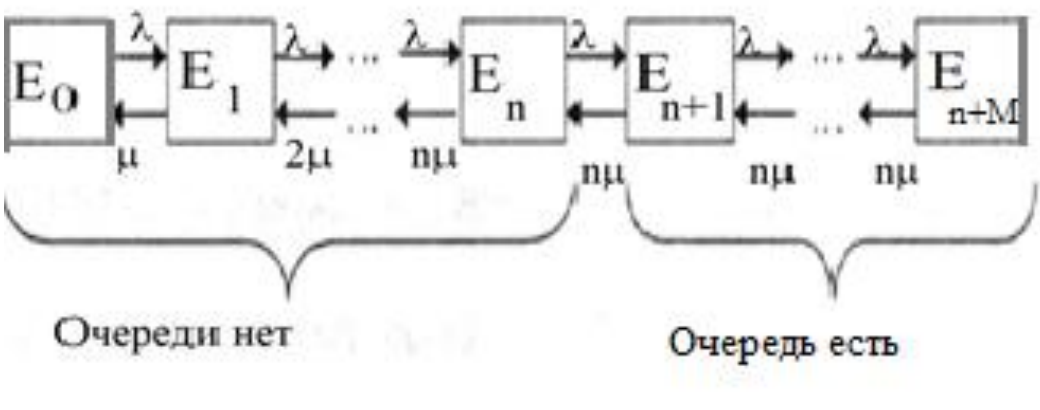


Рис. 2.18. Многоканальная СМО с ограниченной длиной очереди

Предельные вероятности состояний:

$$p_0 = \left(1 + \frac{y}{1!} + \frac{y^2}{2!} + \ldots + \frac{y^n}{n!} + \frac{y^{n+1}}{n \cdot n!} \cdot \frac{1 - \left(\frac{y}{n} \right)^M}{1 - y/n} \right)^{-1}; \qquad (2.23)$$

$$p_1 = y p_0, \quad p_2 = \frac{y^2}{2!} p_0, \ldots, \quad p_n = \frac{y^n}{n!} p_0; \qquad (2.24)$$

$$p_{n+1} = \frac{y^{n+1}}{n \cdot n!} p_0, \quad p_{n+2} = \frac{y^{n+2}}{n^2 \cdot n!} p_0, \ldots, \; p_{n+M} = \frac{y^{n+M}}{n^M \cdot n!} p_0; \quad (2.25)$$

$$\pi = P_{отк} = p_{n+M} = \frac{y^{n+M}}{n^M \cdot n!} p_0. \qquad (2.26)$$

Вероятность образования очереди:

$$P_{оч} = \sum_{i=0}^{M-1} p_{n+i} = \frac{y^n}{n!} \cdot \frac{1 - \left(\frac{y}{n} \right)^M}{1 - y/n} p_0. \qquad (2.27)$$

Относительная пропускная способность:

$$Q = 1 - \pi. \qquad (2.28)$$

Абсолютная пропускная способность:

$$\lambda' = \lambda Q. \qquad (2.29)$$

Среднее число занятых каналов: $\bar{k}_{зан} = \lambda' / \mu$

Среднее число заявок, находящихся в очереди:

$$l = \frac{y^{n+1}}{n \cdot n!} \cdot \frac{1 - \left(\frac{y}{n} \right)^M \left(M + 1 - \frac{M}{n} y \right)}{\left(1 - \frac{y}{n} \right)^2} p_0. \qquad (2.30)$$

Среднее время ожидания в очереди:

$$w = l/\lambda. \qquad (2.31)$$

Среднее число заявок в системе:

$$m = l + \bar{k}_{зан}. \qquad (2.32)$$

Среднее время пребывания заявки в СМО: $u = m/\lambda$.

Пример 16. На некоторую базу в среднем через каждые 30 мин прибывают автомашины с продукцией. Среднее время разгрузки одной машины составляет 1,5 ч. Разгрузку производят две бригады грузчиков. На территории базы могут находиться в очереди в ожидании разгрузки не более 4 автомашин. Определить показатели работы СМО.

Решение. СМО двухканальная, $n = 2$. Число мест в очереди $M = 4$. Интенсивность входящего потока $\lambda = 2$ машины в 1 ч. Интенсивность обслуживания $\mu = 1/b = 2/3$ машины в 1 ч. Приведенная интенсивность потока заявок: $y = \lambda/\mu = 3$, $y/n = 1{,}5$.

Вероятность того, что все бригады не загружены: $p_0 = 0{,}0158$.

Вероятность отказа:

$$\pi = P_{\text{отк}} = p_{n+M} = \frac{y^{n+M}}{n^M \cdot n!} p_0 = \frac{3^6}{2^4 \cdot 2!} \cdot 0{,}0158 = 0{,}36.$$

Относительная пропускная способность: $Q = 1 - \pi = 1 - 0{,}36 = 0{,}64$.

Абсолютная пропускная способность: $\lambda' = 2 \cdot 0{,}64 = 1{,}28$.

Среднее число занятых бригад: $\bar{k}_{\text{зан}} = \dfrac{\lambda'}{\mu} = 3 \cdot 0{,}64 = 1{,}92$.

Среднее число заявок, находящихся в очереди:

$$l = \frac{3^3}{2 \cdot 2!} \cdot \frac{1 - \left(\frac{3}{2}\right)^4 \left(4 + 1 - \frac{4}{2} \cdot 3\right)}{\left(1 - \frac{3}{2}\right)^2} \cdot 0{,}0158 = 2{,}6 \text{ машин.}$$

Среднее время ожидания в очереди: $w = 2{,}6/2 = 1{,}3$ ч.

Среднее время пребывания заявки в СМО: $u = 4{,}52/2 = 2{,}26$ ч.

Вопросы для самоконтроля

1. Чему равно число состояний n-канальной СМО с числом мест в очереди, равным M?

2. Нарисуйте размеченный граф состояний для n-канальной СМО с числом мест в очереди, равным M.

3. С вероятностью какого состояния совпадает вероятность отказа?

4. Сформулируйте условие существования финальных вероятностей для n-канальной СМО с числом мест в очереди, равным M.

Задачи для самоконтроля

1. В двухканальную СМО поступают заявки с интенсивностью 2 заявки в 1 ч. Поток обслуживания имеет интенсивность 4 заявки в 1 ч. Потоки поступления заявок и обслуживания – простейшие. Ожидать обслуживания в системе могут не более двух заявок. Определить показатели работы СМО.

Ответ: $p_0 = 0{,}6009$, $\pi = 0{,}0047$, $Q = 0{,}9953$, $\lambda' = 1{,}9906$, $\bar{k}_{зан} =$ $= 0{,}4977$, $l = 0{,}0282$, $w = 0{,}0141$ ч, $m = 0{,}5258$, $u = 0{,}2629$ ч.

2. В парикмахерской работают 3 мастера, а в зале ожидания расположены 3 стула. Поток клиентов имеет интенсивность $\lambda = 12$ клиентов в 1 ч. Среднее время обслуживания $b = 20$ мин. Определить относительную и абсолютную пропускную способность системы, среднее число занятых кресел, среднюю длину очереди, среднее время, которое клиент проводит в парикмахерской.

Ответ: $p_0 = 0{,}012$, $\pi = 0{,}307$, $Q = 0{,}693$, $\lambda' = 8{,}32$, $\bar{k}_{зан} = 2{,}78$, $l = 1{,}76$, $w = 0{,}13$ ч, $m = 4{,}34$, $u = 0{,}36$ ч.

2.8. Многоканальная система массового обслуживания с неограниченной очередью

Рассмотрим n-канальную систему массового обслуживания с неограниченной очередью, в которую поступает простейший поток заявок с интенсивностью λ; интенсивность обслуживания μ (т.е. в среднем непрерывно занятый канал будет выдавать $y = \lambda/\mu$ – обслуженных заявок в единицу времени).

Длительность обслуживания – случайная величина, подчиненная показательному закону распределения.

Поток обслуживания является простейшим пуассоновским потоком событий.

Заявка, поступившая в момент, когда все n каналов заняты, становится в очередь и ожидает обслуживания.

В качестве показателей эффективности одноканальной СМО с ограниченной длиной очереди будем рассматривать:

λ' – абсолютную пропускную способность СМО;

Q – относительную пропускную способность;

π – вероятность отказа;

$P_{оч}$ – вероятность образования очереди;

m – среднее число находящихся в системе заявок;

u – среднее время пребывания заявки в системе;
l – среднюю длину очереди;
w – среднее время ожидания в очереди;
$\bar{k}_{зан}$ – среднее число занятых каналов.

Размеченный граф состояний представлен на рис. 2.19.

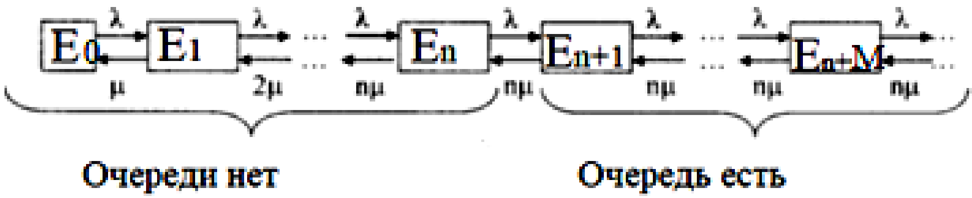


Рис. 2.19. Многоканальная СМО с неограниченной длиной очереди

На рис. 2.19:
E_0 – все каналы свободны, $k = 0$;
E_1 – занят один канал, остальные свободны, $k = 1$;
E_n – заняты все n каналов, очереди нет, $k = n$;
E_{n+1} – заняты все n каналов, одна заявка в очереди, $k = n + 1$; …;
E_{n+M} – заняты все n каналов, M заявок в очереди, $k = n + M$.

Поскольку ограничение на длину очереди отсутствует, то любая заявка может быть обслужена, поэтому $P_{обс} = 1$, следовательно, относительная пропускная способность $Q = P_{обс} = 1$, следовательно, $\pi = 0$, а абсолютная пропускная способность $\lambda' = \lambda Q = \lambda$.

Предельные вероятности:

$$p_0 = \left(1 + \frac{y}{1!} + \frac{y^2}{2!} + \ldots + \frac{y^{n-1}}{(n-1)!} + \frac{y^n}{n!} + \frac{y^{n+1}}{n!} \cdot \frac{1}{n-y}\right)^{-1}; \qquad (2.33)$$

$$p_1 = yp_0, \quad p_2 = \frac{y^2}{2!}p_0, \ldots, \quad p_n = \frac{y^n}{n!}p_0; \qquad (2.34)$$

$$p_{n+1} = \frac{y^{n+1}}{n \cdot n!}p_0, \quad p_{n+2} = \frac{y^{n+2}}{n^2 \cdot n!}p_0, \ldots, \quad p_{n+M} = \frac{y^{n+M}}{n^M \cdot n!}p_0.$$

Вероятность образования очереди:

$$P_{оч} = \frac{y^{n+1}}{n!(n-y)} \cdot p_0. \qquad (2.35$$

Среднее число занятых каналов:

$$\bar{k}_{зан} = \frac{A}{\mu}. \qquad (2.36)$$

Средняя длина очереди:

$$l = P_{\text{оч}} \cdot \frac{n}{n-y}. \tag{2.37}$$

Среднее время ожидания в очереди: $w = l/\lambda$.

Среднее число заявок в системе: $m = l + y$.

Среднее время пребывания заявки в СМО: $u = m/\lambda$.

Если $y < n$, то процесс обслуживания устойчив. Если $y > n$, то СМО работает неустойчиво.

Пример 17. В магазине работают 3 продавца. Покупатели магазина образуют простейший поток требований с интенсивностью 60 человек в 1 ч. Интенсивность обслуживания одного покупателя составляет 30 человек в 1 ч. Найти характеристики обслуживания.

Решение: $y = \lambda/\mu = 60/30 = 2$. По условию $y < n$, следовательно, очередь не будет возрастать до бесконечности и в системе наступит предельный стационарный режим работы.

Найдем вероятность того, что у касс отсутствуют покупатели:

$$p_0 = \left(1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{3!} \cdot \frac{1}{3-2}\right)^{-1} = 0{,}111.$$

Вероятность того, что у касс обслуживаются один, два, три покупателя, находим по формулам:

$$p_1 = 2 \cdot 0{,}111 = 0{,}222;$$
$$p_2 = (2^2/2!) \cdot 0{,}111 = 0{,}222;$$
$$p_3 = (2^3/3!) \cdot 0{,}111 = 0{,}148.$$

Вероятность того, что у касс стоят в очередь один, два покупателя, находим по формулам:

$$p_4 = 2^4/(3 \cdot 3!) \cdot 0{,}111 = 0{,}099.$$
$$p_5 = 2^5/(3^2\ 3!) \cdot 0{,}111 = 0{,}066.$$

Вероятность того, что заявка окажется в очереди, определяется по формуле

$$P_{\text{оч}} = 2^4/((3-2) \cdot 3!) \cdot 0{,}111 = 0{,}296.$$

Среднее число занятых касс: $\bar{k}_{\text{зан}} = \dfrac{A}{\mu} = \dfrac{60}{30} = 2$ кассира.

Среднее число покупателей в очереди:

$$l = 0{,}296 \cdot \frac{3}{3-2} = 0{,}296 \cdot 3 = 0{,}888.$$

Среднее число покупателей, обслуживаемых кассирами и стоящих в очереди: $m = 0{,}888 + 2 = 2{,}888$.

Среднее время пребывания заявки в очереди: $w = \dfrac{0{,}888}{60} = 0{,}0148 \dfrac{\text{человек}}{\text{ч}}$.

Среднее время пребывания заявки в системе: $u = \dfrac{2{,}888}{60} = 0{,}0481 \dfrac{\text{человек}}{\text{ч}}$.

Вопросы для самоконтроля

1. Чему равно число состояний n-канальной СМО с неограниченным ожиданием?

2. Нарисуйте размеченный граф состояний для n-канальной СМО с неограниченным ожиданием.

3. Сформулируйте условие существования финальных вероятностей для n-канальной СМО с неограниченным ожиданием.

4. Чему равны абсолютная и относительная пропускные способности n-канальной СМО с неограниченным ожиданием?

5. С какими характеристиками эффективности n-канальной СМО с ожиданием совпадает среднее число занятых каналов данной системы?

6. Как связаны между собой временные характеристики *«среднее время обслуживания одной заявки, относящееся ко всем заявкам»* и *«среднее время обслуживания одной заявки, относящееся только к обслуженным заявкам»* для n-канальной СМО с неограниченным ожиданием?

Задачи для самоконтроля

1. В многоканальную СМО с двумя каналами обслуживания поступают заявки с интенсивностью 0,8 заявок в 1 ч (поток заявок простейший). Поток обслуживания имеет интенсивность 0,5 заявки в 1 ч. Очередь заявок на обслуживание может расти практически неограниченно. Определить все средние характеристики системы.

Ответ: p_0 = 0,111; $P_{\text{оч}}$ = 0,431; l = 0,71; m = 2,31; w = 0,89; u = 2,89; $\bar{k}_{\text{зан}} = 1{,}6$.

2. В морском порту три причала, интенсивность входного потока – 2,5 судов в 1 сут. Интенсивность погрузочно-разгрузочных

работ – 2 судна в 1 день. Поток заявок и поток обслуживания пуассоновские. Очередь судов может расти практически неограниченно. Имея в виду стационарный режим работы, определить все средние характеристики системы.

Ответ: $p_0 = 0{,}279$; $p_1 = 0{,}349$; $p_2 = 0{,}218$; $p_3 = 0{,}091$; $p_4 = 0{,}028$; $P_{\text{оч}} = 0{,}063$; $l = 0{,}111$ судна в 1 сут; $m = 1{,}361$ судна в 1 сут; $w = 0{,}044$ сут; $u = 0{,}544$ сут; $\bar{k}_{\text{зан}} = 1{,}25$.

Задачи для самостоятельного решения

1. Автозаправочная станция имеет 4 бензоколонки. Среднее время заправки 2 мин. Входящий поток автомашин – простейший с интенсивностью 1,5 машин/мин. При всех занятых колонках требование теряется. Определить вероятность отказа и среднее число занятых колонок.

2. Покупатели магазина образуют простейший поток требований с интенсивностью 150 человек/ч. Определить наименьшее число продавцов, при которых среднее число покупателей, ожидающих обслуживания, не превысит 3.

3. В нефтеналивном порту 4 причала для заправки танкеров, которые приходят в среднем через 18 ч, а время загрузки составляет в среднем 2 сут. В очереди могут стоять не более 2 танкеров. Определить пропускную способность и холостой ход порта.

4. Определить закон распределения промежутка времени между приходом двух требований в простейшем потоке требований интенсивностью λ.

5. Поток желающих оформить вызов врача на дом – простейший. В среднем абоненты звонят через каждые 10 с. Время приема вызова распределено по показательному закону со средним значением 12 с. Определить наименьшее число телефонов в регистратуре, при котором вызов принимается не менее чем от 90 % абонентов. Считается, что в случае неудачи абонент не предпринимает больше попыток дозвониться.

6. Доказать, что выходящий из показательного канала (на входе которого всегда имеются заявки) поток является простейшим.

7. Автоматическая мойка может принять на обслуживание одновременно 4 автомашины. В среднем машины прибывают через 2 мин, а средняя продолжительность мойки – 10 мин. В очереди могут находиться не более 6 машин. Определить вероятность

того, что в системе находится хотя бы одна машина, и загруженность одной установки для мойки машин.

8. В магазине имеется 3 справочных телефона. В среднем обращаются за справками 40 человек/ч. Средняя продолжительность справочного разговора 3 мин. Издержки, связанные с работой одного телефона, – *а* руб./мин. Определить минимальную стоимость 1 мин разговора по телефону, при которой система неубыточна.

9. Платная стоянка для легковых машин имеет 7 мест. Найти вероятность того, что прибывшая машина найдет свободное место, если машины в среднем прибывают через 10 мин, а занимают место на стоянке в среднем 1 ч.

10. Поток деталей, сходящих с конвейера, простейший с интенсивностью 2 детали/мин. Время проверки детали контролером имеет показательный закон распределения со средним 2 детали/мин. Определить долю непроверенных деталей.

11. Город обслуживают 4 машины скорой помощи. Вызовы поступают в среднем через 4 ч. Вероятность того, что хотя бы одна машина занята, равна 0,25. Определить среднее число занятых машин и среднюю долю простоя машин.

12. В парикмахерской работают два мастера. Время обслуживания распределено по показательному закону со средним 12 мин. Ожидать обслуживание могут не более трех человек. Поток клиентов – простейший с интенсивностью 10 клиентов/ч. Найти важнейшие операционные характеристики этой системы.

13. Система автоматической посадки самолетов одновременно может хранить данные только о шести самолетах, находящихся в воздухе. Самолеты, подлетающие к аэродрому, образуют простейший поток с интенсивностью 6 самолетов/ч. Если в момент запроса посадки система заполнена, то самолет улетает к запасному аэродрому. Аэродром имеет 3 посадочные полосы, самолет занимает полосу в среднем 20 мин. Найти пропускную способность СМО, загруженность одной полосы, среднее число занятых полос, среднее время ожидания начала посадки после запроса.

14. Рассматривается работа автозаправочной станции, на которой имеется 2 заправочные колонки. Предположим, что она описывается процессом размножения и гибели в стационарном режиме. Заправка каждой машины длится в среднем 3 мин. В среднем на АЗС каждые 2 мин прибывает машина, нуждающаяся в заправке. Число мест в очереди не ограничено. Все машины, ожидающие заправку, терпеливо дожидаются своей очереди. Определить:

1) вероятность того, что на заправке находится 5 машин;

2) вероятность того, что вновь прибывшей машине придется ждать обслуживания.

15. Закусочная на АЗС имеет один прилавок. Автомобили прибывают в соответствии с пуассоновским распределением, в среднем 2 автомобиля за 5 мин. Для выполнения заказа в среднем достаточно 1,5 мин, хотя продолжительность обслуживания распределена по экспоненциальному закону.

Найти:

а) вероятность простоя прилавка;

б) средние характеристики;

в) вероятность того, что количество прибывших автомобилей будет не менее 10.

16. Рентгеновский аппарат позволяет обследовать в среднем 7 человек в 1 ч. Интенсивность посетителей составляет 5 человек в 1 ч. Предполагая стационарный режим работы, определить средние характеристики.

17. В речном порту один причал, интенсивность входного потока – 5 судов в 1 день. Интенсивность погрузочно-разгрузочных работ – 6 судов в 1 день. Имея в виду стационарный режим работы, определить все средние характеристики системы.

18. Какое оптимальное число каналов обслуживания должна иметь СМО, если интенсивность потока заявок равна 3, среднее число заявок, обслуженных в единицу времени, равно 2, штраф за каждый отказ равен 5, а стоимость простоя одной линии равна 2?

19. Какое оптимальное число каналов обслуживания должна иметь СМО, если интенсивность потока заявок равна 3, среднее число заявок, обслуженных в единицу времени, равно 1, штраф за каждый отказ равен 7, а стоимость простоя одной линии равна 3?

20. Какое оптимальное число каналов обслуживания должна иметь СМО, если интенсивность потока заявок равна 4, среднее число заявок, обслуженных в единицу времени, равно 2, штраф за каждый отказ равен 5, а стоимость простоя одной линии равна 1?

21. Определить число взлетно-посадочных полос для самолетов с учетом требования, что вероятность ожидания должна быть меньше, чем 0,05. При этом интенсивность входного потока 27 самолетов в 1 сут, а интенсивность их обслуживания – 30 самолетов в 1 сут.

22. Сколько равноценных независимых конвейерных линий должен иметь цех, чтобы обеспечить ритм работы, при котором вероятность ожидания обработки изделий должна быть меньше

0,03 (каждое изделие выпускается одной линией). Известно, что интенсивность поступления заказов – 30 изделий в 1 ч, а интенсивность обработки изделия одной линией – 36 изделий в 1 ч.

23. Среднее число вызовов, поступающих на АТС за 1 мин, равно 3. Считая поток пуассоновским, найдите вероятность того, что за 2 мин поступит:

а) два вызова;

б) меньше двух вызовов;

в) не менее двух вызовов.

24. Сколько каналов должна иметь СМО с отказами, если интенсивность потока заявок равна 2 треб./ч, среднее число заявок, обслуженных в единицу времени, равно 1 треб./ч, штраф за каждый отказ составляет 8 тыс. руб., стоимость простоя одной линии – 2 тыс. руб. в 1 ч?

25. Система массового обслуживания представляет собой автоматическую телефонную станцию, которая может обеспечить не более пяти переговоров одновременно. Заявка-вызов, поступившая в тот момент, когда все каналы заняты, получает отказ и покидает систему. В среднем на станцию поступает 0,8 вызовов в 1 мин, а средняя продолжительность одних переговоров равна 1,5 мин. Для стационарного режима функционирования системы необходимо определить:

а) вероятности состояний системы;

б) вероятность отказа;

в) абсолютную и относительную пропускные способности;

г) среднее число занятых каналов.

26. Автозаправочная станция имеет одну бензоколонку с площадкой, допускающей пребывание в очереди на заправку не более трех автомашин одновременно. Если в очереди на заправку уже находятся три автомашины, то очередная автомашина, прибывшая на станцию, проезжает мимо. В среднем на заправку прибывает одна машина в 1 мин, а сам процесс заправки в среднем длится 1,25 мин. Для стационарного режима функционирования автозаправочной станции необходимо определить:

а) вероятность отказа;

б) относительную и абсолютную пропускные способности;

в) среднее число автомашин в очереди на заправку;

г) среднее время ожидания в очереди.

27. Рабочий обслуживает три однотипных станка. Каждый станок останавливается в среднем 2 раза в 1 ч, а процедура налад-

ки занимает в среднем 10 мин. В стационарном режиме функционирования системы нужно определить:

а) вероятности состояний системы;

б) вероятность занятости рабочего;

в) среднее количество неисправных станков;

г) среднее число налаживаемых станков.

28. Рассматривается система массового обслуживания $M|M|1$ с простейшим входящим потоком с параметром $\lambda_k = (m - k)\lambda$ ($k = 0,1,...,m$) (ограниченный входящий поток). Поведение системы описывается марковским процессом, $\xi(t)$ – число заявок в системе в момент времени t. Время обслуживания заявки имеет показательное распределение с параметром $\mu_k = \mu$ $\{k =1, 2, \ldots, m)$. Требуется:

а) построить граф переходов процесса;

б) найти предполагаемые стационарные вероятности;

в) доказать, что стационарное распределение имеет найденный вид.

29. Рассматривается система массового обслуживания $M|M|n$ с простейшим входящим потоком с параметром $\lambda_k = (m - k)\lambda$ ($k = 0,1,...,m$) (ограниченный входящий поток). Поведение системы описывается марковским процессом, $\xi(t)$ – число заявок в системе в момент времени t. Время обслуживания заявки имеет показательное распределение с параметром

$$\mu_k = \begin{cases} k\mu, \ 1 \le k < n, \\ n\mu, \ n \le k < m. \end{cases}$$

Требуется:

а) построить граф переходов процесса;

б) найти предполагаемые стационарные вероятности;

в) доказать, что стационарное распределение имеет найденный вид.

30. Специализированный пост диагностики представляет собой одноканальную СМО. Число стоянок для автомобилей, ожидающих проведения диагностики, ограниченно и равно 3. Если все стоянки заняты, то очередной автомобиль, пришедший на диагностику, в очередь на обслуживание не становится. Поток прибывающих автомобилей – пуассоновский и имеет интенсивность 0,85 автомобиля в 1 ч. Время диагностики распределено по показательному закону и в среднем составляет 1,05 ч. Требуется провести сравнительный анализ работы СМО при $E = 3$ и $E = 4$.

2.9. Одноканальная система массового обслуживания с неоднородным потоком заявок и относительными приоритетами

Рассмотрим одноканальную СМО, в которую поступает неоднородный поток заявок. Ожидающие обслуживания заявки разнесены по разным накопителям ограниченной емкости. Между заявками разных классов установлены относительные приоритеты (ОП), означающие, что всякий раз из накопителей на обслуживание выбирается заявка с самым высоким приоритетом. При этом при поступлении в систему высокоприоритетной заявки обслуживание низкоприоритетной не прерывается. При заполненных накопителях поступившая заявка теряется.

1. Описание системы (рис. 2.20).

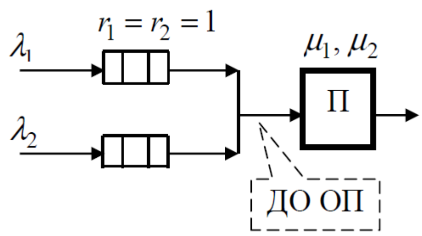


Рис. 2.20. СМО с неоднородным потоком заявок

1.1. Система *одноканальная*.

1.2. Входящий поток заявок – *неоднородный*: в систему поступает два класса заявок.

1.3. Накопители для заявок каждого класса – *ограниченной емкости*: $r_1 = r_2 = 1$.

1.4. Дисциплина буферизации – *без вытеснения* заявок: если при поступлении в систему заявки любого класса соответствующий накопитель заполнен до конца, то заявка теряется.

1.5. Дисциплина обслуживания – *с относительными приоритетами*: заявки первого класса имеют приоритет по отношению к заявкам второго класса.

2. Предположения и допущения.

2.1. Поступающие в систему заявки двух классов образуют *простейшие* потоки с интенсивностями λ_1 и λ_2 соответственно.

2.2. Длительности обслуживания заявок каждого класса распределены по *экспоненциальному* закону с интенсивностями $\mu_1 = 1/b_1$ и $\mu_2 = 1/b_2$, где b_1 и b_2 – средние длительности обслуживания заявок класса 1 и 2 соответственно.

В СМО всегда существует стационарный режим, так как не может быть бесконечных очередей.

3. Кодирование состояний случайного процесса.

Для описания состояний марковского процесса будем использовать распределение заявок между прибором и накопителями. Закодируем состояния следующим образом: (П/$О_1$, $О_2$), где П = {0, 1, 2} – состояние обслуживающего прибора, задаваемое классом заявки, находящейся на обслуживании («0» – прибор свободен; «1» или «2» – на обслуживании в приборе находится заявка класса 1 или 2 соответственно); $О_1$, $О_2$ = {0, 1} – состояние накопителей 1 и 2 соответственно («0» означает отсутствие заявки в накопителе, «1» означает наличие одной заявки в накопителе соответствующего класса).

При выбранном способе кодирования система может находиться в следующих состояниях:

E_0: (0/0,0) – в системе нет ни одной заявки;

E_1: (1/0,0) – на обслуживании в приборе находится заявка класса 1;

E_2: (2/0,0) – на обслуживании в приборе находится заявка класса 2;

E_3: (1/1,0) – на обслуживании находится заявка класса 1 и одна заявка класса 1 ожидает обслуживания в первом накопителе;

E_4: (1/0,1) – на обслуживании находится заявка класса 1 и одна заявка класса 2 ожидает обслуживания соответственно во втором накопителе;

E_5: (2/1,0) – на обслуживании находится заявка класса 2 и одна заявка класса 1 ожидает обслуживания в первом накопителе;

E_6: (2/0,1) – на обслуживании находится заявка класса 2 и одна заявка класса 2 ожидает обслуживания во втором накопителе;

E_7: (1/1,1) – на обслуживании находится заявка класса 1 и по одной заявке каждого класса ожидают обслуживания в соответствующих накопителях;

E_8: (2/1,1) – на обслуживании находится заявка класса 2 и по одной заявке каждого класса ожидают обслуживания в соответствующих накопителях.

Отметим, что при кодировании случайных процессов могут быть применены различные способы кодирования.

В рассматриваемом примере состояния случайного процесса вместо представленного выше способа можно закодировать, например, следующим образом: (П,О), где П = {0, 1, 2} – состояние обслуживающего прибора, задаваемое классом заявки, находящейся на обслуживании («0» прибор свободен; «1» или «2» – на обслуживании в приборе находится заявка класса 1 или 2 соответственно); O = {0, 1, 2, 3} – состояние накопителей 1 и 2 соответственно («0» – означает отсутствие заявок в обоих накопителях; «1» – наличие в первом накопителе заявки класса 1; «2» – наличие во втором накопителе заявки класса 2; «3» – наличие в первом и втором накопителях по одной заявке соответственно класса 1 и 2).

Представленные способы кодирования не применяются, если для заявок обоих классов используется общий накопитель емкостью $r = 2$. В этом случае количество состояний случайного процесса увеличится, поскольку в накопителе могут находиться 2 заявки одного и того же класса, и состояние накопителя может быть представлено следующим образом: O = {0, 1, 2, 11, 12, 22}, где «0» – означает отсутствие заявок в накопителе; «1» – наличие в накопителе только одной заявки класса 1; «2» – наличие в накопителе заявки класса 2; «11» – наличие в накопителе двух заявок класса 1; «22» – наличие в накопителе двух заявок класса 2 и «12» – наличие в накопителе одной заявки класса 1 и одной заявки класса 2. Заметим, что состояние «12» не различает, в какой последовательности эти заявки поступили в систему, что обусловлено наличием относительного приоритета между ними – независимо от момента поступления на обслуживание первой всегда будет выбрана заявка класса 1. В случае бесприоритетного обслуживания, когда заявки разных классов выбираются на обслуживание в порядке поступления, следует ввести еще одно состояние накопителя – «21», означающее, что заявка класса 2 поступила в систему раньше заявки класса 1, в то время как состояние «12» означает, что в систему раньше поступила заявка класса 1.

4. Размеченный граф переходов случайного процесса представлен на рис. 2.21.

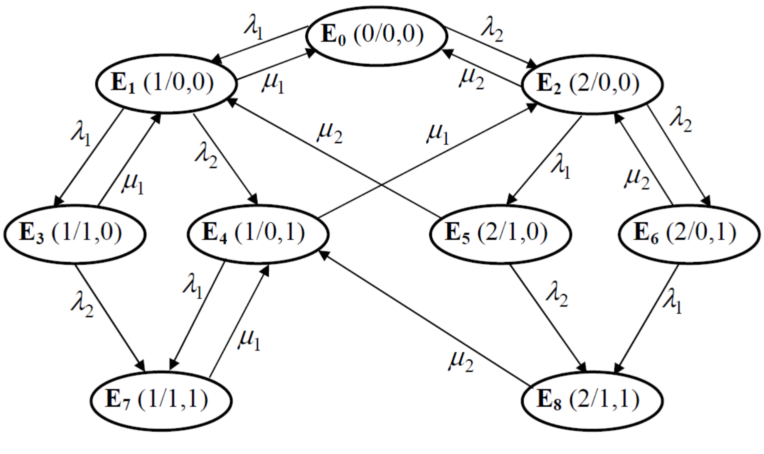


Рис. 2.21. Размеченный граф переходов

В каждый момент времени может произойти только одно событие (или поступление заявки какого-либо класса, или завершение обслуживания заявки, находящейся в приборе), поскольку вероятность появления двух и более событий в один и тот же момент времени равна нулю.

При наличии в накопителях заявок 1 и 2 классов (состояния E_7 и E_8) после завершения обслуживания некоторой заявки в приборе случайный процесс переходит в состояние $\mathbf{E}_4$, означающее, что на обслуживание всегда выбирается высокоприоритетная заявка класса 1.

По графу переходов составим систему уравнений для определения стационарных вероятностей:

$$
\begin{aligned}
&(\lambda_1+\lambda_2)p_0=\mu_1 p_1+\mu_2 p_2;\\
&(\lambda_1+\lambda_2+\mu_1)p_1=\lambda_1 p_0+\mu_1 p_3+\mu_2 p_5;\\
&(\lambda_1+\lambda_2+\mu_2)p_2=\lambda_2 p_0+\mu_1 p_4+\mu_2 p_6;\\
&(\lambda_2+\mu_1)p_3=\lambda_1 p_1;\\
&(\lambda_1+\mu_1)p_4=\lambda_2 p_1+\mu_1 p_7+\mu_2 p_8;\\
&(\lambda_2+\mu_2)p_5=\lambda_1 p_2;\\
&(\lambda_1+\mu_2)p_6=\lambda_2 p_2;\\
&\mu_1 p_7=\lambda_2 p_8+\lambda_1 p_4;\\
&\mu_2 p_8=\lambda_2 p_5+\lambda_1 p_6;\\
&\sum_{i=}^{8} p_i=1.
\end{aligned}
$$

5. Расчет характеристик СМО.

Характеристики обслуживания заявок в СМО с неоднородным потоком заявок делятся на две группы:

– характеристики обслуживания заявок каждого класса;

– характеристики обслуживания заявок суммарного потока.

Расчет характеристик обслуживания заявок *каждого класса* выполняется по следующим формулам:

1) нагрузка: $y_1=\lambda_1/\mu_1=\lambda_1 b_1$; $y_2=\lambda_2/\mu_2=\lambda_2 b_2$;

2) загрузка, создаваемая заявками, которая может трактоваться как вероятность того, что на обслуживании в приборе находится заявка класса 1 и 2 соответственно:

$$\rho_1=p_1+p_3+p_4+p_7;$$

$$\rho_2=p_2+p_5+p_6+p_8;$$

3) среднее число заявок в очереди: $l_1=p_3+p_5+p_7+p_8$; $l_2=$ $=p_4+p_6+p_7+p_8$;

4) среднее число заявок в системе:

$$m_1 = p_1 + 2\,p_3 + p_4 + p_5 + 2\,p_7 + p_8 = l_1 + r_1;$$

$$m_2 = p_2 + p_4 + p_5 + 2p_6 + p_7 + 2p_8 = l_2 + r_2;$$

5) вероятность потери заявок (вероятность отказа $P_{\text{отк}}$):

$$\pi_1 = p_3 + p_5 + p_7 + p_8;\ \pi_2 = p_4 + p_6 + p_7 + p_8;$$

6) производительность по каждому классу заявок (интенсивность непотерянных заявок):

$$\lambda'_1 = \lambda_1(1 - \pi_1),\ \lambda'_2 = \lambda_2(1 - \pi_2);$$

7) среднее время ожидания заявок: $w_1 = l_1 / \lambda'_1$, $w_2 = l_2 / \lambda'_2$;

8) среднее время пребывания заявок:

$$u_1 = m_1/\lambda'_1 = w_1 + b_1;$$

$$u_2 = m_2/\lambda'_2 = w_2 + b_2.$$

Расчет характеристик обслуживания заявок *суммарного потока* выполняется по следующим формулам:

1) суммарная нагрузка системы: $Y = y_1 + y_2$;

2) загрузка системы: $R = \rho_1 + \rho_2$;

3) коэффициент простоя системы: $\eta = 1 - R = t_{\text{пр}}$;

4) суммарное число заявок во всех очередях: $L_{\text{оч}} = l = l_1 + l_2$;

5) суммарное число заявок в системе:

$$m = m_1 + m_2 = l + R = L_{\text{сист}};$$

6) вероятность потери заявок: $\pi = \pi_1 + \pi_2 = P_{\text{отк}}$;

7) производительность системы (интенсивность суммарного потока обслуженных заявок, абсолютная пропускная способность): $A = \lambda Q = \lambda'_1 + \lambda'_2$;

8) среднее время ожидания заявок суммарного потока:

$$w = T_{\text{оч}} = \frac{\lambda'_1 w_1 + \lambda'_2 w_2}{\lambda'} = \frac{L_{\text{оч}}}{\lambda'};$$

9) среднее время пребывания заявок суммарного потока:

$$u = (\lambda'_1 u_1 + \lambda'_2 u_2)/\lambda' = m/\lambda' = w + b.$$

Приоритетные системы массового обслуживания

Рассмотрим однолинейную СМО с ожиданием. На вход системы поступают r независимых простейших потоков; k-й поток имеет интенсивность $\lambda_k,\ k = 1,\ldots,r$. Будем обозначать: $\lambda_k = \sum_{i=1}^{k} \lambda_i$.

Времена обслуживания запросов из k-го потока характеризуются функцией распределения $B_{\text{k}}(t)$ с преобразованием Лапласа – Стилтьеса $\beta_k(s)$ и конечными начальными моментами $b_k^{(m)} = \int_0^\infty t^m dB_{\text{k}}(t), \quad m = 1, 2,\ k = 1, \ldots, r.$ Запросы из k-го потока назовем запросами приоритета k.

Считаем, что запросы из i-го потока более приоритетны, чем запросы из j-го потока, если $i < j$. Приоритетность проявляется в том, что в момент окончания обслуживания следующим на обслуживание выбирается из очереди запрос, имеющий максимальный приоритет. Запросы, имеющие один и тот же приоритет, выбираются согласно установленной дисциплине обслуживания, например, согласно дисциплине FIFO.

Рассматриваются различные варианты поведения системы в ситуации, когда во время обслуживания запроса некоторого приоритета в систему поступает запрос более высокого приоритета.

Система называется СМО с относительным приоритетом, если поступление такого запроса не прерывает обслуживание запроса. Если же такое прерывание происходит, то система называется СМО с **абсолютным приоритетом**. В этом случае, однако, требуется уточнить дальнейшее поведение запроса, обслуживание которого оказалось прерванным. Различают следующие варианты: прерванный запрос уходит из системы и теряется; прерванный запрос возвращается в очередь и продолжает обслуживание с места прерывания после ухода из системы всех запросов, имеющих более высокий приоритет; прерванный запрос возвращается в очередь и начинает обслуживание заново после ухода из системы всех запросов, имеющих более высокий приоритет. Прерванный запрос обслуживается прибором после ухода из системы всех запросов, имеющих более высокий приоритет, в течение времени, имеющего прежнее или некоторое другое распределение. Возможен вариант, когда требуемое время обслуживания в последующих попытках идентично времени, которое требовалось для полного обслуживания данного запроса в первой попытке.

Таким образом, имеется достаточно большое число вариантов поведения системы с приоритетом. Общим в анализе всех систем с приоритетами является использование понятия периода занятости системы запросами приоритета k и выше. При этом основным методом исследования этих систем является метод введения дополнительного события.

Проиллюстрируем особенности нахождения характеристик систем с приоритетами. Будем считать, что это система с относительным приоритетом, и найдем стационарное распределение времени $w_k(t)$ ожидания запроса приоритета k, $k=1,\ldots,r$, если бы он поступил в систему в момент времени t (так называемого виртуального времени ожидания), для системы с относительными приоритетами.

Обозначим

$$W_k(x)=\lim_{t\to\infty} P\{w_k(t)<x\},\quad x>0.$$

Условием существования этих пределов является выполнение неравенства

$$y<1,$$

где величина y вычисляется по формуле

$$y=\sum_{k=1}^{r}\lambda_k b_1^{(k)}.$$

Обозначим также $w_k(s)=\int_0^{\infty} e^{-sx}dW_k(x).$

Утверждение. Время $w_k(s)$ ожидания запроса приоритета k стационарного распределения виртуального времени ожидания запроса приоритета k определяется следующим образом:

$$w_k(s)=\frac{(1-y)\mu_k(s)+\sum_{j=k+1}^{r}\lambda_j(1-\beta_j(\mu_k(s)))}{s-\lambda_k+\lambda_k\beta_k(\mu_k(s))},$$

где функции $\mu_k(s)$ задаются формулой

$$\mu_k(s)=s+\Lambda_{k-1}-\Lambda_{k-1}\pi_{k-1}(s), \qquad (2.38)$$

а функции $\pi_l(s)$, $l=1,\ldots,r$, находятся как решения функциональных уравнений:

$$\Lambda_l\pi_l(s)=\sum_{i=1}^{l}\lambda_i\beta_i(s+\Lambda_l-\Lambda_l\pi_l(s)),\ l=1,\ldots,r. \qquad (2.39)$$

Обозначим $W_1^{(k)}$ как математическое ожидание виртуального времени ожидания в системе запроса приоритета k.

Следствие. Величины $W_1^{(k)}$ высчитываются следующим образом:

$$W_1^{(k)} = \frac{\sum_{i=1}^{r} \lambda_i b_2^{(i)}}{2\left(1-\sum_{i=1}^{k} \lambda_i b_1^{(i)}\right)\left(1-\sum_{i=1}^{k-1} \lambda_i b_1^{(i)}\right)}.$$

Отметим, что из систем с абсолютным приоритетом наиболее легко исследуется система с двумя потоками запросов и дообслуживанием прерванных запросов. В этой системе характеристики процесса обслуживания приоритетного потока совершенно не зависят от наличия второго потока и вычисляются по обычным формулам для системы $M|G|1$ при интенсивности входящего потока, равной λ_1, и распределении $B_1(x)$ времени обслуживания запроса. В свою очередь характеристики процесса обслуживания неприоритетного потока вычисляются по формулам для ненадежной системы $M|G|1$ при интенсивности входящего потока, равной λ_2, и распределении $B_2(x)$ времени обслуживания запроса. Приход приоритетного запроса здесь трактуется как поломка прибора, а время ремонта прибора распределено как период занятости системы, обслуживающей приоритетные запросы.

В системах с абсолютным приоритетом без ограничения на длину очереди заявка, обладающая приоритетом, немедленно принимается к обслуживанию каналом, занятым обслуживанием заявки без приоритета в обслуживании. После того как требование, обладающее приоритетом, обслужено и в очереди нет других приоритетных заявок, возобновляется прерванное обслуживание заявки, не обладающей приоритетом.

При этом может быть несколько ситуаций:

- заявка, обслуживание которой было прервано, начинает обслуживаться заново;
- обслуживание заявки начинается с того места, где оно было прервано;
- заявка, обслуживание которой было прервано, покидает систему (теряется).

В системах с относительным приоритетом заявка, не обладающая приоритетом, обслуживается до конца, и только после завершения ее обслуживания начинает обслуживаться приоритетная заявка. Например, для электроэнергетики наиболее характерной ситуацией является одноканальная СМО с абсолютным приоритетом. Так выглядит устранение аварийных ситуаций в распределительных электрических сетях, когда район электрических сетей поделен на эксплуатационные участки, и каждый из них имеет бригаду из 4–5 человек для устранения возникающих неисправностей. При этом в случае отключения районной больницы, птицефабрики, узла связи восстановление электроснабжения для них осуществляется в первую очередь.

Рассмотрим одноканальную СМО с абсолютным приоритетом без ограничения на длину очереди. На вход системы поступает два независимых простейших потока заявок с интенсивностями: λ_1 – для заявок с приоритетом, λ_2 – без приоритета. Среднее время обслуживания заявок обоих видов соответственно $1/\mu_1$ и $1/\mu_2$. Время обслуживания распределено по показательному закону. Основные параметры рассматриваемой системы (в стационарном режиме) определяются по следующим формулам:

– среднее число приоритетных заявок, находящихся в очереди:

$$\bar{k} = \frac{y_1^2}{1-y_1}; \tag{2.40}$$

– а среднее время пребывания в очереди заявки, обладающей приоритетом:

$$\bar{t}_{\text{оч.1}} = \frac{\bar{k}_1}{\lambda_1} = \frac{y_1}{\mu_1(1-y_1)}; \tag{2.41}$$

– среднее время пребывания в системе заявки, обладающей приоритетом:

$$\bar{t}_{\text{сист.1}} = \frac{1}{\mu_1(1-y_1)} = \bar{t}_{\text{оч.1}} + \bar{t}_{\text{обсл.1}}; \tag{2.42}$$

– среднее число неприоритетных заявок, находящихся в системе:

$$\bar{r}_2 = \frac{y_2}{1-y}\left(1+\frac{\mu_2}{\mu_1}\cdot\frac{y_1}{1-y_1}\right),\quad y = y_1 + y_2; \tag{2.43}$$

– среднее время ожидания в очереди заявки, не обладающей приоритетом:

$$\bar{t}_{\text{оч.2}} = \frac{1}{\mu_2}\cdot\frac{\dfrac{\mu_2}{\mu_1}\cdot\dfrac{y_1}{1-y}+y}{1-y}; \tag{2.44}$$

– среднее время пребывания в системе заявки, не обладающей приоритетом:

$$\bar{t}_{\text{сист.2}} = \bar{t}_{\text{оч.2}} + \frac{1}{\mu_2}. \tag{2.45}$$

Пример 17. По надежности электроснабжения потребители городской электрической сети поделены условно на две категории. Поток отказов фидеров линии электропередачи, обеспечивающих электроснабжение потребителей I категории $\lambda_1 = 0{,}1$ в 1 ч, а для потребителей II категории $\lambda_2 = 0{,}2$ в 1 ч. Среднее время устранения неисправностей в электрических сетях соответствующих категорий составляет $\bar{t}_{\text{обсл.1}} = 1$, $\bar{t}_{\text{обсл.2}} = 2{,}5$. Устранением неисправностей занимается одна бригада дежурных электромонтеров. В случае возникновения неисправности в фидере, обеспечивающем электрической энергией потребителей I категории, обслуживание потребителей II категории прекращается. Определить среднее время ожидания ремонта и нахождения в отключенном состоянии для потребителей I и II категорий.

Решение:

1. По условию задачи рассматривается одноканальная СМО с абсолютным приоритетом без ограничения на длину очереди со следующими параметрами: $\lambda_1 = 0{,}1$ в 1 ч; $\lambda_2 = 0{,}2$ в 1 ч; $\mu_1 = 1$ ч; $\mu_2 = 0{,}4$ ч. Тогда

$$y_1 = \frac{\lambda_1}{\mu_1} = \frac{0{,}1}{1} = 0{,}1; \quad y_2 = \frac{\lambda_2}{\mu_2} = \frac{0{,}2}{0{,}4} = 0{,}5; \quad y = y_1 + y_2 = 0{,}1 + 0{,}5 = 0{,}6.$$

2. Среднее время пребывания в очереди заявок, обладающих приоритетом, составит

$$\bar{t}_{\text{оч.1}} = \frac{y_1}{\mu_1(1 - y_1)} = \frac{0{,}1}{1 \cdot (1 - 0{,}1)} = 0{,}11 \text{ ч.}$$

3. Среднее время пребывания в отключенном состоянии потребителей I категории составит

$$\bar{t}_{\text{сист.1}} = \bar{t}_{\text{оч.1}} + \bar{t}_{\text{обсл.1}} = 0{,}11 + 1 = 1{,}11 \text{ ч.}$$

4. Среднее время ожидания в очереди заявки, не обладающей приоритетом, составит

$$\bar{t}_{\text{оч.2}} = \frac{1}{\mu_2} \cdot \frac{\frac{\mu_2}{\mu_1} \cdot \frac{y_1}{1 - y} + y}{1 - y} = \frac{1}{0{,}4} \cdot \frac{\frac{0{,}4}{1} \cdot \frac{0{,}1}{1 - 0{,}6} + 0{,}6}{1 - 0{,}6} = 1{,}75 \text{ ч.}$$

5. Среднее время пребывания в системе заявки, не обладающей приоритетом, составит

$$\bar{t}_{\text{сист.2}} = \bar{t}_{\text{оч.2}} + \frac{1}{\mu_2} = 1{,}75 + \frac{1}{0{,}4} = 4{,}25 \text{ ч.}$$

2.10. Система массового обслуживания с ограниченным временем ожидания

До сих пор мы рассматривали СМО с ожиданием, ограниченным только длиной очереди (числом m заявок, одновременно находящихся в очереди). В такой СМО заявка, раз ставшая в очередь, уже не покидает ее и «терпеливо» дожидается обслуживания. На практике нередко встречаются СМО другого типа, в которых заявка, подождав некоторое время, может уйти из очереди (так называемые «нетерпеливые» заявки).

Рассмотрим СМО подобного типа, оставаясь в рамках марковской схемы. Предположим, что имеется n-канальная СМО с ожиданием, в которой число мест в очереди не ограничено, но время пребывания заявки в очереди ограничено некоторым случайным сроком $T_{\text{оч}}$ со средним значением $\bar{t}_{\text{оч}}$. Таким образом, на каждую заявку, стоящую в очереди, действует как бы «поток уходов» с интенсивностью

$$\nu = 1/\bar{t}_{\text{оч}}.$$

Если этот поток пуассоновский, то процесс, протекающий в СМО, будет марковским. Найдем для него вероятности состояний. Будем снова нумеровать состояния системы по числу заявок, связанных с системой – как обслуживаемых, так и стоящих в очереди:

$$\left\{\begin{array}{l} E_0 - \text{все каналы свободны,} \\ E_1 - \text{занят один канал,} \\ E_2 - \text{заняты два канала,} \qquad - \text{очереди нет,} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ E_n - \text{заняты все } n \text{ каналов,} \end{array}\right.$$

E_{n+1} – заняты все n каналов, одна заявка стоит в очереди,

……………………………………………………………………………

E_{n+r} – заняты все n каналов, r заявок стоит в очереди,

……………………………………………………………………………

и т.д.

Граф состояний системы показан на рис. 2.22.

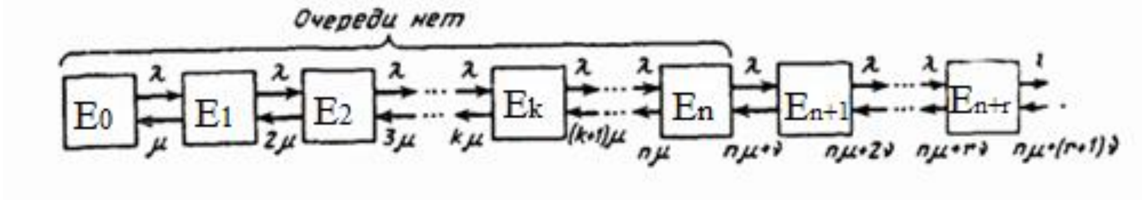


Рис. 2.22. Граф состояний системы

Разметим этот граф, т.е. проставим у стрелок соответствующие интенсивности. Снова, как и раньше, у всех стрелок, ведущих слева направо, будет стоять интенсивность потока заявок λ. Для состояний без очереди у стрелок, ведущих из них справа налево, будет, как и раньше, стоять суммарная интенсивность потока обслуживаний всех занятых каналов. Что касается состояний с очередью, то у стрелок, ведущих из них справа налево, будет стоять суммарная интенсивность потока обслуживаний всех n каналов $n\mu$ плюс соответствующая интенсивность потока уходов из очереди. Если в очереди стоят r заявок, то суммарная интенсивность потока уходов будет равна $r\nu$. Как видно из графа, перед нами опять схема гибели и размножения; применяя общие выражения для предельных вероятностей состояний в этой схеме, напишем:

$$p_1 = \frac{\lambda/\mu}{1!} p_0,$$

$$p_2 = \frac{\left(\frac{\lambda}{\mu}\right)^2}{2!} p_0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$p_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} p_0,$$

$$p_{n+1} = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!(n\mu + \nu)} p_0,$$

$$p_{n+2} = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} \cdot \frac{\lambda^2}{(n\mu+\nu)(n\mu+2\nu)} p_0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$p_{n+r} = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} \cdot \frac{\lambda^r}{(n\mu + v)(n\mu + 2v)\ldots(n\mu + rv)} p_0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$p_0 = \left\{ \begin{array}{l} 1 + \frac{\lambda/\mu}{1!} + \frac{\left(\frac{\lambda}{\mu}\right)^2}{2!} + \ldots + \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} + \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} \cdot \frac{\lambda}{n\mu + v} + \\ \left[+ \frac{\lambda^2}{(n\mu + v)(n\mu + 2v)} + \ldots + \frac{\lambda^r}{(n\mu + v)(n\mu + 2v)\ldots(n\mu + rv)} \right] \end{array} \right\}^{-1},$$

или, вводя обозначения, $y = \frac{\lambda}{\mu}, \beta = v/\mu$:

$$p_0 = \left\{ \begin{array}{l} 1 + \frac{y}{1!} + \frac{y^2}{2!} + \ldots + \frac{y^n}{n!} + \frac{y^n}{n!} \cdot \\ \left[\frac{y}{n+\beta} + \frac{y^2}{(n+\beta)(n+2\beta)} + \ldots + \frac{y^r}{(n+\beta)(n+2\beta)\ldots(n+r\beta)} \right] \end{array} \right\}^{-1}, \quad (2.46)$$

$$p_1 = \frac{y}{1!} p_0,$$

$$p_2 = \frac{y}{2!} p_0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$p_n = \frac{y^n}{n!} p_0,.$$

$$p_{n+1} = \frac{y^n}{n!} \frac{y}{(n+\beta)} p_0,$$

$$p_{n+2} = \frac{y^n}{n!} \cdot \frac{y^2}{(n+\beta)(n+2\beta)} p_0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$p_{n+r} = \frac{y^n}{n!} \cdot \frac{y^r}{(n+\beta)(n+2\beta)\ldots(n+r\beta)} p_0.$$

Отметим некоторые особенности рассмотренной СМО с «нетерпеливыми» заявками по сравнению с ранее рассмотренной СМО с «терпеливыми» заявками.

Если длина очереди не ограничена заранее никаким числом и заявки «терпеливы» (не уходят из очереди), то стационарный предельный режим существует только в случае $y<n$ (при $y \geq n$ соответствующая бесконечная геометрическая прогрессия расходится, что физически соответствует неограниченному росту очереди при $t \to \infty$). Напротив, в СМО с «нетерпеливыми» заявками, уходящими рано или поздно из очереди, установившийся режим обслуживания при $t \to \infty$ достигается всегда, независимо от приведенной интенсивности потока заявок y. Это следует из того, что ряд в знаменателе первой формулы (2.46) сходится при любых положительных значениях y и β. Для СМО с «нетерпеливыми» заявками понятие «вероятность отказа» не имеет смысла – каждая заявка становится в очередь, но может и не дождаться обслуживания, уйдя раньше времени.

Относительную пропускную способность Q такой СМО можно подсчитать следующим образом. Очевидно, обслужены будут все заявки, кроме тех, которые уйдут из очереди досрочно. Подсчитаем, какое в среднем число заявок покидает очередь досрочно. Для этого вычислим среднее число заявок в очереди:

$$\bar{r} = 1 \cdot p_{n+1} + 2 \cdot p_{n+2} + 3 \cdot p_{n+3} + \ldots + r \cdot p_{n+r} + \ldots \tag{2.47}$$

На каждую из этих заявок действует «поток уходов» с интенсивностью v. Значит, из среднего числа $\bar{r}$ заявок в очереди в среднем будет уходить, не дождавшись обслуживания, $v\bar{r}$ заявок в единицу времени; всего в единицу времени в среднем будет обслужено

$$A = \lambda - v\bar{r} \tag{2.48}$$

заявок.

Относительная пропускная способность (Q) СМО будет равна

$$Q = \frac{A}{\lambda} = \frac{\lambda - v\bar{r}}{\lambda} = 1 - \frac{v}{\lambda}\bar{r}. \tag{2.49}$$

Среднее число занятых каналов $(\bar{k}_{\text{зан}})$ по-прежнему получим, деля абсолютную пропускную способность на μ:

$$\bar{k}_{\text{зан}} = \frac{A}{\mu} = \frac{\lambda - v\bar{r}}{\mu} = y - \beta\bar{r}. \tag{2.50}$$

Это позволяет вычислить среднее число заявок в очереди $\bar{r}$, не суммируя бесконечного ряда (2.47). Действительно, из (2.50) получим:

$$\bar{r} = \frac{y}{\beta} - \frac{\bar{k}_{\text{зан}}}{\beta}, \tag{2.51}$$

а входящее в эту формулу среднее число занятых каналов можно найти как математическое ожидание случайной величины K принимающей значения $0,1,2,\ldots,n$ вероятностями $p_0, p_1, p_2, \ldots,$ $(1-(p_0+p_1+\ldots+p_{n-1}))$.

$$\begin{aligned}\bar{k}_{\text{зан}} &= 0\cdot p_0 + 1\cdot p_1 + 2\cdot p_2 + \ldots + n(1-(p_0+p_1+\ldots+p_{n-1})) = \\ &p_1 + 2p_2 + \ldots + n(1-(p_0+p_1+\ldots+p_{n-1})).\end{aligned} \tag{2.52}$$

Мы не будем выводить формул для среднего времени ожидания в очереди, так как для этого требуются сравнительно сложные выкладки.

Заметим, что в формуле (2.46) фигурирует сумма бесконечного ряда, не являющегося прогрессией. Однако эта сумма вычисляется приближенно, причем достаточно легко, так как члены ряда быстро убывают с увеличением их номера. В качестве приближенного значения для бесконечной суммы берется сумма конечного числа $(r-1)$ членов, а остаток оценивается следующим образом:

$$\begin{aligned}&\frac{y^n}{n!}\left[\frac{y^r}{(n+\beta)(n+2\beta)\ldots(n+r\beta)} + \frac{y^{r+1}}{(n+\beta)(n+2\beta)\ldots(n+(r+1)\beta)} + \ldots\right] < \\ &< \frac{y^n}{n!}\left[\frac{y^r}{(\beta)(2\beta)\ldots(r\beta)} + \frac{y^{r+1}}{(\beta)(2\beta)\ldots(r+1)\beta} + \ldots\right] = \\ &= \frac{y^n}{n!}\left[\frac{(y/\beta)^r}{r!} + \frac{(y/\beta)^{r+1}}{(r+1)!} + \ldots\right].\end{aligned} \tag{2.53}$$

Можно доказать, что бесконечная сумма в квадратных скобках меньше, чем $\dfrac{\left(\dfrac{y}{\beta}\right)^r}{r!}$, и выражение (2.53) меньше, чем

$$\frac{y^n}{n!}\cdot\frac{\left(\dfrac{y}{\beta}\right)^r}{r!}e^{\frac{y}{\beta}}.$$

В заключение заметим, что если в формулах (2.46) перейти к пределу при $\nu \to 0$ (или, что то же самое, при $\beta \to 0$), то при $y < n$ «нетерпеливые» заявки станут «терпеливыми».

2.11. Замкнутые системы массового обслуживания

До сих пор мы рассматривали такие системы массового обслуживания, где заявки приходили откуда-то извне и интенсивность потока заявок не зависела от состояния самой системы. В настоящем параграфе мы рассмотрим системы массового обслуживания другого типа – такие, в которых интенсивность потока поступающих заявок зависит от состояния самой СМО. Такие системы массового обслуживания называются замкнутыми.

В качестве примера замкнутой СМО рассмотрим следующую систему. Рабочий-наладчик обслуживает n станков. Каждый станок может в любой момент выйти из строя и потребовать обслуживания со стороны наладчика. Интенсивность потока неисправностей каждого станка равна λ. Вышедший из строя станок останавливается. Если в этот момент рабочий свободен, он берется за наладку станка; на это он тратит среднее время

$$\bar{t}_{\text{об}} = \frac{1}{\mu},$$

где μ – интенсивность потока обслуживаний (наладок).

Если в момент выхода станка из строя рабочий занят, станок становится в очередь на обслуживание и ждет, пока рабочий не освободится.

Требуется найти вероятности состояний данной системы и ее характеристики:

– вероятность того, что рабочий не будет занят,

– вероятность наличия очереди,

– среднее число станков, ожидающих очереди на ремонт и т.д.

Перед нами – своеобразная система массового обслуживания, где источниками заявок являются станки, имеющиеся в ограниченном количестве и подающие или не подающие заявки в зависимости от своего состояния: при выходе станка из строя он перестает быть источником новых заявок. Следовательно, интенсивность общего потока заявок, с которым приходится иметь дело рабочему, зависит от того, сколько имеется неисправных станков, т.е. сколько заявок связано с процессом обслуживания (непосредственно обслуживается или стоит в очереди).

Характерным для замкнутой системы массового обслуживания является наличие ограниченного числа источников заявок.

В сущности, любая СМО имеет дело только с ограниченным числом источников заявок, но в ряде случаев число этих источников так велико, что можно пренебречь влиянием состояния самой СМО на поток заявок. Например, поток вызовов на АТС крупного города исходит, в сущности, от ограниченного числа абонентов, но это число так велико, что практически можно считать интенсивность потока заявок независимой от состояний самой АТС (сколько каналов занято в данный момент). В замкнутой же системе массового обслуживания источники заявок, наряду с каналами обслуживания, рассматриваются как элементы СМО.

Рассмотрим сформулированную выше задачу о рабочем-наладчике в рамках общей схемы марковских процессов.

Система, включающая рабочего и n станков, имеет ряд состояний, которые мы будем нумеровать по числу неисправных станков (станков, связанных с обслуживанием):

E_0 – все станки исправны (рабочий свободен);

E_1 – один станок неисправен, рабочий занят его наладкой;

E_2 – два станка неисправны, один налаживается, другой ожидает очереди, …;

E_n – все n станков неисправны, один налаживается, $(n-1)$ станков стоят в очереди.

Граф состояний приведен на рис. 2.23. Интенсивности потоков событий, переводящих систему из состояния в состояние, проставлены у стрелок. Из состояния E_0 в E_1 систему переводит поток неисправностей всех работающих станков; его интенсивность равна $n\lambda$. Из состояния E_1 в E_2 систему переводит поток неисправностей уже не n, а $(n-1)$ станков (работают всего $(n-1)$) и т.д. Что касается интенсивностей потоков событий, переводящих систему по стрелкам справа налево, то они все одинаковы – работает все время один рабочий с интенсивностью обслуживания μ.

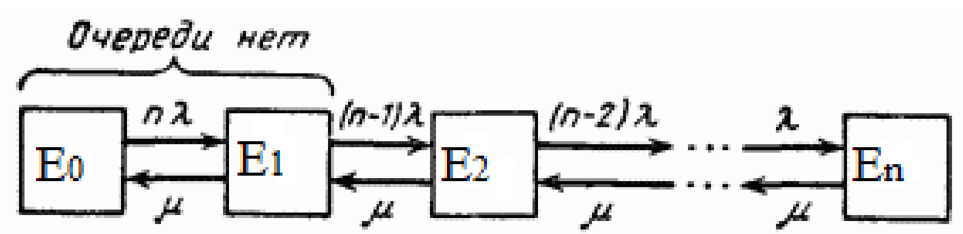


Рис. 2.23. Замкнутые системы массового обслуживания

Пользуясь, как обычно, общим решением задачи о предельных вероятностях состояний для схемы гибели и размножения, напишем предельные вероятности состояний:

$$\begin{cases} p_1 = nyp_0, \\ p_2 = n(n-1)y^2 p_0, \\ \dots\dots\dots\dots\dots\dots \\ p_n = n(n-1)(n-2)\dots 1\cdot y^n p_0, \\ p_0 = (1+ny+n(n-1)y^2+\dots+n(n-1)(n-2)\dots 1\cdot y^n)^{-1}. \end{cases} \tag{2.54}$$

Итак, вероятности состояний СМО найдены.

В силу своеобразия замкнутой СМО, характеристики ее эффективности будут отличны от тех, которые мы применяли ранее для СМО с неограниченным количеством источников заявок.

Роль «абсолютной пропускной способности» в данном случае будет играть среднее количество неисправностей, устраняемых рабочим в единицу времени. Вычислим эту характеристику. Рабочий занят наладкой станка с вероятностью

$$P_{\text{зан}} = 1 - p_0. \tag{2.55}$$

Если он занят, он обслуживает μ станков (ликвидирует μ неисправностей) в единицу времени; значит, абсолютная пропускная способность системы равна

$$A = (1 - p_0)\mu. \tag{2.56}$$

Относительную пропускную способность для замкнутой СМО мы не вычисляем, так как каждая заявка, в конце концов, будет обслужена: $Q = 1$. Вероятность того, что рабочий не будет занят, составляет

$$P_{\text{своб}} = 1 - P_{\text{зан}} = p_0. \tag{2.57}$$

Вычислим среднее число неисправных станков, иначе – среднее число станков, связанных с процессом обслуживания. Обозначим это среднее число $\overline{w}(\overline{k}_{\text{зан}})$. Вообще говоря, величину $\overline{w}$ можно вычислить непосредственно по формуле

$$\overline{k}_{\text{зан}} = 1\cdot p_1 + 2\cdot p_2 + \dots + n\cdot p_n,$$

но проще будет найти ее через абсолютную пропускную способность A.

Действительно, каждый работающий станок порождает поток неисправностей с интенсивностью λ; в нашей СМО в среднем работает n станков; порождаемый ими средний поток неисправностей будет иметь среднюю интенсивность $(n-\bar{k}_{\text{зан}})\lambda$; все эти неисправности устраняются рабочим, следовательно,

$$(n-\bar{k}_{\text{зан}})\lambda=(1-p_0)\mu,$$

откуда

$$\bar{k}_{\text{зан}}=n-\frac{\mu}{\lambda}(1-p_0)$$

и

$$\bar{k}_{\text{зан}}=n-\frac{(1-p_0)}{y}. \qquad (2.58)$$

Определим теперь среднее число станков $(n-\bar{r})$, ожидающих наладки в очереди. Будем рассуждать следующим образом: общее число станков W, связанных с обслуживанием, складывается из числа станков R, стоящих в очереди, плюс число станков Ω, непосредственно находящихся под обслуживанием:

$$W=R+\Omega.$$

Число станков Ω, находящихся под обслуживанием, равно единице, если рабочий занят, и нулю, если он свободен, т.е. среднее значение Ω равно вероятности того, что рабочий занят:

$$\bar{\varpi}=1-p_0.$$

Вычитая эту величину из среднего числа w станков, связанных с обслуживанием (неисправных), получим среднее число станков, ожидающих обслуживания в очереди:

$$\bar{r}=n-\frac{1-p_0}{y}-(1-p_0)=n-(1-p_0)\left(1+\frac{1}{y}\right). \qquad (2.59)$$

Остановимся еще на одной характеристике эффективности СМО: на производительности группы станков, обслуживаемых рабочим.

Зная среднее число неисправных станков $\bar{k}_{\text{зан}}$ и производительность l исправного станка за единицу времени, можно оценить среднюю потерю L производительности группы станков в единицу времени за счет неисправностей:

$$L=\bar{k}_{\text{зан}}l.$$

Пример 18. Рабочий обслуживает группу из трех станков. Каждый станок останавливается в среднем 2 раза в 1 ч. Процесс наладки занимает у рабочего в среднем 10 мин. Определить характеристики замкнутой СМО: вероятность занятости рабочего; его абсолютную пропускную способность А; среднее количество неисправных станков; среднюю относительную потерю производительности группы станков за счет неисправностей.

Решение. Имеем $n = 3$, $\lambda = 2$, $\mu = 1/(1/6) = 6$, $y = \lambda/\mu = 1/3$.

По формулам (2.54) получаем

$$p_0 = (1+3\cdot 1/3+3\cdot 2\cdot 1/3^2+\ldots+3\cdot 2\cdot 1\cdot 1/3^3)^{-1} \approx 0,346.$$

Вероятность занятости рабочего $P_{\text{зан}} = 1 - p_0 = 0,654$.

Абсолютная пропускная способность рабочего (среднее число неисправностей, которое он ликвидирует в 1 ч) $A = 0,654 \cdot 6 = 3,94$.

Среднее число неисправных станков находим по формуле (2.58): $\bar{k}_{\text{зан}} = 3 - \dfrac{0,654}{\frac{1}{3}} = 1,04$.

Средняя относительная потеря производительности группы станков за счет неисправностей $\dfrac{\bar{k}_{\text{зан}}}{n} = 0,347$, т.е. за счет неисправностей группа станков теряет около 35 % производительности.

Рассмотрим теперь более общий пример замкнутой СМО: бригада из m рабочих обслуживает n станков, $m < n$. Перечислим состояния системы:

$$\left.\begin{array}{l} E_0 - \text{все станки работают, рабочие не заняты,} \\ E_1 - \text{один станок остановился, один рабочий занят,} \\ E_2 - \text{два станка остановились, два рабочих заняты,} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ E_m - m \text{ станков остановились, все рабочие заняты,} \end{array}\right\} - \text{очереди нет,}$$

$E_{m+1} - (m+1)$ станок остановился, m налаживаются, один ждет очереди,

……………………………………………………………

$E_n -$ все n станков остановились, m налаживаются, $(n-m)$ ждет очереди.

Граф состояний системы показан на рис. 2.24 (интенсивности потоков событий проставлены у стрелок).

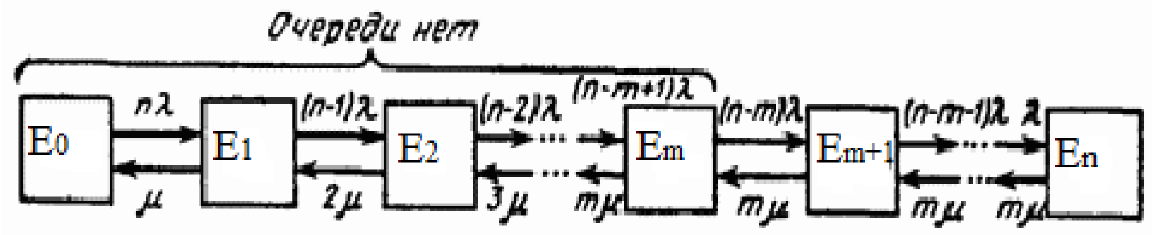


Рис. 2.24. Граф состояний системы

Применяя общее решение для схемы гибели и размножения, находим предельные вероятности состояний $\left(y = \dfrac{\lambda}{\mu}\right)$:

$$
\left\{
\begin{aligned}
&p_1 = n\frac{y}{1!}p_0,\\
&p_2 = n(n-1)\frac{y^2}{2!}p_0,\\
&p_3 = n(n-1)(n-2)\frac{y^3}{2!}p_0,\\
&\dots\dots\dots\dots\\
&p_m = n(n-1)(n-2)\dots(n-m+1)\frac{y^m}{m!}p_0,\\
&p_{m+1} = n(n-1)(n-2)\dots(n-m)\frac{y^{m+1}}{m!m}p_0,\\
&p_{m+2} = n(n-1)(n-2)\dots(n-m-1)\frac{y^{m+2}}{m!m^2}p_0,\\
&\dots\dots\dots\dots\\
&p_n = \frac{n!y^n}{m!m^{n-m}}p_0,\\
&p_0 = \begin{bmatrix} 1+\frac{n}{1}y+\frac{n(n-1)}{2!}y^2+\dots+\frac{n(n-1)(n-2)\dots(n-m+1)}{m!}y^m + \\ \frac{n(n-1)(n-2)\dots(n-m)}{m!m}y^{m+1}+\dots+\frac{n!}{m!m^{n-m}}y^n. \end{bmatrix}^{-1}.
\end{aligned}
\right.
\tag{2.60}
$$

Через эти вероятности выражается среднее число $\bar{k}_{\text{зан}}$ занятых рабочих:

$$\bar{k}_{\text{зан}} = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + \ldots + m \cdot (p_m + p_{m+1} + \ldots + p_n) =$$
$$= p_1 + 2p_2 + \ldots + (m-1)p_{m-1} + m(1 - p_0 - p_1 - \ldots - p_{m-1}).$$

Через $\bar{k}_{\text{зан}}$ выражается, в свою очередь, среднее число станков, обслуживаемых бригадой в единицу времени (абсолютная пропускная способность):

$$A = \bar{k}_{\text{зан}} \mu,$$

а также среднее число неисправных станков:

$$\bar{\omega} = n - \frac{\bar{k}_{\text{зан}} \mu}{\lambda} = n - \frac{\bar{k}_{\text{зан}}}{y}.$$

Отсюда же находится и средняя потеря производительности группы станков в единицу времени за счет неисправностей: нужно умножить среднее число неисправных станков ω на производительность l одного станка в единицу времени.

Пример 19. Два рабочих обслуживают группу из шести станков. Остановки каждого (работающего) станка случаются в среднем через каждые полчаса. Процесс наладки занимает у рабочего в среднем 10 мин. Определить характеристики замкнутой СМО:

– среднее число занятых рабочих,
– абсолютную пропускную способность,
– среднее количество неисправных станков.

Решение. Имеем:

$n = 6,\ m = 2,\ \lambda = 2,\ \ \mu = 1/(1/6) = 6,\ \ y = \lambda/\mu = 1/3.$

По формулам (2.60) получаем

$$p_0 = \left(1 + \frac{6}{1} \cdot \frac{1}{3} + \frac{6 \cdot 5}{1 \cdot 2} \cdot \frac{1}{3^2} + \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 2} \cdot \frac{1}{3^3} + \frac{6 \cdot 5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 2^2} \cdot \frac{1}{3^4} \right. +$$

$$\left. + \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{1 \cdot 2 \cdot 2^3} \cdot \frac{1}{3^5} + \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 2^3} \cdot \frac{1}{3^6} \right)^{-1} = \frac{1}{6{,}549} \approx 0{,}153.$$

$$p_1 = \frac{6}{1} \cdot \frac{1}{3} \cdot 0{,}153 \approx 0{,}306.$$

Отсюда среднее число занятых рабочих равно

$$\bar{k}_{\text{зан}} = 1 \cdot p_1 + 2 \cdot (1 - p_0 - p_1) = 1 \cdot 0{,}306 + 2 \cdot 0{,}541 = 1{,}388.$$

По формуле (2.60) находим абсолютную пропускную способность:

$$A = 1{,}388 \cdot 6 = 8{,}328.$$

По формуле

$$\bar{\omega} = n - \frac{\bar{k}_{\text{зан}}\mu}{\lambda} = n - \frac{\bar{k}_{\text{зан}}}{y}$$

находим среднее число неисправных станков:

$$\bar{\omega} = 6 - 1{,}388 \cdot 3 = 1{,}836.$$

Пример 20. Один ремонтный рабочий обслуживает 6 подъемных устройств на станции технического обслуживания автомобилей. Интенсивность поломок каждого устройства равна 0,1 в 1 сут. Среднее время, которое тратит рабочий на обслуживание одного подъемника, равно X ч. Найти предельные вероятности состояний данной СМО и найти среднее число неисправных подъемников, т.е. находящихся в ремонте и ожидающих ремонта. Определить также следующие характеристики:

1) вероятность того, что наладчик занят ремонтом подъемников;

2) абсолютную пропускную способность СМО;

3) относительную пропускную способность СМО;

4) среднее число подъемников в ремонте;

5) среднее число подъемников, ожидающих в очереди ремонта.

Пример 21. СМО состоит из n идентичных приборов, каждый из которых выходит из строя в случайные моменты времени с интенсивностью λ. В случае выхода прибора из строя он начинает сразу восстанавливаться одним из m свободных восстанавливающих устройств (ВУ) с интенсивностью μ. Если все ВУ заняты, то прибор встает в очередь и ждет до тех пор, пока не освободится ВУ. Каждое ВУ в любой момент времени может восстанавливать не более одного прибора. Требуется оценить надежность работы системы и дать предложения по повышению эффективности ее функционирования.

Характеристики функционирования замкнутой одноканальной СМО

Показатель (коэффициент) нагрузки системы, порождаемой каждым источником заявок	$y = \lambda / \mu$
Показатель (коэффициент) нагрузки системы, порождаемой всеми k источниками заявок	$ky = k\lambda / \mu$

Вероятность того, что канал свободен	$p_0 = \left(\sum_{i=0}^{k} \frac{k!}{\left(k-i\right)!} y^i \right)^{-1}$
Вероятность состояний СМО	$p_i = \frac{k!}{(k-i)!} y^i p_0, i = 1, \ldots, k$
Вероятность того, что канал занят	$P_{\text{зан}} = 1 - \text{p}_0$
Абсолютная пропускная способность СМО	$A = P_{\text{зан}} \mu = (1 - p_0)\mu$
Интенсивность выходящего потока обслуженных заявок	$V = A = P_{\text{зан}} \mu = (1 - p_0)\mu$
Относительная пропускная способность	$Q = 1$
Среднее число заявок в системе (т.е. среднее число источников, находящихся в пассивном состоянии)	$\bar{L}_{\text{сис}} = \bar{L}_{\text{пас}} = i - \frac{1 - \text{p}_0}{y} = \sum_{i=1}^{k} i \text{ p}_i$
Средняя интенсивность среднего суммарного входящего потока заявок	$\bar{\Lambda} = \left(k - \bar{L}_{\text{пас}}\right)\lambda$
Среднее число заявок, находящихся под обслуживанием	$\bar{L}_{\text{об}} = 1 - \text{p}_0$
Среднее число заявок, находящихся в очереди	$\bar{L}_{\text{оч}} = \bar{L}_{\text{сис}} = \bar{L}_{\text{об}} = = k - \left(1 - \text{p}_0\right)\left(1 + \frac{1}{y}\right)$
Коэффициент готовности – вероятность того, что произвольный источник находится в активном состоянии	$P_{\text{акт}} = 1 - \frac{1}{k} \bar{L}_{\text{сис}} = \frac{\bar{\Lambda}}{k\lambda}$
Вероятность того, что в момент поступления заявки СМО находилась в состоянии E_k	$\tilde{p}_i = \frac{(k-i) p_i}{k - L_{\text{сис}}}$ $i = 0,\ 1,\ \ldots, k-1$
Среднее время ожидания заявки в очереди	$\bar{T}_{\text{оч}} = \frac{1}{\mu(k - \bar{L}_{\text{сис}}} \sum_{i=1}^{k-1} i(k-i) p_i = = \frac{1}{\bar{\Lambda}} \bar{L}_{\text{оч}}$

Среднее время обслуживания одной заявки	$\bar{T}_{\text{об}} = \dfrac{1}{\mu} = \dfrac{1}{\bar{\Lambda}} \bar{L}_{\text{об}}$
Среднее время пребывания заявки в системе	$\bar{T}_{\text{сис}} = \bar{T}_{\text{оч}} + \bar{T}_{\text{об}} = \dfrac{\bar{L}_{\text{сис}}}{\bar{\Lambda}}$
Средняя производительность группы источников, находящихся в активном состоянии	$\bar{N}_{\text{акт}} \cdot k = (k - \bar{N}_{\text{пас}})k = \dfrac{1 - p_0}{y} k$
Средняя потеря производительности за счет группы источников, находящихся в пассивном состоянии	$\bar{N}_{\text{пас}} \cdot k = \left(k - \dfrac{1 - p_0}{y} \right) k$

2.12. Многофазные системы массового обслуживания

В предыдущих разделах были рассмотрены модели СМО, в которых обслуживание запросов производится одним прибором либо одним из нескольких параллельных идентичных приборов. Вместе с тем процесс обработки запросов во многих реальных системах состоит из их последовательной обработки в нескольких обслуживающих устройствах. Системы обслуживания такого вида явились прототипом сетей массового обслуживания и получили название многофазных СМО.

Многофазные СМО принято кодировать в виде последовательности символов типа:

$$A_1|B_1|n_1|m_1 \rightarrow |B_2|n_2|m_2 \rightarrow \ldots \rightarrow |B_r|n_r|m_r.$$

Здесь символ A_1 описывает входящий поток на вход цепочки из r последовательных обслуживающих устройств, символы B_k, n_k, m_k, $k = 1, ..., r$, описывают соответственно распределение времени обслуживания, число параллельных каналов и число мест в буфере для ожидания в k-м звене цепочки. Символы A_1, B_k, $k = 1, \ldots, r$, принимают значение в тех же множествах, как и соответствующие символы в описании однофазных систем, изученных нами выше. Символ → означает переход запроса на вход следующего обслуживающего устройства по окончании его обслуживания в предыдущем. В некотором смысле этот символ является заменителем символа, задающего вид входящего потока в соответствующую систему обслуживания, поскольку входящий поток в данную СМО определяется выходящим потоком запросов из предыдущей СМО.

Если какой-либо символ принимает конечное значение, то возникает вопрос о поведении многофазной СМО в ситуации, когда буфер перед соответствующей однофазной СМО уже полон, а на вход этой СМО поступает следующий запрос. Обычно рассматриваются два варианта: поступающий запрос теряется и поступающий запрос остается на приборе, где он закончил обслуживание, временно блокируя дальнейшую работу этого прибора.

Кратко коснемся трех простых многофазных СМО.

1. *Многофазная СМО с бесконечным буфером перед первой фазой*, на вход которой поступает простейший поток интенсивности λ, а время обслуживания в каждой из фаз имеет показательное распределение с параметром μ. Предполагается, что одновременно в системе может обслуживаться только один запрос. Только по завершении прохождения запросом всей цепочки приборов на обслуживание может быть выбран следующий запрос.

Эрланговская случайная величина с параметрами (μ, r) есть сумма r независимых случайных величин с параметром μ. Все характеристики рассматриваемой СМО легко получить из характеристик соответствующей однофазной СМО типа $M|E_r|1$. Анализ такой СМО можно провести с использованием так называемого метода фаз Эрланга. Соответствующие результаты можно получить также из формул для системы $M|G|1$.

Введем формулу Поллячека – Хинчина для производящей функции распределения числа запросов в системе $M|G|1$:

$$\Pi(z) = (1-\rho)\frac{(1-z)\beta(\lambda(1-z))}{\beta(\lambda(1-z))-z},$$

где $\beta(s)$ – преобразование Лапласа – Стилтьеса.

Для производящей функции $P(z)=\sum_{i=0}^{\infty} p_i z^i$ распределения $p_i, i \ge 0$, числа запросов в системе из формулы Поллячека – Хинчина с учетом равенства $\beta(s)=\left(\frac{\mu}{\mu+s}\right)^r$ получаем

$$P(z) = (1-\rho)\frac{(1-z)}{1-z(1+y(1-z))^r}, \tag{2.61}$$

где коэффициент нагрузки ρ определяется как $\rho = \frac{r\lambda}{\mu} = ry$.

Обозначим через z_i, $i = 1,\ldots,r$, ноли знаменателя в (2.61), по модулю большие единицы:

$$1 - z\left(1 + \frac{\lambda}{\mu}(1-z)\right)^r. \qquad (2.62)$$

Разлагая правую часть уравнения (2.61) на простые дроби и используя свойства производящей функции, получаем

$$p_0 = 1 - \rho,$$

$$p_i = (1-p)\sum_{i=1}^{r} K_l z_l^{-ir} \frac{1-z_l^r}{1-z_l}, \qquad i \ge 1, \qquad (2.63)$$

где коэффициенты K_m, $m = 1,\ldots,r$, определяются следующим образом:

$$K_m = \prod_{l=1, l \ne m}^{r} \frac{1}{1 - \frac{z_m}{z_l}}, \quad m = 1,\ldots,r. \qquad (2.64)$$

2. *Многофазная СМО из r последовательных многоканальных СМО с бесконечным буфером:*

$$M|M|n_1|\infty \to M|n_2|\infty \to \ldots M|n_r|\infty.$$

Пусть λ – интенсивность входящего потока, μ_k, $k = 1,\ldots,r$, – интенсивность обслуживания в каждом из этих каналов k-й системы, $k = 1,\ldots,r$.

Обозначим $i_k(t)$ – число запросов в k-й системе в момент времени t, $t \ge 0$, $k = 1,..,r$; $p(i_1,..,i_r)$ – стационарная вероятность состояния $\{i_1,..,i_r\}$ рассматриваемой СМО, т.е.

$$p(i_1,\ldots,i_r) = \lim_{t\to\infty} P\{i_1(t) = i_1,\ldots,i_r(t) = i_r\}, \; i_k \ge 0, \; k = 1,\ldots,r. \qquad (2.65)$$

Можно показать, что условием существования пределов (2.65) является выполнение неравенств

$$\rho_i = \frac{\lambda}{n_i \mu_i} < 1, \quad i = 1,\ldots,r. \qquad (2.66)$$

Далее считаем эти неравенства выполненными.

Используя «Δt-метод», можно получить систему дифференциальных уравнений для вероятностей $P\{i_1(t)=i_1,..,i_r(t)=i_r\}$, откуда, переходя к пределу при $t\to\infty$, получим следующую систему уравнений для стационарных вероятностей $p(i_1,..,i_r)$:

$$\left(\lambda+\sum_{k=1}^{r}\alpha_k(i_k)\mu_k\right)p(i_1,\ ...,\ i_r)=\lambda p(i_1-1,\ ...,\ i_r)(1-\delta_{i_1,0})+$$

$$+\sum_{k=1}^{r-1}p(i_1,\ ...,\ i_k+1,\ i_{k+1}-1,i_{k+2},\ ...,\ i_r)\alpha_k(i_k+1)(1-\delta_{i_{k+1},0})\mu_k+$$

$$+\ p(i_1,\ldots,i_{r-1},i_r+1)\alpha_r(i_r+1)\mu_r,\quad i_k\geq 0,\quad k=1,\ldots,r, \tag{2.67}$$

где

$$\alpha_k(i)=\begin{cases}i, 0\leq i\leq n_k,\\ n_k, i>n_k.\end{cases}\qquad \delta_{ig}=\begin{cases}1, i=j,\\ 0, i\neq j.\end{cases}$$

Непосредственной подстановкой несложно убедиться, что решение системы (2.67) имеет следующий вид:

$$p(i_1,..,i_r)=p(0,..,0)\prod_{k=1}^{r}c_k(i_k), \tag{2.68}$$

где

$$c_k(i)=\begin{cases}\dfrac{(n_k\rho_k)^i}{i!},\ \ 0\leq i\leq n_k,\\[2ex] \dfrac{(n_k)^{n_k}}{n_k!}\rho_k^i,\ \ i>n_k.\end{cases}$$

Вероятность $p(0,..,0)$ находится из условия нормировки

$$\sum_{i_1=0}^{\infty}\ldots\sum_{i_r=0}^{\infty}p(i_1,..,i_r)=1 \tag{2.69}$$

и имеет вид

$$p(0,...,0)=\prod_{k=1}^{r}\left(\sum_{i_k=0}^{\infty}c_k\left(i_k\right)\right)^{-1}. \tag{2.70}$$

Из (2.68), (2.69) и формул для n-канальной системы с бесконечной очередью ожидания получаем представление стационарных вероятностей $p(i_1,..,i_r)$, рассматриваемой СМО в виде

$$p(i_1,..,i_r)=\prod_{k=1}^{r}\lim_{t\to\infty}P\{i_k(t)=i_k\}\ , \tag{2.71}$$

то есть совместная вероятность того, что в произвольный момент времени на k-й фазе находится i_k запросов, $k=1,\dots,r,$ равна произведению вероятностей того, что i_k запросов в данный момент находится на k-й фазе, независимо от числа запросов на других фазах, $k=1,\dots,r.$

Этот факт, позволяющий рассчитывать распределение вероятностей состояний многофазной системы как произведение вероятностей состояний однофазных СМО, образующих данную СМО, следует из теоремы Берка, которая формулируется следующим образом.

Теорема Берка. В системе с r параллельными каналами, простейшим входящим потоком интенсивности λ и одинаковым для каждого канала показательным распределением времени обслуживания с параметром μ в стационарном состоянии выходящий поток является простейшим потоком интенсивности λ.

Таким образом, обслуживание простейшего потока запросов на каждой фазе многофазной СМО с показательным распределением времени обслуживания не изменяет характера потока, и в результате совместное распределение вероятностей числа запросов в соответствующей многофазной системе имеет мультипликативный вид.

3*. *Многофазная СМО типа* $M|G|1|\infty\to M|n|0$.

Обозначим интенсивность входящего потока – λ, функцию распределения времени обслуживания – $B(t)$, интенсивность обслуживания любым прибором на второй фазе – μ. Предполагаем, что в случае занятости всех приборов на второй фазе в момент окончания обслуживания запроса на первой фазе с вероятностью Θ, $0<\Theta<1$, запрос уходит из системы недообслуженным (теряется), а с дополнительной вероятностью первый прибор блокируется и не обслуживает следующий запрос, пока не освободится прибор на второй фазе. Как крайние случаи при $\Theta=0$, имеем систему с блокировкой прибора, при $\Theta=1$ – систему с потерями.

Будем рассматривать двумерный процесс $i_{t_k} v_{t_k}, k\geq 1,$ где t_k есть k-й момент окончания обслуживания запроса на первой фазе; $i_{t_k}, i_{t_k}\geq 0,$ – число запросов на первой фазе; $v_{t_k}, v_{t_k}=0,\dots,n,$ – число запросов на второй фазе в момент времени $t_k + 0$. Этот процесс является двумерной цепью Маркова с дискретным временем.

Обозначим одношаговые вероятности переходов этой цепи:

$$P\{i_{t_{k+1}} = l, v_{t_{k+1}} = v' | i_{t_k} = i, v_{t_k} = v\} = P\{(i,v) \to (l,v')\},\ i > 0.$$

Введем производящую функцию:

$$R_{v,v'}(z) = \sum_{l=i-1}^{\infty} P\{(i,v) \to (l,v')\} z^{l-i+1}.$$

Аналогично вложенной цепи для системы $M|G|1$ вероятности переходов $P\{(i,v) \to (l,v')\}$ зависят от значения $l - i$, но не зависят от i и l отдельно. Это делает введенное определение производящей функции $R_{v,v'}(z)$ корректным.

Анализируя переходы двумерной цепи Маркова, можно убедиться, что производящие функции $R_{v,v'}(z)$ определяются следующим образом:

$$R_{v,v'}(z) = \begin{cases} \Gamma(z,n,n)\left(\theta + \dfrac{(1-\theta)n\mu}{n\mu + \lambda(1-z)}\right), v = n,\ v' = n, \\ \Gamma(z,v,v'-1),\ v' \le v \le n,\ v' \ne n, \end{cases}$$

где

$$\Gamma(z,v,v') = \int_0^{\infty} e^{-\lambda(1-z)t} C_v^{v'} e^{-\mu v' t} (1 - e^{-\mu t})^{v-v'} dB(t),$$

$$0 \le v' \le v \le n.$$

Составим из производящих функций $R_{v,v'}(z)$ матричную производящую функцию $R(z)$, введем в рассмотрение матрицу Δ, состоящую из элементов

$$\Delta_{v,v'} = \int_0^{\infty} \lambda e^{-\lambda t} C_v^{v'} e^{-\mu v' t} (1 - e^{-\mu t})^{v-v'} dt,$$

$$0 \le v' \le v \le n.$$

Матрица Δ характеризует вероятности переходов числа занятых каналов на второй фазе системы за время, когда прибор на первой фазе простаивает, ожидая прихода запроса.

Обозначим:

$$\pi(i,v) = \lim_{k \to \infty} P\{(i_{t_k}(t) = i,\ v_{t_k} = v)\},$$

$$\vec{\pi}(i) = (\pi(i,0), \pi(i,1), \dots, \pi(i,n)),$$

$$\vec{\Pi}(z) = \sum_{i=0}^{\infty} \vec{\pi}(i) z^i, \quad |z| < 1. \tag{2.72}$$

Утверждение. Векторная производящая функция $\vec{\Pi}(z)$ удовлетворяет матричному функциональному уравнению:

$$\vec{\Pi}(z)(R(z) - I) = \vec{\Pi}(0)(I - \Delta z)R(z), \tag{2.73}$$

здесь I – тождественная матрица.

Условия существования пределов и алгоритмы решения уравнения (2.73) можно найти, например, в [12, 13]. (http://scask.ru/q_book_tpn.php?id=21)

Рассмотрим другой подход.

Допустим, что специфика ремонтно-эксплуатационного обслуживания предопределяет использование многофазных СМО, когда обслуживание поступившей заявки осуществляется в несколько этапов. Характерная ситуация возникает в электрических сетях при возникновении аварийной ситуации. Проведение ремонтных работ обычно осуществляется в два этапа – отключение поврежденного участка и подготовка рабочего места выполняются оперативно-выездной бригадой, а непосредственно восстановительные работы проводит ремонтное подразделение. При этом в зависимости от сложности повреждения последний вид работ может проводить одна или несколько бригад. Специфичным может быть также отсутствие очереди по причине малой интенсивности потока отключений электрических сетей.

Рассмотрим простейшую СМО без очереди с подготовкой каналов. На n-канальную СМО поступает простейший поток заявок с интенсивностью λ. Время обслуживания заявок распределено по показательному закону с параметром μ. До начала обслуживания канал должен быть подготовлен. Время подготовки $t_{\text{подг}}$ имеет показательное распределение с параметром φ и не зависит от того, как давно канал прекратил работу. При поступлении заявки вначале выполняется операция подготовки, а затем она поступает на обслуживание. Заявка, заставшая все каналы занятыми, на обслуживание не принимается.

В рассматриваемом случае обслуживание заявки состоит из двух фаз: подготовки со временем подготовки $t_{\text{подг}}$ и самого обслуживания в течение времени $t_{\text{обсл}}$. В результате общее среднее время обслуживания равно $\tilde{t}_{\text{обсл}} = \bar{t}_{\text{подг}} + \bar{t}_{\text{обсл}}$.

Случайная величина $\tilde{t}_{\text{обсл}}$ обычно распределена по обобщенному закону Эрланга второго порядка с параметрами μ и φ.

Известно, что формулы Эрланга справедливы не только для показательного, но и для любого другого распределения времени обслуживания. Для решения практических задач нам необходимо найти величину $\tilde{\mu}$:

$$t_{\text{обсл}} = \frac{1}{\mu} + \frac{1}{\varphi} = \frac{\mu + \varphi}{\mu\varphi}, \tag{2.74}$$

следовательно,

$$\tilde{\mu} = \frac{\mu\varphi}{\mu + \varphi}. \tag{2.75}$$

Определив $\tilde{y} = \dfrac{\lambda}{\tilde{\mu}}$ и подставив его в формулы Эрланга, получим

$$p_0 = \left(1 + \frac{\tilde{y}}{1!} + \ldots + \frac{\tilde{y}^k}{k!} + \ldots + \frac{\tilde{y}^n}{n!}\right)^{-1}; \tag{2.76}$$

$$p_k = \frac{\tilde{y}^k}{k!} p_0, \ (1 \le k \le n); \tag{2.77}$$

$$P_{\text{отк}} = \tilde{p}_n = \frac{\tilde{y}^n}{n!} p_0. \tag{2.78}$$

Показатели эффективности СМО определяются по известным формулам:

$$Q = 1 - \frac{\tilde{y}^n}{n!} p_0; \qquad A = \lambda Q = \lambda\left(1 - \frac{\tilde{y}^n}{n!} p_0\right). \tag{2.79}$$

Для определения среднего числа занятых каналов надо разделить A на $\tilde{\mu}$:

$$\bar{k}_{\text{зан}} = \frac{A}{\tilde{\mu}} = \frac{\lambda}{\tilde{\mu}}\left(1 - \frac{\tilde{y}^n}{n!} p_0\right) = \tilde{y}\left(1 - \frac{\tilde{y}^n}{n!} p_0\right). \tag{2.80}$$

Пример 22. Ввиду сложности возникшей аварийной ситуации в электрической сети на ее устранение были направлены оперативно-выездная бригада и две ремонтные бригады электромонтеров. Интенсивность потока отключений электрической сети $\lambda = 0{,}2$ 1/ч, среднее время подготовки к ремонту $t_{\text{подг}} = 0{,}5$ ч, среднее время ремонта $t_{\text{обсл}} = 2$ ч. Определить вероятности состояний и показатели эффективности СМО.

Решение:

1. По условию задачи определяем: $n = 3$; $\lambda = 0{,}2$ $1/$ ч; $\mu_1 = 2$ ч; $\mu_2 = 0{,}5$ ч. Тогда по (2.75)

$$\tilde{\mu} = \frac{\mu\varphi}{\mu + \varphi} = 0{,}5 \cdot \frac{2}{0{,}5 + 2} = 0{,}4 \text{ и } \tilde{y} = \frac{0{,}2}{0{,}4} = 0{,}5.$$

2. Рассчитываем вероятности состояний:

$$p_0 = \left(1 + \frac{\tilde{y}}{1!} + \frac{\tilde{y}^2}{2!}\right)^{-1} = \left(1 + \frac{0{,}5}{1} + \frac{0{,}25}{2}\right)^{-1} = 0{,}62,$$

$$P_{\text{отк}} = \tilde{p}_n = \frac{\tilde{y}^2}{2!} p_0 = \frac{0{,}25}{2} 0{,}62 = 0{,}08.$$

Определяем показатели эффективности СМО:

$$Q = 1 - \frac{\tilde{y}^n}{n!} p_0 = 1 - 0{,}08 = 0{,}92; \ A = \lambda Q = 0{,}2 \cdot 0{,}92 = 0{,}184,$$

$$\bar{k}_{\text{зан}} = \frac{A}{\tilde{\mu}} = \frac{0{,}184}{0{,}4} = 0{,}46.$$

Как видим, СМО недостаточно загружена, так как число занятых каналов составляет 0,46.

2.13. Системы массового обслуживания со «взаимопомощью» между каналами

До сих пор мы рассматривали только такие СМО, в которых каждая заявка может обслуживаться только одним каналом; незанятые каналы не могут «помогать» занятому в обслуживании.

Вообще, это не всегда бывает так: встречаются системы массового обслуживания, где одна и та же заявка может одновременно обслуживаться двумя и более каналами. Например, один и тот же вышедший из строя станок могут обслуживать два рабочих сразу. Такая «взаимопомощь» между каналами может иметь место как в открытых, так и в замкнутых СМО.

При рассмотрении СМО со взаимопомощью между каналами необходимо учитывать два фактора:

1. Насколько убыстряется обслуживание заявки, когда над ним работает не один, а сразу несколько каналов?

2. Какова «дисциплина взаимопомощи», т.е. когда и как несколько каналов берут на себя обслуживание одной и той же заявки?

Рассмотрим сначала первый вопрос. Естественно предположить, что, если над обслуживанием заявки работает не один канал, а несколько (k) каналов, интенсивность потока обслуживаний не будет убывать с увеличением k, т.е. будет представлять собой некоторую неубывающую функцию числа k работающих каналов. Обозначим эту функцию $\mu(k)$. Возможный вид функции $\mu(k)$ показан на рис. 2.25.

Очевидно, что неограниченное увеличение числа одновременно работающих каналов не всегда ведет к пропорциональному увеличению скорости обслуживания; естественнее предположить, что при некотором критическом значении $k = k_{кр}$ дальнейшее увеличение числа занятых каналов уже не повышает интенсивности обслуживания.

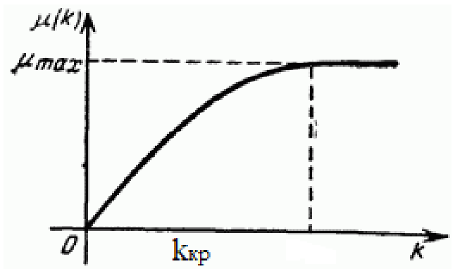


Рис. 2.25. Возможный вид функции $\mu(k)$

Для того чтобы проанализировать работу СМО со взаимопомощью между каналами, нужно прежде всего задать вид функции $\mu(k)$.

Самым простым для исследования будет случай, когда функция $\mu(k)$ возрастает пропорционально k при $k \le k_{кр}$, а при $k > k_{кр}$ остается постоянной и равной $\mu_{max} = k_{кр}$ (рис. 2.26). Если при этом общее число каналов n, которые могут помогать друг другу, не превосходит $k_{кр}$

$$n \le k_{\text{кр}},$$

то можно считать интенсивность обслуживания заявки несколькими каналами пропорциональной числу каналов.

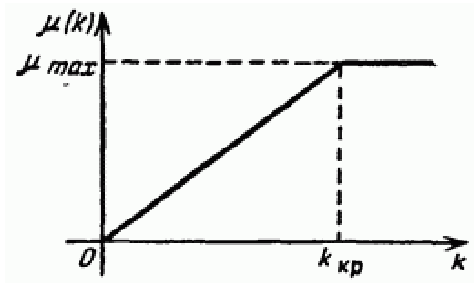


Рис. 2.26. Функция $\mu(k)$ возрастает пропорционально k при $k \le k_{\text{кр}}$, а при $k > k_{\text{кр}}$ остается постоянной

Остановимся теперь на втором вопросе: дисциплине взаимопомощи. Самый простой случай этой дисциплины мы обозначим условно «все как один». Это означает, что при появлении одной заявки ее начинают обслуживать все n каналов сразу и остаются занятыми, пока не закончится обслуживание этой заявки; затем все каналы переключаются на обслуживание другой заявки (если она есть) или ждут ее появления, если ее нет, и т.д. Очевидно, в этом случае все n каналов работают как один, СМО становится одноканальной, но с более высокой интенсивностью обслуживания.

Возникает вопрос: выгодно или невыгодно вводить такую взаимопомощь между каналами? Ответ на этот вопрос зависит от того, какова интенсивность потока заявок, каков вид функции $\mu(k)$, каков тип СМО (с отказами, с очередью), какая величина выбирается в качестве характеристики эффективности обслуживания.

Пример 23. Имеется трехканальная СМО с отказами: интенсивность потока заявок $\lambda = 4$ заявки в 1 мин, среднее время обслуживания одной заявки одним каналом $\bar{t}_{\text{об}} = 4$ мин, функция $\mu(k) = k\mu$. Спрашивается, выгодно ли с точки зрения пропускной способности СМО вводить взаимопомощь между каналами по типу «все как один»? Выгодно ли это с точки зрения уменьшения среднего времени пребывания заявки в системе?

Решение:

1. *Без взаимопомощи*: $n = 3,\ \mu = 1/0{,}5 = 2,\ \lambda = 4, y = \lambda/\mu = 2,$

$$p_0 = \left(1 + \frac{2}{1} + \frac{4}{2!} + \frac{8}{3!}\right)^{-1} = \frac{3}{19} \approx 0{,}158;$$

$$P_{\text{отк}} = p_3 = \frac{8}{3!} \cdot \frac{3}{19} = 0,21.$$

По формулам Эрланга имеем:

– относительную пропускную способность СМО:

$$Q = 1 - P_{\text{отк}} = 1 - 0,21 = 0,79;$$

– абсолютную пропускную способность:

$$A = \lambda Q = 4 \cdot 0,79 = 3,16.$$

Среднее время пребывания заявки в СМО найдется, как вероятность того, что заявка будет принята к обслуживанию, умноженная на среднее время обслуживания:

$$\bar{t}_{\text{сист}} = 0,79 \cdot 0,5 = 0,5 \text{ мин.}$$

Не нужно забывать, что это среднее время относится ко всем заявкам – как обслуженным, так и необслуженным Нас же может интересовать среднее время, которое пробудет в системе обслуженная заявка. Это время равно:

$$\bar{t}_{\text{сис}}^{(\text{об})} = \bar{t}_{\text{об}} = 0,5 \text{мин.}$$

2. *Со взаимопомощью*:

$$n^* = 1, \quad \lambda = 4, \qquad \mu^* = 3\mu = 6, \quad y^* = \frac{\lambda}{\mu^*} = \frac{2}{3};$$

$$p_0 = \frac{1}{1 + \frac{2}{3}} = \frac{3}{5}; \; p_1 = \frac{2}{3} \cdot \frac{3}{5} = \frac{2}{5}; \; p_{\text{отк}} = p_1 = \frac{2}{5}; \; Q = 1 - \frac{2}{5} = \frac{3}{5} = 0,6;$$

$$A = \lambda Q = 4 \cdot 0,6 = 2,4.$$

Среднее время пребывания заявки в СМО:

$$\bar{t}_{\text{сист}} = p_1 \cdot \frac{1}{3\mu} = \frac{2}{5 \cdot 6} = \frac{1}{15} = 0,06667 \text{мин.}$$

Среднее время пребывания обслуженной заявки в СМО:

$$\bar{t}_{\text{сис}}^{(\text{об})} = \frac{1}{3\mu} = \frac{1}{6} = 0,167 \text{мин.}$$

Таким образом, при наличии взаимопомощи «все как один» пропускная способность СМО заметно уменьшилась. Это объясняется увеличением вероятности отказа: за то время, пока все каналы заняты обслуживанием одной заявки, могут прийти другие заявки и, естественно, получить отказ. Что касается среднего времени пребывания заявки в СМО, то оно, как и следовало ожидать, уменьшилось. Если по каким-то соображениям мы стремимся ко всемерному уменьшению времени, которое заявка проводит в СМО (например, если пребывание в СМО опасно для заявки), может оказаться, что, несмотря на уменьшение пропускной способности, все же будет выгодно объединить три канала в один.

Рассмотрим теперь влияние взаимопомощи типа «все как один» на работу СМО с ожиданием. Возьмем для простоты только случай неограниченной очереди. Естественно, влияния взаимопомощи на пропускную способность СМО в этом случае не будет, так как при любых условиях обслужены будут все пришедшие заявки. Возникает вопрос о влиянии взаимопомощи на характеристики ожидания: среднюю длину очереди, среднее время ожидания, среднее время пребывания в СМО.

В силу формул для обслуживания без взаимопомощи получаем:

– среднее число заявок в очереди:

$$\bar{L}_{\text{оч}} = \frac{y^{n+1} p_0}{n \cdot n!(1-\chi)^2}; \tag{2.81}$$

– среднее время ожидания:

$$\bar{t}_{\text{ож}} = \frac{y^n p_0}{n \cdot \mu \cdot n!(1-\chi)^2}; \tag{2.82}$$

– среднее время пребывания в системе:

$$\bar{t}_{\text{сист}} = \bar{t}_{\text{ож}} + \frac{1}{\mu}, \tag{2.83}$$

где

$$p_0 = \frac{1}{1 + \frac{y}{1!} + \frac{y^2}{2!} + \ldots + \frac{y^n}{n!} + \frac{y^{n+1}}{n!(n-y)}}, \quad \chi = \frac{y}{n}. \tag{2.84}$$

Если же применяется взаимопомощь типа «все как один», то система будет работать как одноканальная с параметрами

$$y^* = \frac{\lambda}{\mu^*} = \frac{\lambda}{n\mu} = \frac{y}{n} = \chi$$

и ее характеристики определятся формулами для одноканальной очереди с ожиданием:

$$\overline{L}_{\text{оч}} = \frac{\chi^2}{1-\chi}\ , \tag{2.85}$$

$$\overline{t}_{\text{ож}} = \frac{1}{(n\mu)} \cdot \frac{\chi}{(1-\chi)}, \tag{2.86}$$

$$\overline{t}_{\text{сист}} = \overline{t}_{\text{ож}} + \frac{1}{n\mu} = \frac{\chi}{n\mu(1-\chi)}. \tag{2.87}$$

Пример 24. Имеется трехканальная СМО с неограниченной очередью; интенсивность потока заявок $\lambda = 4$ заявки в 1 мин, среднее время обслуживания $\overline{t}_{\text{об}} = 0,5$ мин. Функция $\mu(k) = k\mu$, $k_{\text{кр}} > 3$. Выгодно ли вводить взаимопомощь между каналами типа «все как один», с учетом:

– средней длины очереди;

– среднего времени ожидания обслуживания;

– среднего времени пребывания заявки в СМО?

Решение:

1. *Без взаимопомощи*:

$$n = 3,\ \lambda = 4,\ \ \mu = \frac{1}{0,5} = 2,\ \ y = \frac{\lambda}{\mu} = 2,\ \ \chi = \frac{\rho}{n} = \frac{2}{3}.$$

По формулам (2.81)–(2.84) получаем:

$$p_0 = \frac{1}{1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^n}{3!} + \frac{y^{n+1}}{3!(3-2)}} = \frac{1}{9};$$

$$\overline{L}_{\text{оч}} = \frac{2^4 \cdot \frac{1}{9}}{3 \cdot 3!(1-2/3)^2} = \frac{8}{9} = 0,889;$$

$$\overline{t}_{\text{ож}} = \frac{y^n p_0}{n \cdot \mu \cdot n!(1-\chi)^2} = \frac{\overline{L}_{\text{оч}}}{\lambda} = \frac{2}{9} = 0,222;$$

$$\overline{t}_{\text{сист}} = \overline{t}_{\text{ож}} + \frac{1}{\mu} = \frac{2}{9} + \frac{1}{2} = 0,722.$$

2. *Со взаимопомощью*:

$$n^* = 1,\ \lambda = 4,\quad \mu^* = 3\mu = 6,\quad y^* = \chi = \frac{y}{\mu^*} = \frac{2}{3}.$$

По формулам (2.85)–(2.87) находим:

$$\bar{L}_{\text{оч}} = \frac{\chi^2}{1-\chi} = \frac{\left(\frac{2}{3}\right)^2}{1/3} = \frac{4}{3};$$

$$\bar{t}_{\text{ож}} = \frac{\bar{L}_{\text{оч}}}{\lambda} = \frac{\frac{4}{3}}{4} = \frac{1}{3};$$

$$\bar{t}_{\text{сист}} = \bar{t}_{\text{ож}} + \frac{1}{n\mu} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2} = 0,5.$$

Таким образом, средняя длина очереди и среднее время ожидания в очереди в случае взаимопомощи больше, но среднее время пребывания заявки в системе – меньше.

Из рассмотренных примеров видно, что взаимопомощь между каналами типа «все как один», как правило, не способствует повышению эффективности обслуживания: время пребывания заявки в СМО уменьшается, но зато ухудшаются другие характеристики обслуживания.

Поэтому желательно изменить дисциплину обслуживания так, чтобы взаимопомощь между каналами не мешала принимать к обслуживанию новые заявки, если они появятся за время, пока все каналы заняты.

Назовем условно «равномерной взаимопомощью» следующий тип взаимопомощи. Если заявка приходит в момент, когда все каналы свободны, то все n каналов принимаются за ее обслуживание; если в момент обслуживания заявки приходит еще одна, часть каналов переключается на ее обслуживание; если, пока обслуживаются эти две заявки, приходит еще одна, часть каналов переключается на ее обслуживание и т.д., до тех пор, пока не окажутся занятыми все n каналов; если это так, вновь пришедшая заявка получает отказ (в СМО с отказами) или становится в очередь (в СМО с ожиданием).

При такой дисциплине взаимопомощи заявка получает отказ или становится в очередь только тогда, когда нет возможности ее обслужить. Что касается «простоя» каналов, то он в этих условиях минимален: если в системе имеется хотя бы одна заявка, все каналы работают.

Выше мы упомянули, что при появлении новой заявки часть занятых каналов освобождается и переключается на обслуживание вновь прибывшей заявки. Какая часть? Это зависит от вида функции $\mu(k)$. Если она имеет вид линейной зависимости, как показано на рис. 2.26, и $k_{кр} > n$, то все равно, какую часть каналов выделить на обслуживание вновь поступившей заявки, лишь бы все каналы были заняты (тогда суммарная интенсивность обслуживаний при любом распределении каналов по заявкам будет равна $n\mu$). Можно доказать, что если кривая $\mu(k)$ выпукла кверху, как показано на рис. 2.25, то нужно распределять каналы по заявкам как можно более равномерно.

Рассмотрим работу n-канальной СМО при «равномерной» взаимопомощи между каналами.

2.13.1. Система массового обслуживания с отказами

Будем нумеровать состояния СМО по числу заявок, находящихся в состоянии обслуживания (рис. 2.27):

E_0 – СМО свободна;

E_1 – одна заявка обслуживается всеми n каналами;

E_2 – две заявки обслуживаются всеми n каналами;

E_k – k заявок обслуживаются всеми n каналами;

E_n – n заявок обслуживаются всеми n каналами.

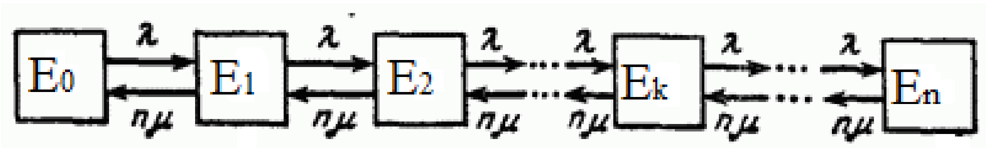


Рис. 2.27. Граф состояний

Граф состояний (см. рис. 2.27) здесь тот же, что для одноканальной СМО с производительностью $\mu^* = n\mu$ и ограниченной очередью, имеющей $(n - 1)$ мест. Поэтому для определения характеристик системы мы можем воспользоваться формулами для одноканальной СМО с ожиданием, подставляя в них $\chi = \dfrac{\lambda}{\mu^*}$ вместо $y = \lambda/\mu$:

$$P_{\text{отк}} = \frac{\chi^n(1-\chi)}{1-\chi^{n+1}}; \tag{2.88}$$

$$Q = \frac{1-\chi^n}{1-\chi^{n+1}}; \tag{2.89}$$

$$A = \lambda Q = \lambda \frac{1-\chi^n}{1-\chi^{n+1}}. \qquad (2.90)$$

Пример 25. В условиях примера 23 сравнить относительную и абсолютную пропускную способности СМО, а также среднее число занятых каналов:

а) при отсутствии взаимопомощи;

б) при наличии равномерной взаимопомощи между каналами.

Решение.

1. *Без взаимопомощи.* Из примера 23 имеем: Q = 0,79, A = 3,16. Среднее число занятых каналов $\bar{k}_{\text{зан}} = \frac{A}{\mu} = \frac{3,16}{2} = 1,58.$

2. *С равномерной взаимопомощью*:

$$\chi = \frac{\lambda}{\mu*} = \frac{4}{2} \cdot 3 = 2/3.$$

По формуле (2.89) получаем:

$$Q = \frac{1-(2/3)^3}{1-(2/3)^4} = 0,887; \;\; A = 4Q = 3,51; \;\; \bar{k}_{\text{зан}} = \frac{A}{\mu} = \frac{3,51}{2} = 1,76.$$

Таким образом, за счет применения разумно организованной взаимопомощи между каналами пропускная способность СМО несколько повысилась. Соответственно увеличилась и средняя занятость каналов.

2.13.2. Система массового обслуживания с очередью

Рассмотрим СМО с очередью и максимальным числом заявок в очереди m. Предположим, что между каналами имеется «равномерная» взаимопомощь и $\mu(k) = k\mu$. Состояния системы опять будем нумеровать по числу заявок, находящихся в СМО:

E_0 – система свободна;

E_1 – одна заявка обслуживается всеми n каналами;

E_2 – две заявки обслуживаются всеми n каналами;

E_k – k заявок обслуживаются всеми n каналами, очереди нет;

E_n – n заявок обслуживаются всеми n каналами, очереди нет;

E_{n+1} – n заявок обслуживаются всеми n каналами, одна заявка стоит в очереди;

E_{n+m} – заявок обслуживаются всеми n каналами, в очереди стоит m заявок.

Граф состояний СМО приведен на рис. 2.28.

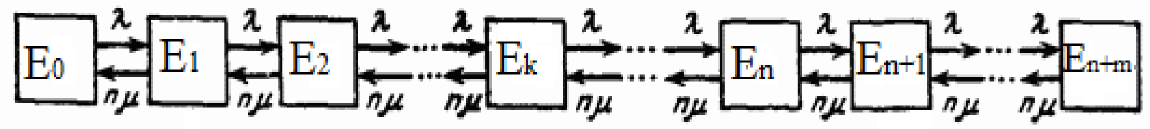


Рис. 2.28. Граф состояний СМО

Мы опять получили граф того же вида, что и на рис. 2.27, но с увеличенным на m числом состояний. Значит, нам нужно воспользоваться формулами для одноканальной СМО с производительностью $\mu^* = n\mu$ и числом мест в очереди $n + m - 1$. Получим:

$$P_{\text{отк}} = \frac{\chi^{n+m}(1-\chi)}{1-\chi^{n+m+1}}; \tag{2.91}$$

$$Q = \frac{1-\chi^{n+m}}{1-\chi^{n+m+1}}; \tag{2.92}$$

$$A = \lambda Q = \lambda \frac{1-\chi^{n+m}}{1-\chi^{n+m+1}}. \tag{2.93}$$

Пример 26. В условиях примера 23 сравнить абсолютную и относительную пропускные способности для случая отсутствия взаимопомощи и наличия равномерной взаимопомощи, если в очереди имеется два места ($m = 2$).

Решение:

1. *Без взаимопомощи*. Из примера 23 имеем $Q \approx 0{,}79$, $A \approx 3{,}16$.

2. *С равномерной взаимопомощью*. По формулам (2.91)–(2.93) для $n = 3,\ \lambda = 4,\ \mu = 2,\ \chi = \dfrac{y}{n}$ имеем:

$$Q = \frac{1-\left(\frac{2}{3}\right)^3}{1-\left(\frac{2}{3}\right)^4} = \frac{57}{65} = 0{,}88; \qquad A = \lambda Q = 3{,}52;$$

Предоставляем читателю самостоятельно подсчитать среднее число заявок в очереди, среднее время ожидания и среднее время пребывания в системе для обоих вариантов примера 26 и убедиться, что при наличии равномерной взаимопомощи между каналами все характеристики СМО меняются только в желательном для нас направлении.

Не следует, однако, забывать, что организация такой взаимопомощи между каналами далеко не для всех СМО осуществима.

2.14. Оптимизация структуры системы массового обслуживания

2.14.1. Постановка оптимизационной задачи

Более 100 лет назад П. Л. Чебышев писал: «Математика создавалась и развивалась под влиянием общей и основной задачи всей человеческой деятельности: распорядиться имеющимися под руками средствами для достижения наибольшей выгоды». Теория массового обслуживания является прикладным разделом математики, и к ней в полной мере относятся слова выдающегося российского ученого.

Как известно, теория массового обслуживания изучает явления, носящие случайный характер, когда отдельные люди или объекты пользуются ограниченными средствами. Иногда вследствие случайных колебаний количество требований может превысить имеющиеся средства, и некоторые требования будут вынуждены либо ожидать обслуживания в очереди, либо вовсе отказаться от обслуживания. Как первое, так и второе приводит к нежелательным издержкам.

Для наглядности изложения воспользуемся уровнем микроэкономики. Будем полагать, что некоторая фирма или организация располагает средствами для удовлетворения конкретных потребностей в обслуживании и извлекает этот доход, удовлетворяя эти потребности. Речь может идти о парикмахерской, магазине, мастерской и т.д.

Общая схема процесса обслуживания приведена на рис. 2.29. Она позволяет описать многие реальные ситуации. Обычно при этом экономическая сторона деятельности фирмы может быть представлена так. Расходы фирмы складываются из постоянной и переменной составляющих. Первая практически не зависит от того, как организован процесс обслуживания требований, и, как правило, включает:

– затраты на приобретение и установку оборудования;
– расходы на подготовку помещения;
– арендную плату, оплату коммунальных услуг и пр.

входящий поток требований на обслуживание → *накопитель для ожидающих обслуживания требований* → *средства для удовлетворения потребностей* → *обслуженные требования*

Рис. 2.29. Общая схема процесса обслуживания

Переменная составляющая пропорциональна числу обслуженных требований и может существенно зависеть от организации процесса обслуживания.

Доходы фирмы складываются из платы за обслуживание требований, и в большинстве ситуаций рациональная организация процесса обслуживания может их существенно повысить.

Фирма стремится организовать процесс обслуживания таким образом, чтобы максимизировать получаемую прибыль, представляющую собой разность между доходами и затратами фирмы.

Исследуем, на какие параметры системы обслуживания можно влиять, чтобы увеличить прибыль:

1. Входящий поток требований. Очевидно, что на его интенсивность можно влиять, активизируя рекламу, снижая стоимость услуг для привлечения новых клиентов и сохранения старых и т.д. Все эти методы выходят за рамки теории массового обслуживания, поэтому входящий поток требований будем считать заданным.

2. Объем накопителя, измеряемый максимальным числом требований, которые одновременно могут ожидать обслуживания. Увеличение объема накопителя увеличивает расходы фирмы на его организацию и содержание, но одновременно увеличивает и доход фирмы, так как при этом уменьшается доля требований, получивших отказ в обслуживании из-за отсутствия свободных мест в накопителе.

3. Дисциплина очереди, т.е. правило, по которому из нескольких требований, находящихся в данный момент в очереди, выбирается очередное требование на обслуживание. Если с точки зрения доходов и расходов все требования одинаковы, то дисциплина их обслуживания для фирмы не играет никакой роли. Однако, если предусмотрена плата за срочность обслуживания, за об-

служивание без очереди или если нахождение в очереди одних требований увеличивает издержки фирмы более, чем других, то выбор дисциплины очереди может сильно влиять на прибыль.

4. Время обслуживания отдельного требования и число каналов обслуживания. Данные параметры существенно влияют на характеристики системы обслуживания, а следовательно, на расходы и доходы фирмы.

Итак, все параметры системы обслуживания, за исключением входящего потока требований, могут рассматриваться в качестве управляемых, и оптимизационную задачу можно сформулировать так: в системе, изображенной на рис. 2.29, выбрать управляемые параметры таким образом, чтобы прибыль фирмы была максимальна.

К сожалению, решение задачи в столь общей постановке чрезвычайно сложно. Как правило, решают частные задачи оптимизации, когда определяют оптимальное значение одного из управляемых параметров, а значение всех остальных параметров считают постоянными. Иногда найденное частное оптимальное решение оказывается оптимальным при любых значениях остальных параметров. Например, оптимальная дисциплина очереди часто не зависит ни от емкости накопителя, ни от числа каналов обслуживания, ни от времени обслуживания требований.

Рассмотрим основные экономические критерии, используемые при оптимизации СМО. Отметим, что обычно выбор критерия оптимизации – это задача не математика, а руководства фирмы.

2.14.2. Критерий минимума себестоимости продукции

Критерий минимума себестоимости продукции чаще всего используется при исследовании производственных ситуаций, когда в качестве источников требований на обслуживание выступают рабочие станки или механизмы основного производства, а роль обслуживающих устройств играют вспомогательные службы и хозяйства фирмы или цеха (ремонтные подразделения, инструментальные кладовые, транспортные службы и пр.). Хотя вспомогательные службы непосредственного участия в основном производстве не принимают, об оптимальности их организации следует судить по общим результатам работы фирмы.

Очевидно, что из различных вариантов организации обслуживающей системы оптимальным будет тот, при котором себестоимость продукции, выпускаемой цехом, минимальна. В этом случае себестоимость считают функцией от варианта организации обслуживающей системы при постоянном значении прочих факторов.

Варианты организации обслуживающей системы часто задаются количеством каналов обслуживания, которыми являются вспомогательные рабочие или единицы обслуживания вспомогательных служб.

Следовательно, критерий оптимизации может быть записан в виде функции связи себестоимости единицы выпускаемой фирмой продукции с числом каналов обслуживающей системы:

$$C = f\left(N\right) \to \min, \tag{2.94}$$

где C – себестоимость единицы продукции; N – число каналов обслуживающей системы.

Себестоимость продукции высчитывается по формуле

$$C = \frac{З}{B}, \tag{2.95}$$

где З – затраты на производство; B – объем продукции, выпускаемой за фиксированный период.

Обслуживающие системы не имеют прямого отношения к выпускаемой продукции, однако могут влиять на ее объем, так как от качества ее работы зависит длительность простоев обслуживаемых рабочих мест.

Теория массового обслуживания позволяет определить длительность простоев обслуживаемых рабочих мест при различном числе каналов обслуживания. Например, в моделях с ожиданием, с ограниченным потоком требований и с конечным числом каналов обслуживания это будут показатели:

u – коэффициент простоя требования, т.е. относительная доля времени, проводимая требованием в обслуживании и в ожидании его;

d – коэффициент простоя рабочего места в ожидании обслуживания.

Предположим, что в момент проведения исследования фирмы специалистом по ТМО число каналов обслуживания равнялось N_0, а коэффициент простоя требования имел значение u_0. При изменении числа каналов обслуживания до N_1 и сохранении остальных параметров системы неизменными величина u изменится до значения u_1. При этом доля времени, когда рабочие единицы основного производства находятся вне системы обслуживания, т.е.

заняты выпуском продукции, изменится от $(1-u_0)$ до $(1-u_1)$. Объем выпускаемой продукции пропорционален времени работы рабочих единиц основного производства. Поэтому влияние числа каналов обслуживания на объем выпускаемой продукции может быть описано соотношением

$$B_1 = B_0 \frac{1-u_1}{1-u_0}, \qquad (2.96)$$

где B_0 и B_1 – объемы выпускаемой продукции при числе каналов обслуживания N_0 и N_1 соответственно.

Величина N влияет также и на величину затрат на производство З, которые делятся на переменные расходы П и условно-постоянные расходы K.

Переменные расходы меняются пропорционально объему выпускаемой продукции или, что то же самое, пропорционально времени работы, т.е.

$$\Pi_1 = \Pi_0 \frac{1-u_1}{1-u_0}, \qquad (2.97)$$

где Π_0 и Π_1 – переменные расходы при числе каналов обслуживания N_0 и N_1 соответственно.

Условно-постоянные расходы K, включающие в себя расходы на содержание вспомогательных служб и хозяйств фирмы, при изменении N меняются лишь в части заработной платы обслуживающих рабочих с соответствующими отчислениями.

Если эти расходы составляют в среднем на одного рабочего Q рублей, то изменение условно-постоянных расходов K определяется формулой

$$\Delta K = Q(N_1 - N_0), \qquad (2.98)$$

а условно-постоянные расходы можно представить в виде

$$K_1 = K_0 + \Delta K = K_0 + Q(N_1 - N_0), \qquad (2.99)$$

где K_0 и K_1 – значения K, соответствующие N_0 и N_1.

Общие затраты на производство равны сумме переменных и условно-постоянных расходов:

$$З_1 = \Pi_1 + K_1 = \Pi_0 \frac{1-u_1}{1-u_0} + [Q(N_1 - N_0)], \qquad (2.100)$$

где $З_1$ – затраты, соответствующие величине N_1.

Подставив выражения для объема выпуска продукций и затрат на выпуск в формулу (2.95), получим зависимость

$$C_1 = \frac{З_1}{B_1} = \frac{П_0}{B_0} + \frac{[K_0 + Q(N_1 - N_0)](1-u_0)}{B_0(1-u_1)} = f(N_1). \quad (2.101)$$

Переменными в данной функции являются число каналов N_1 и коэффициент простоя рабочих единиц основного производства u_1, который, в свою очередь, зависит от N_1.

Положив в формуле (2.101) $N_1 = N_0$, получим значение себестоимости, соответствующее существующему числу каналов обслуживания N_0, т.е.

$$C_0 = \frac{З_0}{B_0} = \frac{\Pi_0 + K_0}{B_0}.$$

Вычислив ряд значений C_1 при различных значениях N_1, можно выбрать наилучший вариант организации обслуживающей системы, соответствующий минимуму себестоимости.

Приведем окончательные расчетные формулы для системы, удовлетворяющей следующим условиям:

1) число единиц основного оборудования в фирме равно $H < \infty$;

2) число каналов или число единиц вспомогательного оборудования равно N;

3) каждая единица основного оборудования, когда она находится в рабочем состоянии, является источником простейшего потока требований на обслуживание с интенсивностью λ;

4) время обслуживания требования каждым каналом – показательно (экспоненциально) распределенная случайная величина с параметром ν.

Основные характеристики системы обслуживания определяются по следующим формулам:

а) вероятность того, что обслуживанием занято k каналов

$$P_k = \begin{cases} \dfrac{H!}{k!(H-k)!} p^k P_0, & 1 \le k \le N, \\ \dfrac{H!}{N^{k-N}(H-k)!N!} p^k P_0, & N \le k \le H; \end{cases} \quad (2.102)$$

б) коэффициент нахождения требования в обслуживаемой системе:

$$U = \frac{1}{H}\sum_{k=1}^{H} kP_k;$$

фактически это и есть коэффициент простоя основного оборудования, если считать, что во время обслуживания требования его источник – рабочее место основного производства – не производит продукции; величина $(1 - u)$ – это коэффициент загрузки рабочего места;

в) коэффициент простоя канала:

$$d = \frac{1}{N}\sum_{k=0}^{N-1} (N-k)P_k .$$

Рассмотрим пример выбора оптимального числа обслуживающих каналов.

Пример 27. Предположим, что число обслуживаемых станков $H = 6$. Каждый из них требует внимания ремонтного рабочего в среднем 1 раз в 1 ч, т.е. $\lambda = 1$ ч$^{-1}$. Среднее время ремонта составляет 6 мин, т.е. $\nu = 10$ ч$^{-1}$. Значения остальных параметров системы таковы: $П_0$ = 5 у.д.е. (условных денежных единиц), B_0 = 12 изделий, Q = 0,1 у.д.е., K_0= 3 у.д.е.

Предположим, что в момент исследования N_0 = 1, т.е. все станки обслуживаются одним ремонтником, и сведем результаты расчетов в табл. 2.1.

Таблица 2.1

N	P_0	U	C, у.д.е.
1	0,485	0,141	0,667
2	0,560	0,094	0,664
3	0,564	0,091	0,669
4	0,567	0,089	0,671

Из табл. 2.1 видно, что минимум себестоимости продукции достигается, если обслуживание станков ведут два ремонтника.

Используют и другие критерии оптимальности, например максимальный выпуск продукции на основном производстве. При этом не принимается во внимание, что себестоимость продукции может и не быть минимальной. Такой критерий имеет вид

$$B = f(N) = B_0 \frac{1-u_1}{1-u_0} \to \max. \qquad (2.103)$$

Указанный критерий целесообразно принять, например, в том случае, если обнаруживаемый цех или участок является «узким местом» и недостаточное производство изделий на данном участке может приводить к большим экономическим потерям. Например, участок производства комплектующих изделий для автомобиля тормозит работу главного сборочного конвейера. Единственная переменная в данном критерии – это коэффициент простоя рабочих единиц основного производства *u*, зависящий от числа каналов *N*.

Анализ расчетных формул показывает, что эти простои с увеличением *N* уменьшаются до определенного предела. При дальнейшем увеличении *N* скорость уменьшения простоев резко снижается и после достижения некоторого граничного значения *N* практически перестает меняться. Число каналов *N*, после которого падение величины простоев практически прекращается, и выбирают в качестве оптимального значения.

Исследование качества функционирования СМО по критерию минимума себестоимости можно проводить не только при изменении числа каналов обслуживания *N*, но и при уменьшении или увеличении числа объектов обслуживания *Н*.

В этом случае себестоимость единицы продукции основного производства будет функцией от числа обслуживаемых единиц основного производства при фиксированном значении *N*:

$$C = f(H) = \frac{З(H)}{B(H)} \to \min. \qquad (2.104)$$

Очевидно, что объем выпуска продукции будет изменяться пропорционально изменению числа рабочих единиц основного производства:

$$B_1 = B_0 \frac{H_1}{H_0},$$

где H_0 и H_1 – число обслуживаемых единиц основного производства при существующем и планируемом вариантах организации производства.

Однако при изменении *Н* изменяется и качество обслуживания. В частности, меняется время нахождения требований в обслуживающей системе. Следовательно, полное изменение объема выпускаемой продукции определяется формулой

$$B_1 = B_0 \frac{H_1}{H_0} \frac{1-u_1}{1-u_0}. \qquad (2.105)$$

Условно-постоянные расходы в этом случае можно считать неизменными, т.е. $K = K_0$; переменные расходы П, как и в предыдущем случае, будут изменяться пропорционально изменению объема выпуска продукции, т.е.

$$П_1 = П_0 \frac{H_1}{H_0} \frac{1-u_1}{1-u_0}.$$

Затраты на весь объем выпускаемой продукции будут меняться следующим образом:

$$З_1 = П_1 + K_1 = П_0 \frac{H_1}{H_0} \frac{1-u_1}{1-u_0} + K_0. \qquad (2.106)$$

Подставив (2.105) и (2.106) в формулу для себестоимости, получим выражение зависимости себестоимости от числа рабочих единиц основного производства:

$$C = f(H_1) = \frac{З_1}{B_1} = \frac{П_0}{B_0} + \frac{K_0(1-u_0)H_0}{B_0(1-u_1)H_1}. \qquad (2.107)$$

Изменяя значения H_1 по отношению к существующему значению H_0 и учитывая, что коэффициент u_1 также зависит от H_1, выбираем такое значение H_1, при котором функция (2.107) достигает минимума. Это значение и следует выбрать в качестве оптимального.

Сравнив выражения (2.101) и (2.107), можно заметить, что они различаются лишь отношением $\frac{H_1}{H_0}$ и выражением $Q(N_1 - N_0)$. Поэтому если в СМО возможно одновременное изменение как количества обслуживания H, так и числа каналов обслуживания N, то ожидаемую себестоимость единицы продукции при различных вариантах организации такой системы можно определить по формуле

$$C_1 = f(H_1, N_1) = \frac{П_0}{B_0} + \frac{[K_0 + Q(N_1 - N_0)](1-u_0)H_0}{B_0(1-u_1)H_1}.$$

Отметим, что изменение количественного состава обслуживаемой и обслуживающей систем, если объектами обслуживания будут не только рабочие, но и используемые ими машины и оборудование, связано с единовременными затратами на приобретение и установку нового оборудования или с доходами от реализации излишнего оборудования.

Для полного экономического обоснования рекомендуемых для внедрения вариантов организации СМО необходимо рассчитать также и эти показатели. Для экономического обоснования единовременных затрат на приобретение оборудования может быть определена окупаемость расходов за счет снижения себестоимости единицы выпускаемой продукции $(C_0 - C_1)$:

$$Q_k = \frac{E}{(C_0 - C_1)B},$$

где Q_k – срок окупаемости нововведений; E – единовременные затраты на приобретение оборудования; B – объем выпуска продукции в течение планового периода при данном варианте организации СМО.

2.14.3. Критерий минимума экономических потерь от ожидания обслуживания

При анализе некоторых систем обслуживания необходимо учитывать, что ожидание обслуживания требований и простой обслуживающих каналов вызывают экономические потери.

Потери от ожидания обслуживания могут проявляться различным образом: штраф и пени за то, что требование не поступило вовремя в канал обслуживания; дополнительные текущие затраты; потеря качества продукции; удорожание технологии; уменьшение объема производимой продукции и т.д.

Перечисленные виды потерь могут быть измерены: простои каналов обслуживания вызывают потери в объеме выпускаемой продукции, штрафы за простои зачастую дорогостоящих каналов обслуживания, удорожание продукции и пр.

Для постановки оптимизационной задачи необходимо, чтобы все учитываемые виды потерь в процессе обслуживания были соизмеримы (т.е. измерялись в одних и тех же единицах) и сопоставимы (т.е. относились к одинаковым интервалам времени работы системы).

Необходимо определить число каналов обслуживания, при котором суммарные потери от простоев в единицу времени минимальны:

$$C = C_1 M_1 + C_2 M_2 = C_1 \sum_{k=N+1}^{H} (k-N) P_k + C_2 \sum_{k=0}^{N-1} (N-k) P_k, \qquad (2.108)$$

где C_1 – экономические потери от ожидания в очереди, приходящиеся на одно требование в единицу времени; C_2 – экономические потери от простоя каналов, приходящиеся на один канал обслуживания в единицу времени; M_1 – среднее число требований в очереди; M_2 – среднее число незанятых каналов обслуживания.

Решение задачи оптимизации сводится к вычислению P_k вероятностей по формулам (2.102) и значения C по формуле (2.108) для последовательности значений $N = 1, 2, \ldots$ и выбору значения N, соответствующего минимальному значению C. Если величины постоянны и положительны, то функция $C = f(N)$ будет унимодальной.

Пример 28. В СМО поступает простейший поток требований на обслуживание, длительность обслуживания имеет показательное распределение. Загрузка СМО в пересчете на один канал для всех вариантов неизменна, число мест для ожидания неограниченно, число каналов обслуживания $N \geq 2$. Результаты расчетов для некоторых соотношений C_1 и C_2 приведены в табл. 2.2.

Таблица 2.2

C_1, у.д.е.	C_2, у.д.е.	Суммарные потери C, у.д.е., при числе каналов N				
		2	3	4	5	6
500	50	206	122	153	200	250
1000	10	321	64	46	50	60
20	100	106	201	301	400	500

Результаты расчетов показывают, что оптимальное число каналов обслуживания при постоянстве остальных параметров системы существенно зависит от отношения $\frac{C_1}{C_2}$. Так в первом случае число оптимальных каналов равно 3, во втором – 4, в третьем – 2.

2.14.4. Критерий минимума экономических потерь с учетом отказов в обслуживании

Во многих системах экономические потери связаны с отказом клиента от обслуживания в данной системе. Причины такого отказа могут быть различны. Приведем наиболее типичные ситуации.

1. В случае занятости канала обслуживания при заказе услуги по телефону клиент после нескольких попыток соединения может отказаться от услуги вообще или попытаться заказать ее в конкурирующей фирме.

2. Недостаточное число мест для ожидания, например перед автозаправочной станцией, заставляет водителя искать другую АЗС.

3. Длинная очередь за услугой отпугивает клиентов.

Во всех перечисленных случаях плохая организация системы обслуживания приводит к потере клиентов, которые могли бы принести системе прибыль.

Как и ранее, рассмотрим систему, в которую поступает простейший поток требований, при этом длительность обслуживания имеет показательное распределение, число каналов обслуживания равно N, а число мест для ожидания – S. Проанализируем влияние N и S на вероятность отказа в обслуживании некоторого требования $P_{\text{отк}}$. В табл. 2.3 приведены значения $P_{\text{отк}}$ зависимости от N и S, причем интенсивность входящего потока и число мест для ожидания в очереди, приходящиеся на один канал обслуживания, неизменны.

Таблица 2.3

$P_{\text{отк}}$ при $N=1$					
S	0	1	2	3	4
0,25	0,2000	0,0470	0,0118	0,0029	0,0007
0,50	0,3333	0,1429	0,0667	0,0323	0,0159
1,00	0,6667	0,5914	0,5333	0,5161	0,5079
$P_{\text{отк}}$ при $N=2$					
S	0	2	4	6	8
0,50	0,0769	0,0046	0,0003	0,0001	0,0000
1,00	0,2000	0,0436	0,0105	0,0026	0,0007
2,00	0,6154	0,5246	0,5059	0,5015	0,4852

Анализ табл. 2.3 позволяет сделать ряд интересных выводов. Во-первых, наличие хотя бы одного места для ожидания в очереди позволяет значительно уменьшить $P_{\text{отк}}$, особенно при малых загрузках. Во-вторых, зависимость $P_{\text{отк}}$ от числа мест для ожидания в очереди существенно нелинейна. Особенно важна нелинейность зависимости $P_{\text{отк}}$ от числа каналов обслуживания N при неизменном числе мест для ожидания S постоянной интенсивности входящего потока, приходящихся на один канал.

В самом деле, предположим, что имеется некоторая СМО с $N = 2,\ S = 4$. Пусть λ и ν нам известны и неизменны: $\frac{\lambda}{\nu} = 1$. Если теперь уменьшить вдвое интенсивность входящего потока λ, т.е. сделать $\frac{\lambda}{\nu} = 0,5$, то казалось бы, уменьшив вдвое мощность обслуживающей системы (т.е. при $N = 1, S = 2$), мы не изменим вероятностные характеристики системы, в частности величину $P_{\text{отк}}$.

Однако из табл. 2.3 следует, что подобное изменение системы увеличит долю требований, которые получат отказ в обслуживании, более чем в 6 раз (с 0,0105 до 0,0667). Этот простой пример показывает, что к случайным процессам, протекающим в СМО, нельзя подходить, учитывая только их средние характеристики. Необходимо принимать во внимание вероятностную структуру процессов, которая существенно влияет на итоговые зависимости.

Сформулируем оптимизационную задачу так. Пусть в системе обслуживания занято N каналов и стоимость эксплуатации одного канала в единицу времени составляет C_1 денежных единиц. Тогда текущие издержки на функционирование системы в течение времени $t_{\text{ф}}$ составят $C_1 N t_{\text{ф}}$. Если среднее число простаивающих (свободных) каналов равно $C_1 N_{\text{св}} t_{\text{ф}}$, то доля текущих затрат расходуется непроизводительно.

Для того чтобы соизмерить эту величину с потерями от ожидания требований в очереди, необходимо выразить эти потери в соизмеримых единицах, т.е. ввести величину C_2 – оценку времени пребывания одного требования в очереди. Тогда совокупные потери от ожидания требования составят $C_2 T_{\text{ож}}$, где $T_{\text{ож}}$ – среднее время ожидания требования начала обслуживания.

Если доход от обслуживания равен C_3, то потери, связанные с отказами требований пройти обслуживание из-за отсутствия свободных мест для ожидания, составят $C_3 \lambda P_{\text{отк}} t_{\text{ф}}$. Таким образом, совокупные потери системы обслуживания составят

$$C = C_1 N_{\text{св}} t_{\text{ф}} + C_2 T_{\text{ож}} + C_3 \lambda P_{\text{отк}} t_{\text{ф}}. \qquad (2.109)$$

При известных значениях $\lambda, \nu, C_1, C_2, C_3$ задача обычно сводится к определению величин N и S, минимизирующих (2.109).

Пример 29. Около шоссе находится автозаправочная станция с тремя колонками. Наблюдения показали, что на нее в среднем поступает 3 автомашины в 1 мин, причем их поток – простейший. Среднее время заправки – 2 мин, причем длительность заправки – показательно (экспоненциально) распределенная случайная величина.

Около АЗС есть площадка для ожидания автомашин, на которой в очереди может находиться одновременно не более трех автомобилей. Если все места для ожидания заняты, то подъехавший для заправки автомобиль получает отказ и покидает АЗС без обслуживания.

Рядом с АЗС есть свободная площадка, на которой можно либо установить еще одну заправочную колонку либо оборудовать два дополнительных места для ожидания. Расходы на реализацию каждого из этих вариантов одинаковы.

Требуется проанализировать оба варианта реконструкции АЗС, дать их вероятностную и экономическую оценку по критерию (2.109).

Решение. Сначала рассмотрим существующий вариант организации АЗС.

Имеем: $N = 3,\ S = 3,\ \lambda = 3\ \text{мин}^{-1},\ \nu = 0{,}5\ \text{мин}^{-1}$.

Следовательно, $\rho = 3/0{,}5 = 6$.

Вероятность отсутствия на АЗС автомашин равна

$$P_0 = \left[1 + \frac{6}{1!} + \frac{6^2}{2!} + \frac{6^3(1-2^4)}{3!(1-2)}\right]^{-1} = 0{,}00177,$$

т.е. практически все время на заправке либо на площадке для ожидания находятся автомашины (время их отсутствия составляет менее 0,2 % времени работы АЗС).

Определим теперь вероятность отказа в обслуживании, т.е. долю автомашин, которые не могут быть обслужены из-за того, что в момент их поступления все места для ожидания в очереди заняты:

$$P_{\text{отк}} = \frac{6^6}{3^3 \cdot 3!} 0,00177 = 0,51.$$

Это означает, что почти 51 % всех пребывающих на АЗС автомобилей получают отказ в обслуживании и вынуждены искать другую АЗС, чтобы приобрести топливо.

Подсчитаем теперь среднюю длину очереди автомобилей, ожидающих начала обслуживания:

$$M_{\text{оч}} = \frac{6^4 \cdot 0,00177}{3 \cdot 3! \cdot (1-2)^2}(1 - 4 \cdot 2^3 + 2^4 \cdot 3) = 2,156,$$

т.е. в среднем более $\frac{2}{3}$ площади, отведенной для ожидания автомашин в очереди, занято.

Вычислим теперь среднее число автомашин, находящихся на АЗС (как в очереди, так и на обслуживании):

$$M_{\text{тр}} = M_{\text{зан}} + M_{\text{оч}} = 0,00177\left[6 + \frac{6^2}{1!} + \frac{6^3(1-2^4)}{2!(1-2)}\right] + 2,156 = 5,097.$$

Следовательно, среднее число заправочных колонок, занятых обслуживанием, составляет $M_{\text{зан}} = 2,941$, а коэффициент их загрузки равен $K_{\text{загр}} = 2,941/3 = 0,98$, т.е. каждая колонка простаивает в среднем лишь около 2 % рабочего времени.

Подсчитаем вероятность того, что все колонки одновременно заняты обслуживанием:

$$P_{\text{зан}} = 0,00177\frac{6^3}{3!} \cdot \frac{1-2^4}{1-2} = 0,956,$$

т.е. около 96 % всех пребывающих на заправку автомобилей вынуждены ожидать в очереди.

Опираясь только на анализ вероятностных характеристик, можно сделать заключение, что режим работы АЗС чрезмерно интенсивный и она не справляется с обслуживанием более половины поступающих на заправку автомобилей.

Перейдем к анализу вариантов реконструкции АЗС.

Вариант 1. Увеличиваем число бензоколонок на одну, а число мест для ожидания в очереди оставляем без изменения, т.е.

$$N = 4,\ S = 3,\ p = 6,\ \frac{p}{N} = 1,5.$$

Получаем следующие характеристики:

$$P_0 = 0,002;\ P_{\text{отк}} = 0,364;\ M_{\text{оч}} = 1,742;\ M_{\text{тр}} = 5,594;\ M_{\text{зан}} = 3,852;$$
$$K_{\text{загр}} = 0,963;\ P_{\text{зан}} = 0,911.$$

Вариант 2. Число бензоколонок оставляем без изменения, а количество мест в очереди увеличиваем на 2.

В этом варианте вероятностные характеристики таковы:

$$P_0 = 0,0004;\ P_{\text{отк}} = 0,446\ \ M_{\text{ОЧ}} = 3,598;\ M_{\text{тр}} = 6,250;\ M_{\text{зан}} = 2,652;$$
$$K_{\text{загр}} = 0,884;\ P_{\text{зан}} = 0,879.$$

Легко видеть, что первый вариант реконструкции с точки зрения вероятностных характеристик обслуживания автомобилей явно предпочтительнее.

Действительно несмотря на то, что по сравнению со вторым вариантом вероятность ожидания в очереди достаточно велика, значительно уменьшается количество автомобилей, получивших отказ в обслуживании, т.е. увеличивается пропускная способность АЗС.

Средняя очередь к одной бензоколонке $\left(\dfrac{M_{\text{оч}}}{N}\right)$ для первого варианта также значительно меньше (0,435 против 1,199).

Перейдем к сравнению трех вариантов организации обслуживания на АЗС по критерию (2.109), приняв $t_{\text{ф}} = 24$ ч, т.е. АЗС работает круглосуточно. Результаты анализа сведем в табл. 2.4, приняв $C_2 = 0$.

Таблица 2.4

C_1, у.д.е.	C_3, у.д.е.	Существующий вариант			Вариант 1			Вариант 2		
		$N_{\text{св}}$	$P_{\text{отк}}$	C	$N_{\text{св}}$	$P_{\text{отк}}$	C	$N_{\text{св}}$	$P_{\text{отк}}$	C
0,02	0,6	0,06	0,51	1322	0,15	0,36	156	0,35	0,45	1145
0,3	0,3			664,2			79			581
0,6	0,02			45			7,4			43,6

В качестве единицы стоимости в табл. 2.4 принята условная денежная единица. Из табл. 2.4 следует, что при различных соотношениях стоимостей C_1 и C_3 наибольшие экономические потери имеют место при существующем варианте организации АЗС, а наилучшим вариантом оказывается первый вариант реконструкции. Отметим, что при соотношении C_1 и C_3 наиболее соответствующем реальному, первый вариант реконструкции АЗС позволяет уменьшить экономические потери почти в 8,5 раза.

2.14.5. Пример на оптимизацию

При решении многих экономических задач используется теория систем массового обслуживания. Опираясь на теорию вероятностей и математическую статистику, эта область прикладной математики занимается анализом процессов в производстве, обслуживании, управлении, т.е. там, где однородные события повторяются многократно или происходит удовлетворение требований на выполнение каких-либо услуг. При этом часто встает вопрос о нахождении оптимального с точки зрения экономической эффективности числа каналов СМО. В процессе постановки задачи нами были рассмотрены экономические показатели СМО, которые по своей базовой принадлежности можно разделить на две группы: первая связана с издержками обращения системы, а вторая определяется издержками обслуживания заявок, поступающих на обслуживание.

Издержки обращения системы определяются числом занятых каналов, затратами на содержание СМО, интенсивностью обслуживания, степенью загрузки каналов, эффективностью их использования, пропускной способностью СМО и др. Издержки обслуживания заявок связаны с такими показателями, как длина очереди, время ожидания обслуживания, вероятность отказа в обслуживании, время пребывания заявки в СМО и др.

Эти группы показателей противоречивы в том смысле, что улучшение показателей одной группы, например сокращение длины очереди или времени ожидания в очереди путем увеличения числа каналов обслуживания, связано с ухудшением показателей другой группы, поскольку это может привести к увеличению времени простоев каналов обслуживания, затрат на их содержание и т.д. В связи с этим при формализации задач обслуживания вполне естественно стремление построить СМО таким образом, чтобы установить разумный компромисс между показателями, связанными с собственно заявками и полнотой использования возможностей системы. С этой целью необходимо выбрать показатель эффективности СМО, включающий одновременно претензии и возможности обеих групп. В качестве такого показателя был выбран минимум общих затрат.

Можно отметить, что в качестве критерия экономической эффективности в целевой функции могут быть также товарооборот, доход, прибыль, тогда оптимальное значение управляемых показателей СМО находится, очевидно, уже при максимизации указанных показателей.

Рассмотрим СМО с числом каналов n, на вход которой поступает пуассоновский поток заявок с интенсивностью λ. Длительности обслуживания отдельных требований предполагаем случайными с экспоненциальным законом распределения и интенсивностью обслуживания μ. Максимально возможное число мест в очереди ограничим величиной m.

Система может иметь следующие дискретные состояния:

E_0 – все каналы свободны;

E_1 – занят один канал (любой);

- -

E_n – заняты все n каналов;

E_{n+1} – заняты все n каналов и одна заявка стоит в очереди;

- -

E_{n+m} – заняты все n каналов и все m мест в очереди.

Граф состояний n-канальной СМО с очередью, ограниченной m местами, представлен на рис. 2.30.

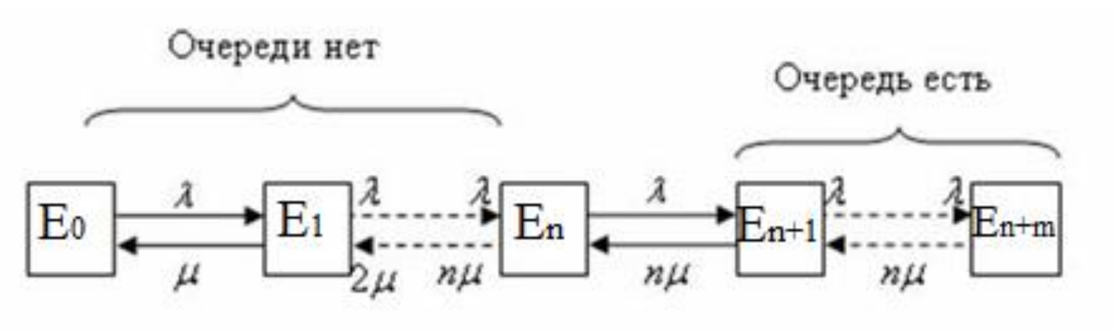


Рис. 2.30. Граф состояний n-канальной СМО с очередью длиной m

В узлах графа указаны состояния системы, дуги графа размечены плотностями вероятностей переходов.

На основании представленного графа можно составить алгебраическую систему уравнений для определения вероятностей p_i состояний системы в установившемся режиме:

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0; \\ -(\lambda + k\mu)\, p_k + \lambda p_{k-1} + (k+1)\mu p_{k+1} = 0; \;\; 1 \le k \le n-1; \\ -(\lambda + n\mu)\, p_n + \lambda p_{n-1} + n\mu p_{n+1} = 0; \\ -(\lambda + n\mu)\, p_{i+n} + \lambda p_{i+n-1} + n\mu p_{i+n+1} = 0; \;\; 1 \le i \le m-1; \\ -n\mu p_{n+m} + \lambda p_{n+m-1} = 0. \end{cases} \qquad (2.110)$$

Условие нормировки вероятностей имеет вид

$$\sum_{k=0}^{n} p_k + \sum_{s=1}^{m} p_{n+s} = 1. \qquad (2.111)$$

Решая систему уравнений, можно определить искомые вероятности, которые, в свою очередь, позволяют вычислить интересующие нас характеристики системы.

Средняя длина очереди определяется из условия, что очередь возникает, когда заняты все каналы и количество заявок в системе варьирует от (n + 1) до (n + m). Поэтому получаем

$$k_{\text{ср}} = \sum_{i=n+1}^{n+m} i p_i. \qquad (2.1112)$$

Отказ в обслуживании происходит, когда заняты все каналы и все места в очереди ожидания. Следовательно, вероятность отказа $p_{\text{отк}} = p_{n+m}$.

Вероятность потерь из-за отказа в обслуживании уменьшается с увеличением длины очереди и особенно сильно с увеличением числа обслуживающих приборов. Наличие очереди в системе позволяет снизить потери из-за отказа в обслуживании, но вызывает необходимость ожидания заявки в очереди перед началом обслуживания.

Среднее число занятых и свободных (простаивающих) каналов будет, соответственно, равно

$$\bar{n}_{\text{з}} = \frac{\lambda(1 - p_{\text{отк}})}{\mu}. \qquad (2.113)$$

$$\bar{n}_{\text{св}} = n - \bar{n}_{\text{з}}. \qquad (2.114)$$

Рассмотрим основные экономические показатели функционирования системы.

Издержки, связанные с обращением системы $C_{\text{ис}}$, включающие затраты, связанные с эксплуатацией СМО и простоем каналов обслуживания, равны

$$C_{\text{ис}} = C_{\text{пр}}\bar{n}_{\text{св}} + C_{\text{экс}}\bar{n}_{\text{з}}, \qquad (2.115)$$

где $C_{\text{пр}}$, $C_{\text{экс}}$ – издержки простоя и эксплуатации одного канала.

Можно заметить, что с увеличением числа обслуживающих каналов СМО издержки обращения системы растут.

Издержки обслуживания заявок включают потери, связанные с пребыванием заявок в очереди и с уходом необслуженных заявок:

$$C_{\text{из}} = C_{\text{оч}}k_{\text{ср}} + C_{\text{отк}}p_{\text{отк}}\lambda / \mu, \qquad (2.116)$$

где $C_{\text{оч}}$ – потери из-за простоя заявки в очереди; издержки $C_{\text{отк}}$ – потери, вследствие отказа в обслуживании (упущенная прибыль).

Издержки обслуживания заявок с увеличением числа каналов падают, поскольку сокращается число заявок в очереди и уменьшается вероятность отказа в обслуживании.

Общие издержки, которые в данной постановке задачи выступают в роли целевой функции:

$$C = C_{\text{ис}} + C_{\text{из}}. \qquad (2.117)$$

Таким образом, математически поставленную задачу можно сформулировать как минимизацию функции (2.101). В качестве варьируемого параметра СМО, подлежащего определению, было взято число каналов n при заданных интенсивностях обслуживания и потока заявок.

В качестве объекта исследования было взято предприятие, занятое перевозкой пассажиров маршрутными такси; оно предоставило некоторые экономические показатели. В частности, были использованы следующие данные. Интенсивность обслуживания пассажиров одним микроавтобусом μ = 25,3 чел./ч. Интенсивность пассажиропотока в «часы пик» λ = 500 чел./ч, в иное время λ = 85 чел./ч. Количество мест в очереди m, т.е. количество человек, которые готовы подождать следующего таксомотора, если им не удалось уехать на предыдущем, было взято равным 2.

В ходе компьютерного эксперимента получены следующие результаты.

Для случая, когда λ = 500 чел./ч, оптимальное число таксомоторов, рассчитанное по модели, равно 29. Вероятность отказа в обслуживании составит 0,5 %, средняя длина очереди – 0,02 чел., средняя загрузка машины – 70 %. Для случая, когда λ = 85 чел./ч, оптимальное число таксомоторов, рассчитанное по модели, равно 10.

Вероятность отказа в обслуживании составит 1,5 %, средняя длина очереди – 0,05 чел., средняя загрузка машины – 66 %.

Анализ результатов показывает, что система в полном объеме обслуживает все поступающие в нее требования. Однако, несмотря на минимальные издержки, она в обоих случаях загружена не полностью.

Попытка определить статистические показатели при полной загрузке микроавтобусов показала следующее. Полная загрузка системы при λ = 500 чел./ч возможна в случае использования 19 таксомоторов. Но в этом случае вероятность отказа в обслуживании возрастет до 14 %, что вызовет повышение издержек, связанных с пребыванием заявок в очереди и с уходом необслуженных клиентов. Аналогичная картина наблюдается и при λ = 85 чел./ч, когда полная загрузка системы происходит для n = 6. Здесь вероятность отказа составит уже 22 %.

В результате проведенного исследования можно сделать выводы. Применение элементов теории массового обслуживания в сочетании с экономическими критериями может позволить снизить затраты предприятия путем оптимизации структуры производства.

ЗАДАЧИ

Многоканальная СМО с отказами в обслуживании

1. АТС имеет k линий связи. Поток вызовов – простейший с интенсивностью λ в 1 мин. Среднее время переговоров составляет t мин. Время переговоров имеет показательное распределение.

Найти:

а) вероятность того, что все линии связи заняты;

б) относительную и абсолютную пропускные способности АТС;

в) среднее число занятых линий связи.

Определить оптимальное число линий связи, достаточное для того, чтобы вероятность отказа не превышала α; $k = 5$; $\lambda = 0,6$; $t = 3,5$; $\alpha = 0,04$.

Решение. Исчисляем показатели обслуживания многоканальной СМО.

Интенсивность потока обслуживания:

$$\mu = 1/3,5 = 0,29.$$

1. Интенсивность нагрузки:

$$y = \lambda \cdot t_{\text{обс}} = 0,6 \cdot 3,5 = 2,1.$$

Интенсивность нагрузки $y = 2,1$ показывает степень согласованности входного и выходного потоков заявок канала обслуживания и определяет устойчивость системы массового обслуживания.

2. Вероятность, что канал свободен (доля времени простоя каналов):

$$p_0 = \frac{1}{1 + \frac{2,1}{1!} + \frac{2,1^2}{2!} + \frac{2,1^3}{3!} + \frac{2,1^4}{4!} + \frac{2,1^5}{5!}} = 0,13.$$

Следовательно, 13 % в течение 1 ч канал не будет занят, время простоя равно $t_{\text{пр}} = 7,8$ мин. Вероятность того, что обслуживанием:

– занят 1 канал:

$$p_1 = \frac{y^1}{1!} \cdot p_0 = \frac{2,1}{1!} \cdot 0,13 = 0,26;$$

– заняты 2 канала:

$$p_2 = \frac{y^2}{2!} \cdot p_0 = \frac{2,1^2}{2!} \cdot 0,13 = 0,28;$$

– заняты 3 канала:

$$p_3 = \frac{y^3}{3!} \cdot p_0 = \frac{2,1^3}{3!} \cdot 0,13 = 0,19;$$

– заняты 4 канала:

$$p_4 = \frac{y^4}{4!} \cdot p_0 = \frac{2,1^4}{4!} \cdot 0,13 = 0,1;$$

– заняты 5 каналов:

$$p_5 = \frac{y^5}{5!} \cdot p_0 = \frac{2,1^5}{5!} \cdot 0,13 = 0,0425 \text{ –}$$

вероятность того, что все линии связи заняты.

3. Доля заявок, получивших отказ:

$$P_{\text{отк}} = \frac{y^n}{n!} p_0 = \frac{2,1^5}{5!} \cdot 0,13 = 0,0425.$$

Значит, 4 % из числа поступивших заявок не принимаются к обслуживанию.

4. Вероятность обслуживания поступающих заявок. В системах с отказами события отказа и обслуживания составляют полную группу событий, поэтому $P_{\text{отк}} + P_{\text{обс}} = 1$. Относительная пропускная способность: $Q = P_{\text{обс}} = 1 - P_{\text{отк}} = 1 - 0,0425 = 0,96$.

Следовательно, 96 % из числа поступивших заявок будут обслужены. Приемлемый уровень обслуживания должен быть выше 90 %.

5. Среднее число занятых линий связи:

$$n_{\text{з}} = y \cdot P_{\text{обс}} = 2,1 \cdot 0,96 = 2,01 \text{ линии.}$$

6. Среднее число простаивающих каналов:

$$n_{\text{пр}} = n - n_{\text{з}} = 5 - 2,01 = 3 \text{ канала.}$$

7. Коэффициент занятости каналов обслуживанием:

$$K_3 = n_3 / n = 2,01 / 5 = 0,4.$$

Следовательно, система на 40 % занята обслуживанием.

8. Абсолютная пропускная способность:

$$A = P_{\text{обс}} \cdot \lambda = 0,96 \cdot 0,6 = 0,57 \text{ заявок/мин.}$$

9. Среднее время простоя СМО:

$$t_{\text{пр}} = P_{\text{отк}} \cdot t_{\text{обс}} = 0,0425 \cdot 3,5 = 0,15 \text{ мин.}$$

10. Среднее число обслуживаемых заявок:

$$L_{\text{обс}} = y \cdot Q = 2,1 \cdot 0,96 = 2,01 \text{ ед.}$$

Для определения оптимального числа линий связи, достаточного для того, чтобы вероятность отказа не превышала 0,04, воспользуемся формулой

$$P_{отк} = \frac{y^n}{n!} p_0 .$$

Для наших данных получаем: $0,04 = \dfrac{2,1^n}{n!} p_0$,

где $p_0 = \sum\limits_{k=0}^{5} \dfrac{y^k}{k!}$.

Подбирая количество линий связей, находим, что при $k = 6$

$$P_{\text{отк}} = 0,0147 < 0,04, \; p_0 = 0,12.$$

2. Коммерческая фирма занимается посреднической деятельностью по продаже автомобилей и осуществляет часть переговоров по трем телефонным линиям. В среднем поступает 75 звонков/ч. Среднее время предварительных переговоров справочного характера составляет 2 мин.

Найти:

а) вероятность того, что все линии связи заняты;

б) относительную и абсолютную пропускные способности АТС;

в) среднее число занятых линий связи.

Рекомендации к решению задачи: здесь $n = 3$; $\lambda = 75$ ед. в час.; $t = 2$ мин или $\mu = 30$ ед. в 1 ч.

3. Пункт по ремонту квартир работает в режиме отказа и состоит из двух бригад. Интенсивность потока заявок $\lambda = 1,5$ ед./ч, производительность пункта $\mu = 1,8$ ед./ч. Определить вероятность того, что оба канала свободны, один канал занят, оба канала заняты, вероятность отказа, относительную и абсолютную пропускные способности, средне число занятых бригад.

Рекомендации к решению задачи: здесь $n = 2$; $\lambda = 1{,}5$; $\mu = 1{,}8$.

4. В вычислительный центр коллективного пользования с тремя ЭВМ поступают заказы от предприятий на вычислительные работы. Если работают все три ЭВМ, то вновь поступающий заказ не принимается, и предприятие вынуждено обратиться в другой вычислительный центр. Среднее время работы с одним заказом составляет 3 ч. Интенсивность потока заявок 0,25 (1/ч). Найти предельные вероятности состояний и показатели эффективности работы вычислительного центра.

Рекомендации к решению задачи: здесь $n = 3$; $\lambda = 0{,}25$ ед./ч; $t_{обс} = 3$ ч.

Многоканальная СМО с ограниченной длиной очереди

5. В мини-маркет поступает поток покупателей с интенсивностью 6 покупателей в 1 мин, которых обслуживают три контролера-кассира с интенсивностью 2 покупателя в 1 мин, длина очереди ограничена 5 покупателями. Построить две модели многоканальной системы массового обслуживания – с бесконечной и ограниченной очередью. Вычислить p_0 – вероятность простаивания всех каналов обслуживания; $L_{оч}$ – среднее число клиентов, ожидающих обслуживания; $T_{оч}$ – среднее время ожидания обслуживания; $P_{оч}$ – вероятность обязательного пребывания в очереди.

Рекомендации к решению задачи: здесь $n = 3$; $m = 5$; $\lambda = 6$ ед./мин; $\mu = 2$ ед./мин.

6. На плодоовощную базу в среднем через 30 мин прибывают автомашины с плодоовощной продукцией. Среднее время разгрузки одной машины составляет 1,5 ч. Разгрузку производят две бригады. На территории базы у дебаркадера могут находиться в очереди в ожидании разгрузки не более 4 автомашин.

Построить две модели многоканальной системы массового обслуживания – с бесконечной и ограниченной очередью. Вычислить p_0 – вероятность простаивания всех каналов обслуживания; $L_{оч}$ – среднее число клиентов, ожидающих обслуживания; $T_{оч}$ – среднее время ожидания обслуживания; $P_{оч}$ – вероятность обязательного пребывания в очереди.

Рекомендации к решению задачи: здесь $n = 2$; $m = 4$; $\lambda = 2$ ед. /ч; $\mu = 2/3 = 0{,}67$ ед./ч.

7. На автомойку в среднем за 1 ч приезжают 9 автомобилей, но, если в очереди уже находятся 4 автомобиля, вновь подъезжающие клиенты, как правило, не встают в очередь, а проезжают мимо. Среднее время мойки автомобиля составляет 20 мин, а мест для мойки всего два. Средняя стоимость мойки автомобиля составляет 70 руб. Определить среднюю величину потери выручки автомойки в течение дня.

Рекомендации к решению задачи: здесь $n = 2$; $r = 4$; $\lambda = 9$ ед./ч; $t_{\text{обс}} = 20$ мин. Величина потери выручки: $S = t_{\text{время работы мойки за день}} \cdot \lambda \cdot P_{\text{отк}} \cdot 70$ руб. (ответ: 5443,2 руб.)

8. Магазин получает овощи из теплиц. Автомобили с грузом прибывают с интенсивностью λ машин в день. Подсобные помещения позволяют обрабатывать и хранить товар, привезенный r автомобилями. В магазине работают n фасовщиков, каждый из которых в среднем может обрабатывать товар с одной машины в течение $t_{\text{обс}}$ [ч]. Продолжительность рабочего дня при сменной работе составляет 12 ч. Определить емкость подсобных помещений при заданной вероятности $P^*_{\text{обс}}$ полной обработки товаров.

9. Имеется автозаправочная станция с двумя колонками. В очереди не может быть больше трех машин. Интенсивность и среднее время заправки равны 2,1 и 0,55. Найти вероятность простоя системы.

Решение. Интенсивность потока обслуживания равна $\mu = 1/0{,}55 = 1{,}82$. Отсюда, интенсивность нагрузки составит $y = \lambda \cdot t_{\text{обс}} = 2{,}1 \cdot 0{,}55 = 1{,}16$. Заметим, что интенсивность нагрузки $y = 1{,}16$ показывает степень согласованности входного и выходного потоков заявок канала обслуживания и определяет устойчивость системы массового обслуживания. Поскольку 1,16 < 2, то процесс обслуживания будет стабилен. Вероятность простоя системы выражается следующей формулой:

$$p_0 = \frac{1}{\sum_{k=0}^{n} \frac{y^k}{k!} + \frac{y^{n+1}}{n!(n-y)}\left(1 - \left(\frac{y}{n}\right)^m\right)} =$$

$$= \frac{1}{\frac{1{,}16^0}{0!} + \frac{1{,}16^1}{1!} + \frac{1{,}16^2}{2!} + \frac{1{,}16^{2+1}}{2!(2-1{,}16)}\left(1 - \left(\frac{1{,}16}{2}\right)^3\right)} = 0{,}28.$$

Следовательно, 28 % времени в течение 1 ч канал будет не занят, время простоя равно $t_{\text{пр}} = 0,28 \cdot 60\,\text{мин} = 16,9$ мин.

10. В типографию с тремя множительными аппаратами поступают заказы от соседних предприятий на размножение рабочей документации. Если все аппараты заняты, то вновь поступающий заказ не принимается. Среднее время работы с одним заказом составляет 2 ч. Интенсивность потока – 0,5 заявки в 1 ч. Найти предельные вероятности состояний и показатели эффективности работы типографии. Здесь: $n = 3$, модель СМО: многоканальная СМО с отказами в обслуживании, $\lambda = 0,5$ заявки в 1 ч, $\overline{t}_{\text{об}} = 2$ ч.

Многоканальная СМО с неограниченной очередью

11. Построить две модели многоканальной системы массового обслуживания – с бесконечной и ограниченной очередью. Вычислить P_0 – вероятность простаивания всех каналов обслуживания, $L_{\text{оч}}$ – среднее число клиентов, ожидающих обслуживания, $T_{\text{оч}}$ – среднее время ожидания обслуживания, $P_{\text{оч}}$ – вероятность обязательного пребывания в очереди.

12. В расчетном узле магазина самообслуживания работают 3 кассы. Интенсивность входного потока составляет 5 покупателей в 1 мин. Интенсивность обслуживания каждого контролера-кассира составляет 2 покупателя в 1 мин.

Рекомендации к решению задачи: здесь $n = 3$; $\lambda = 5$ ед./мин; $\mu = 2$ ед./мин. В качестве количества заявок в очереди можно указать, например, $r = 4$. Тогда будут рассчитаны соответствующие вероятности появления данных заявок.

13. В аудиторскую фирму поступает простейший поток заявок на обслуживание с интенсивностью $\lambda = 1,5$ заявки в день. Время обслуживания распределено по показательному закону и равно в среднем трем дням. Аудиторская фирма располагает пятью независимыми бухгалтерами, выполняющими аудиторские проверки (обслуживание заявок). Очередь заявок не ограничена. Дисциплина очереди не регламентирована. Определить вероятностные характеристики аудиторской фирмы как системы массового обслуживания, работающей в стационарном режиме.

Рекомендации к решению задачи: здесь $n = 5$; $\lambda = 1.5$ ед./ч; $t_{\text{обс}} = 3$ ед./ч. После решения необходимо заменить единицы измерения «час» на «дни».

14. В мастерской по ремонту холодильников работает 3 (n) мастера. В среднем в течение дня поступает в ремонт 6 (λ) холодильников. Поток заявок пуассоновский. Время ремонта подчиняется экспоненциальному закону распределения вероятностей, в среднем в течение дня при семичасовом рабочем дне каждый из мастеров ремонтирует 2 (μ) холодильника.

Требуется определить:

1) вероятность того, что все мастера свободны от ремонта холодильников;

2) вероятность того, что все мастера заняты ремонтом;

3) среднее время ремонта одного холодильника;

4) среднее время ожидания начала ремонта для каждого холодильника;

5) среднюю длину очереди, которая определяет необходимое место для хранения холодильника, требующего ремонта;

6) среднее число мастеров, свободных от работы.

15. Определить, сколько автомобилей следует иметь на станции скорой помощи, если:

1) в среднем в 1 ч поступает 10 заявок (вызовов);

2) средняя продолжительность обслуживания одной заявки 1 ч 20 мин;

3) среднее время ожидания (время от момента вызова до момента выезда бригады) не должно превышать 5 мин.

16. Железнодорожная касса с двумя окошками продает билеты в два пункта A и B. Интенсивность потока пассажиров, желающих купить билеты, для обоих пунктов одинакова: $\lambda_A = \lambda_B = 0{,}45$ (пассажиров в 1 мин). На обслуживание пассажиров кассир тратит в среднем 2 мин. Рассматриваются два варианта продажи билетов: первый – билеты продаются в одной кассе с двумя окошками одновременно в оба пункта A и B, второй – билеты продаются в двух специализированных кассах (по одному окошку в каждой), одна только в пункт A, другая – только в пункт B. Необходимо:

а) сравнить два варианта продажи билетов по основным характеристикам обслуживания;

б) определить, как надо изменить среднее время обслуживания одного пассажира, чтобы по второму варианту продажи пассажиры затрачивали на приобретение билетов в среднем меньше времени, чем по первому варианту.

Решение.

а. *По первому варианту* имеем двухканальную СМО, на которую поступает поток заявок с интенсивностью $\lambda = 0{,}45 + 0{,}45 = 0{,}9$; интенсивность потока обслуживания $\mu = 1/2 = 0{,}5$; $y = \lambda/\mu = 1{,}8$. Так как $y/n = 1{,}8/2 = 0{,}9 < 1$, то предельные вероятности существуют.

Вероятность простоя двух кассиров по формуле

$$p_0 = \left(1 + \frac{y}{1!} + \frac{y^2}{2!} + \ldots + \frac{y^n}{n!} + \frac{y^{n+1}}{n!(n-y)}\right)^{-1};$$

равна

$$p_0 = \left(1 + \frac{1{,}8}{1!} + \frac{1{,}8^2}{2!} +, \frac{1.8^3}{2!(2-1{,}8)}\right)^{-1} \approx 0{,}0526.$$

Среднее число пассажиров в очереди (среднее число заявок в очереди) по формуле

$$L_{\text{оч}} = \frac{y^{n+1} p_0}{n!n}\left(1 - \frac{y}{n}\right)^{-2};$$

равна

$$L_{\text{оч}} = \frac{1{,}8^3 \cdot 0{,}0526}{2!2}\left(1 - \frac{1{,}8}{2}\right)^{-2} = 7{,}67.$$

Среднее число пассажиров у кассы $(L_{\text{сист}} = L_{\text{оч}} + y)$:

$$L_{\text{сист}} = 7{,}67 + 1{,}8 = 9{,}47.$$

Среднее время на ожидание в очереди и покупку билетов вычисляется соответственно по формулам:

$$\left(T_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda}\right) \text{ и } \left(T_{\text{сист}} = \frac{L_{\text{сист}}}{\lambda}\right),$$

получаем:

$$T_{\text{оч}} = \frac{7{,}67}{0{,}9} = 8{,}52 \text{ мин и } T_{\text{сист}} = \frac{9{,}47}{0{,}9} = 10{,}5 \text{ мин.}$$

По второму варианту имеем две одноканальные СМО (два специализированных окошка); на каждую поступает поток заявок с интенсивностью $\lambda = 0{,}45$. По-прежнему $\mu = 0{,}5$, $y = \lambda/\mu = 0{,}9 < 1$, предельные вероятности существуют. По формулам

$$L_{\text{оч}} = \frac{y^2}{1-y};\ L_{\text{сист}} = \frac{y}{1-y};\ T_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda};\ T_{\text{сист}} = \frac{L_{\text{сист}}}{\lambda};$$

получаем

$$L_{\text{оч}} = \frac{0,9^2}{1-0,9} = 8,1;\; L_{\text{сист}} = \frac{0,9}{1-0,9} = 9;$$

$$T_{\text{оч}} = \frac{8,1}{0,45} = 18;\; T_{\text{сист}} = \frac{9}{0,45} = 20.$$

Итак, по второму варианту увеличились и длина очереди, и среднее время ожидания в ней, и в целом время на покупку билетов. Такое различие объясняется тем, что в первом варианте (двухканальная СМО) меньше средняя доля времени, которую простаивает каждый из двух кассиров: если он не занят обслуживанием пассажира, покупающего билет в пункт *A*, он может заняться обслуживанием пассажира, покупающего билет в пункт *B*, и наоборот. Во втором варианте такой взаимозаменяемости нет.

Можно заметить, что среднее время на покупку билетов по второму варианту увеличилось более чем в 2 раза. Такое значительное увеличение связано с тем, что СМО работает на пределе своих возможностей y = 0,9: достаточно незначительно увеличить среднее время обслуживания $\bar{t}_{\text{об}}$, т.е. уменьшить μ, и y превзойдет 1, т.е. очередь начнет неограниченно возрастать.

б. Выше было получено, что по первому варианту продажи билетов при среднем времени обслуживания одного пассажира $\bar{t}_{\text{об}} = b = 2$ мин среднее время на покупку билетов составит $T_{\text{сист.1}} = 10,5$ мин. По условию для второго варианта продажи $T_{\text{сист.2}} < T_{\text{сист.1}}$, или $\frac{1}{\lambda} \cdot \frac{y}{1-y} < T_{\text{сист.1}}$.

Полагая $y = \frac{\lambda}{\mu} = \lambda\, \bar{t}_{\text{об}}$, получим $\frac{\bar{t}_{\text{об}}}{1-\lambda\, \bar{t}_{\text{об}}} < T_{\text{сист.1}}$, откуда найдем $\bar{t}_{\text{об}} < \frac{T_{\text{сист.1}}}{1+\lambda T_{\text{сист.1}}}$ или $\bar{t}_{\text{об}} < \frac{10,5}{1+0,45 \cdot 10,5} = 1,83$ мин.

Итак, средние затраты времени на покупку билетов по второму варианту продажи уменьшатся, если среднее время обслуживания одного пассажира уменьшится более чем на 0,17 мин, или более чем на 8,5 %.

17. В порту имеется один причал для разгрузки судов. Интенсивность потока судов равна 0,4 судов в сутки. Среднее время разгрузки одного судна составляет 2 сут. Предполагается, что очередь может быть неограниченной длины. Найти показатели эффективности работы причала, а также вероятность того, что ожидают разгрузки не более чем 2 судна.

Решение. Имеем $y = \frac{\lambda}{\mu} = \lambda$, $\bar{t}_{об} = 0,4 \cdot 2 = 0,8$. Так как $y = 0,8 < 1$, то очередь на разгрузку не может бесконечно возрастать и предельные вероятности существуют. Найдем их.

Вероятность того, что причал свободен: $p_0 = 1 - 0,8 = 0,2$, вероятность того, что он занят: $P_{зан} = 1 - 0,2 = 0,8$. Находим вероятности того, что у причала находятся 1, 2, 3 судна (т.е. ожидают разгрузки 0, 1, 2 судна):

$$p_1 = 0,8(1 - 0,8) = 0,16; \quad p_2 = 0,16 \cdot 0,8 = 0,128;$$
$$p_3 = 0,128 \cdot 0,8 = 0,1024.$$

Вероятность того, что ожидают разгрузку не более чем 2 судна, равна

$$P = p_1 + p_2 + p_3 = 0,16 + 0,128 + 0,1024 = 0,3904.$$

Среднее число судов, ожидающих разгрузки:

$$L_{оч} = \frac{0,8 \cdot 0,8}{1 - 0,8} = 3,2.$$

Среднее время ожидания разгрузки:

$$T_{оч} = \frac{3,2}{0,8} = 4 \text{ сут.}$$

Среднее число судов, находящихся у причала:

$$L_{сист} = \frac{y}{1 - y} = \frac{0,8}{1 - 0,8} = 4 \text{ сут.}$$

Среднее время пребывания судна у причала:

$$T_{сист} = \frac{L_{сист}}{\lambda} = \frac{4}{0,8} = 5 \text{ сут.}$$

Очевидно, что эффективность разгрузки судов невысокая. Для ее повышения необходимо уменьшение среднего времени разгрузки судна $\bar{t}_{об}$ либо увеличение числа n причалов.

18. Дисплейный зал имеет 5 дисплеев. Поток пользователей простейший. Среднее число пользователей, посещающих дисплейный зал за сутки, равно 140. Время обработки информации одним пользователем на одном дисплее распределено по показательному закону и составляет в среднем 40 мин. Определить, существует ли стационарный режим работы зала; вероятность того, что пользователь застанет все дисплеи занятыми; среднее число пользователей в дисплейном зале; среднее число пользователей в очереди; среднее время ожидания свободного дисплея; среднее время пребывания пользователя в дисплейном зале.

Решение. Рассматриваемая в задаче СМО относится к классу многоканальных систем с неограниченной очередью. Число каналов k = 5. Найдем λ – интенсивность потока заявок (среднее время между двумя последовательными заявками входящего потока пользователей):

$$\lambda = \frac{1}{M(T)},$$

где $M(T) = \frac{24}{140} \approx 0{,}17$.

Тогда $\lambda = \frac{1}{0{,}17} \approx 5{,}85$ польз./ч.

Найдем μ – интенсивность потока обслуживания (среднее время обслуживания одного пользователя одним дисплеем):

$$\mu = \frac{1}{M[T_{\text{обс}}]},$$

где $M\left[T_{\text{обс}}\right] = 40$ мин = 0,67 ч.

Тогда $\mu = \frac{1}{0{,}67} \approx 1{,}49$ польз./ч.

Таким образом, классификатор данной системы имеет вид СМО: 5, ∞; 5,85; 1,49.

Вычислим коэффициент загрузки СМО $y = \frac{\lambda}{\mu} \approx \frac{5{,}85}{1{,}49} \approx 3{,}93.$ Известно, что для СМО такого класса стационарный режим существует, если отношение коэффициента нагрузки системы к числу каналов меньше единицы. Находим это отношение:

$$x = \frac{y}{k} = \frac{3{,}93}{5} \approx 0{,}79 < 1.$$

Следовательно, стационарный режим существует. Предельное распределение вероятностей состояний вычисляется по формулам:

$$p_0 = \left[1 + \frac{y}{1!} + \frac{y^2}{2!} + \ldots + \frac{y^k}{k!} + \frac{y^{k+1}}{k \cdot k!} \cdot \frac{1}{1-x}\right]^{-1},$$

$$p_i = \frac{y^i}{i!} \cdot p_0, i \in \{1, \ldots, k\}, \quad p_{k+r} = \frac{y^{k+r}}{k^r \cdot k!} p_0, \; r \geq 1.$$

Поскольку $k = 5$, имеем

$$p_0 = \left[1 + \frac{y}{1!} + \frac{y^2}{2!} + \frac{y^3}{3!} + \frac{y^4}{4!} + \frac{y^5}{5!} + \frac{y^{5+1}}{5 \cdot 5!} \cdot \frac{1}{1-0,79}\right]^{-1} = 69,76^{-1} = 0,014.$$

Вычислим P^* – вероятность того, что пользователь застанет все дисплеи занятыми. Очевидно, она равна сумме вероятностей таких событий: все дисплеи заняты, очереди нет (p_5); все дисплеи заняты, один пользователь в очереди (p_6); все дисплеи заняты, два пользователя в очереди (p_7) и т.д. Поскольку для полной группы событий сумма вероятностей этих событий равна единице, то справедливо равенство

$$P^* = p_5 + p_6 + p_7 + \ldots = 1 - p_0 - p_1 - p_2 - p_3 - p_4.$$

Найдем эти вероятности:

$$p_0 = 0{,}014;\ p_1 = 3{,}93 \cdot 0{,}014;\ p_2 = 7{,}72 \cdot 0{,}014;$$
$$p_3 = 10{,}12 \cdot 0{,}014;\ p_4 = 9{,}94 \cdot 0{,}014.$$

Выносся за скобки общий множитель, получим:

$$P^* = 1 - 0,0148 \cdot \left(1 + 3,93 + 7,72 + 10,12 + 9,94\right) =$$
$$= 1 - 0,014 \cdot 32,71 = 1 - 0,46 = 0,54.$$

Используя формулы для вычисления показателей эффективности, найдем:

1) среднее число пользователей в очереди:

$$\bar{r} = \frac{y^{k+1}}{k \cdot k!} \cdot \frac{p_0}{(1-x)^2} = \left(3,93\right)^6 \cdot \frac{0.014}{5 \cdot 5!(1-0,79)^2} \approx 1,95 \text{ польз.};$$

2) среднее число пользователей в дисплейном зале:

$$\bar{k} = \bar{r} + y \approx 1,95 + 3,93 = 5,88 \text{ польз.};$$

3) среднее время ожидания свободного дисплея:

$$T_{\text{оч}} = \frac{\bar{r}}{\lambda} \approx \frac{1,95}{5,85} \approx 0,33 \text{ ч} \approx 20 \text{ мин};$$

4) среднее время пребывания пользователя в дисплейном зале:

$$T_{\text{сис}} = \frac{\bar{k}}{\lambda} \approx \frac{5,88}{5,85} \approx 1,01 \text{ ч} \approx 60,3 \text{ мин}.$$

Ответ: стационарный режим работы дисплейного зала существует и характеризуется следующими показателями $P^* = 0{,}54$; $\overline{r} = 1{,}95$ пользователя; $\overline{k} = 5{,}88$ пользователя; $T_{\text{оч}} = 20$ мин; $T_{\text{сис}} = 1$ ч.

При моделировании случайных явлений методом Монте-Карло случайностью пользуются как аппаратом исследования.

19. В двухканальную систему массового обслуживания с отказами поступает стационарный пуассоновский поток заявок. Время между поступлениями двух последовательных заявок распределено по показательному закону с параметром λ = 5 заявок в 1 мин. Длительность обслуживания каждой заявки равна 0,5 мин. Методом Монте-Карло найти среднее число обслуженных заявок за время 4 мин. *Указание*: провести три испытания.

Для решения этой задачи необходим метод Монте-Карло, или метод статистических испытаний. При реализации этого метода не требуется проведения реальных опытов. Для этой цели служат специальные компьютерные программы или датчики случайных чисел.

Применение метода Монте-Карло в системах массового обслуживания состоит в том, что вместо аналитического описания СМО проводится «розыгрыш» случайного процесса, проходящего в СМО, с помощью специально организованной процедуры. В результате такого «розыгрыша» получается каждый раз новая, отличная от других реализация случайного процесса. Это множество реализаций можно использовать как некий искусственно полученный статистический материал, который обрабатывается методами математической статистики. После такой обработки могут быть получены приближенно любые характеристики обслуживания.

Решение. Изобразим статистическое моделирование работы заданной СМО с помощью временных диаграмм (рис. 2.31). Введем следующие обозначения для временных осей: *Вх* – входящий поток заявок; t_i – моменты поступления заявок; T_i – интервалы времени между двумя последовательными заявками (очевидно, что $t_i = t_{i-1} + T_i$; К1 – первый канал обслуживания; К2 – второй канал обслуживания). На рис. 2.31 жирные линии на временной оси обозначают интервалы занятости канала. Если оба канала свободны, то заявка становится под обслуживание в канал К1, в случае его занятости заявка обслуживается каналом К2.

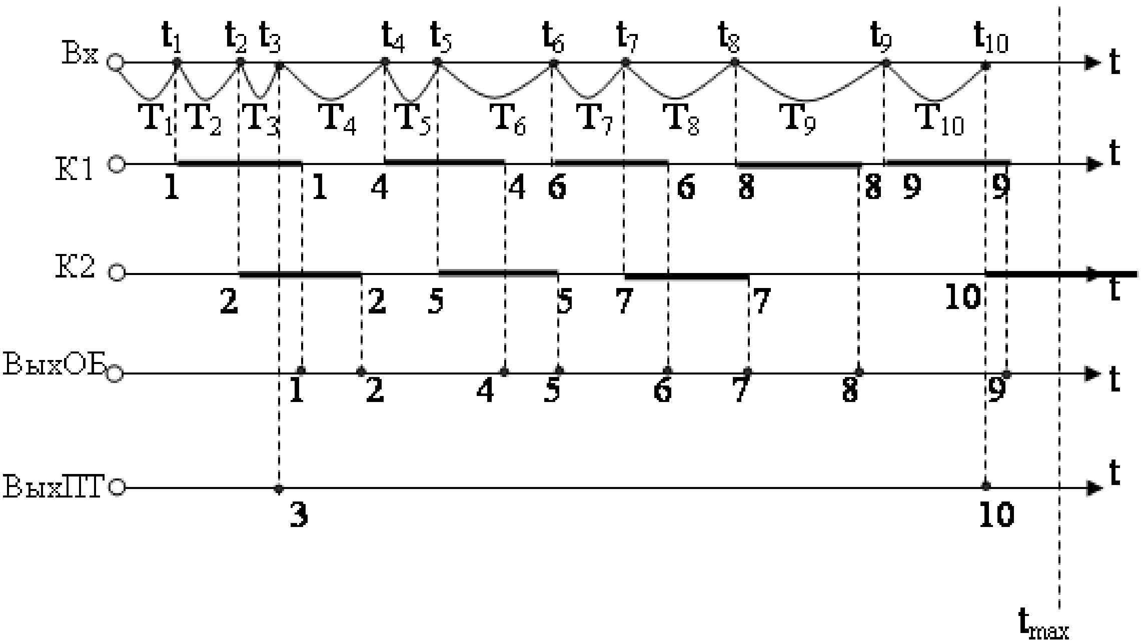


Рис. 2.31. Временная диаграмма статистического моделирования

Если заняты оба канала, то заявка покидает СМО необслуженной.

ВыхОБ – выходящий поток обслуженных заявок.

ВыхПТ – выходящий поток потерянных заявок за счет отказов СМО (случай занятости обоих каналов).

Статистические испытания продолжаются в течение временного интервала $[0;t_{\max}]$. Очевидно, что любое превышение времени $t_{\max}$ влечет за собой сброс заявки в выходящий поток ВыхПТ. Так, на рис. 2.31 заявка № 10, пришедшая в систему в момент t_{10}, не успевает обслужиться до момента $t_{\max}$, так как $t_{10} + T_{\text{обс}} > t_{\max}$. Следовательно, она не принимается свободным каналом К2 на обслуживание и сбрасывается в ВыхПТ, получая отказ.

Из временных диаграмм видно, что необходимо научиться моделировать интервалы T_i. Применим метод обратных функций. Поскольку случайная величина T_i распределена по показательному закону с параметром λ = 5, то плотность распределения имеет вид $f(\tau) = 5e^{-5\tau}$. Тогда значение $F(T_i)$ функции распределения вероятностей определяется интегралом

$$F(T_i) = \int_0^T f(\tau)d\tau = 5\int_0^T e^{-5\tau}d\tau = 5\left(-\frac{1}{5}\right)e^{-5x} \Big|_0^T = 1 - e^{-5T}.$$

Известно, что область значений функции распределения $F(T)$ есть отрезок $\left[0;1\right]$. Выбираем из таблицы случайных чисел число r_i^+, определяем T_i из равенства $1-e^{-5T_i}=r_i^+$, откуда $e^{-5T_i}=1-r_i^+$. Однако, если $r_i^+\in(0;1]$, то $(1-r_i^+)\in[0;1]$. Поэтому можно сразу получать из таблицы случайных чисел реализации $r_i=1-r_i^+$. Следовательно $e^{-5T_i}=r_i$, или $-5T_i=\ln(r_i)$, откуда $T_i=-\frac{1}{5}\ln\left(r_i\right)$. Результаты вычислений удобно заносить в таблицу.

Для проведения испытания 1 были взяты случайные числа из табл. 2.5, начиная с первого числа первой строки. Далее выборка осуществлялась по строкам. Проведем еще два испытания (табл. 2.6, 2.7).

Таблица 2.5

Испытание 1

№ заявки i	Сл. число r_i	$-\ln(r_i)$	$T_i=-\frac{1}{5}\ln(r_i)$	Момент поступления заявки $t_i=t_{i-1}+T_i$	Момент окончания обслужив. $t_i+0,50$		Счетчик заявок	
					К1	К2	Обсл.	Потери
1	0,10	2,30	0,46	0,46	0,96		1	
2	0,09	2,41	0,48	0,94		1,44	1	
3	0,73	0,31	0,06	1,00	1,50		1	
4	0,25	1,39	0,28	1,28				1
5	0,33	1,11	0,22	1,50	2,00		1	
6	0,76	0,27	0,05	1,55		2,05	1	
7	0,52	0,65	0,13	1,68				1
8	0,01	4,61	0,92	2,60	3,10		1	
9	0,35	1,05	0,21	2,81		3,31	1	
10	0,86	0,15	0,03	2,84				1
11	0,34	1,08	0,22	3,06				1
12	067	0,40	0,08	3,14	3,64		1	
13	0,35	1,05	0,21	3,35		3,85	1	
14	0,48	0,73	0,15	3,50				1
15	0,76	0,27	0,05	3,55				1
16	0,37	0,99	0,20	3,75	4,25	STOP		
						$\sum$	9	

Обратите внимание на выборку случайных чисел из табл. 2.5, если в испытании 1 последнее случайное число для заявки 16 было 0,37, то испытание 2 начинается со следующего за ним случайного числа 0,54. Испытание 2 содержит последним ее случайное число 0,53. Следовательно, третье испытание начнется с числа 0,19. Вообще, в пределах одной серии испытаний случайные числа из таблицы выбираются без пропусков и вставок по определенному порядку, например по строкам.

Таблица 2.6

Испытание 2

№ заявки i	Сл. число r_i	$-\ln(r_i)$	$T_i = -\frac{1}{5}\ln(r_i)$	Момент поступления заявки $t_i = t_{i-1} + T_i$	Момент окончания обслужив. $t_i + 0,50$		Счетчик заявок	
					К1	К2	Обсл.	Потери
1	0,54	062	0,12	0,12	0,62		1	
2	0,20	1,61	0,32	0,44		0,94	1	
3	0,48	0,73	0,15	0,59				1
4	0,05	3,00	0,60	1,19	1,69		1	
5	0,64	0,45	0,09	1,28		1,78	1	
6	0,89	0,12	0,02	1,30				1
7	0,47	0,76	0,15	1,45				1
8	0,42	0,87	0,17	1,62				1
9	0,96	0,04	0,01	1,63				1
10	0,24	1,43	0,29	1,92	2,42		1	
11	0,80	0,22	0,04	1,96		2,46	1	
12	0,52	0,65	0,13	2,09				1
13	0,40	0,92	0,18	2,27				1
14	0,37	0,99	0,20	2,47	2,97		1	
15	0,08	2,53	0,51	2,98	3,48		1	
16	0,42	0,87	0,17	3,15		3,65	1	
17	0,26	1,35	0,27	3,42				1
18	0,89	0,12	0,02	3,44				1
19	0,53	0,63	0,13	3,57	4,07	STOP		
						$\sum$	9	

Таблица 2.7

Испытание 3

№ заявки i	Сл. число r_i	$-\ln(r_i)$	$T_i = -\frac{1}{5}\ln(r_i)$	Момент поступления заявки $t_i = t_{i-1} + T_i$	Момент окончания обслужив. $t_i + 0{,}50$		Счетчик заявок	
					К1	К2	Обсл.	Потер
1	0,19	1,66	0,33	0,33	0,83		1	
2	0,64	0,45	0,09	0,42		0,92	1	
3	0,50	0,69	0,14	0,56				1
4	0,93	0,07	0,01	0,57				1
5	0,03	3,51	0,70	1,27	1,77		1	
6	0,23	1,47	0,29	1,56		2,06	1	
7	0,20	1,61	0,32	1,88	2,38		1	
8	0,90	0,11	0,02	1,90				1
9	0,25	1,39	0,28	2,18		2,68	1	
10	0,60	0,51	0,10	2,28				1
11	0,99	0,01	0,00	2,28				1
12	0,01	4,61	0,92	3,20	3,70		1	
13	0,90	0,11	0,02	3,22		3,72	1	
14	0,25	1,39	0,28	3,50				1
15	0,29	1,24	0,25	3,75	4,25	STOP		
						$\sum$	8	

Таким образом, по результатам трех испытаний число обслуженных заявок составило соответственно: $x_1 = 9$, $x_2 = 9$, $x_3 = 8$.

Найдем среднее число обслуженных заявок:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{3}(9+9+8) = 8{,}6.$$

Ответ: среднее число заявок, обслуженных СМО за 4 мин, равно 8,6.

ПРИЛОЖЕНИЕ

Таблица П.1

Показатели одно- и многоканальной СМО с ограниченной очередью

Показатель	Одноканальная СМО с ограниченной очередью	Многоканальная СМО с ограниченной очередью
1	2	3
Предельные вероятности	$p_0 = \dfrac{1-y}{1-y^{m+2}}$, $p_1 = yp_0$, $p_2 = y^2 p_0, \ldots,.$ $p_k = y^k p_0$	$p_0 = \left[1 + \dfrac{y}{1!} + \ldots + \dfrac{y^n}{n!} + \ldots + \dfrac{y^{n+1}\left(1-\left(\dfrac{y}{n}\right)^m\right)}{n \cdot n!\left(1-\dfrac{y}{n}\right)}\right]^{-1}$ $p_1 = \dfrac{y}{1!} p_0, \ldots, p_k = \dfrac{y^k}{k!} p_0, \ldots, p_n = \dfrac{y^n}{n!} p_0,$ $p_{n+1} = \dfrac{y^{n+1}}{n \cdot n!} p_0, \ldots, p_{n+r} = \dfrac{y^{n+r}}{n^r \cdot n!} p_0$ $(r = 1, \ldots, m)$
Вероятность отказа	$P_{\text{отк}} = P_{m+1} = y^{m+1} p_0$	$P_{\text{отк}} = P_{n+m} = \dfrac{y^{n+m}}{n^m \cdot n!} p_0$
Абсолютная пропускная способность	$A = \lambda Q = \lambda\left(1 - y^{m+1} p_0\right)$	$A = \lambda Q = \lambda\left(1 - \dfrac{y^{n+m}}{n^m \cdot n!} p_0\right)$
Относительная пропускная способность	$Q = 1 - P_{\text{отк}} =$ $= 1 - y^{m+1} p_0$	$Q = 1 - P_{\text{отк}} = 1 - \dfrac{y^{n+m}}{n^m \cdot n!} p_0$
Среднее число заявок в очереди	$L_{\text{оч}} =$ $= y^2 \dfrac{\left[1 - y^m (m+1-my)\right]}{\left(1-y^{m+2}\right)(1-y)}$	$L_{\text{оч}} = \dfrac{y^{n+1} p_0 \left[1 - \left(m+1-m\dfrac{y}{n}\right)\left(\dfrac{y}{n}\right)^m\right]}{n \cdot n!\left(1-\dfrac{y}{n}\right)^2}$

Окончание табл. 1

1	2	3
Среднее число заявок под обслуживанием (среднее число занятых каналов)	$L_{\text{об}} = 1 - p_0$	$\bar{k} = y\left(1 - \dfrac{y^{n+m}}{n^m \cdot n!} p_0\right)$
Среднее число заявок в системе	$L_{\text{сист}} = L_{\text{оч}} + L_{\text{об}}$	$L_{\text{сист}} = L_{\text{оч}} + \bar{k}$

Таблица П.2

Показатели эффективности одноканальной СМО с неограниченной очередью

Показатель	Обозначение	Аналитическое выражение
Среднее число заявок в системе	$L_{\text{сист}} = m$	$m = \dfrac{y}{1-y}$
Среднее число заявок в очереди	$L_{\text{оч}}$	$L_{\text{оч}} = \dfrac{y^2}{1-y}$
Среднее время пребывания заявки в системе	$T_{\text{сист}}$	$T_{\text{сист}} = \dfrac{1}{\lambda} L_{\text{сист}} = \dfrac{y}{\lambda(1-y)}$
Среднее время пребывания заявки в очереди	$T_{\text{оч}}$	$T_{\text{оч}} = \dfrac{y^2}{\lambda(1-y)}$

СПИСОК ЛИТЕРАТУРЫ

1. Вишневский В. М. Теоретические основы проектирования компьютерных сетей. М. : Техносфера, 2003. 512 с.

2. Вентцель Е. С. Исследование операций. М. : Советское радио, 1972. 552 с.

3. Волков И. К., Зуев С. М., Цветкова Г. М. Случайные процессы : учебник / под ред. В. С. Зарубина, А. П. Крищенко. М. : Изд-во МГТУ им. Н. Э. Баумана, 1999. 448 с.

4. Хинчин А. Я. Работы по математической теории массового обслуживания / под ред. Б. В. Гнеденко. М. : Физматгиз, 1963. 236 с.

5. Тихонов В. И., Миронов М. А. Марковские процессы. М. : Сов. радио, 1977.

6. Вентцель Е. С. Теория вероятностей : учебник. 6-е изд. стер. М. : Высш. шк., 1999. 576 с.

7. Кошуняева Н. В., Патронова Н. Н. Теория массового обслуживания : практикум по решению задач / САФУ имени М. В. Ломоносова. Архангельск : САФУ, 2013. 107 с.

8. Алиев Т. И. Основы моделирования дискретных систем. СПб. : СПбГУ ИТМО, 2009. 363 с.

9. Алиев Т. И., Муравьева-Витковская Л. А., Соснин В. В. Моделирование: задачи, задания, тесты. СПб. : НИУ ИТМО, 2011. 197 с.

10. Саакян Г. Р. Лекции. Теория массового обслуживания для студентов экономических специальностей очной, заочной и дистанционной форм обучения. Шахты : Изд-во ЮРГУЭС, 2006. 28 с.

11. Теория массового обслуживания : метод. указания, учебная программа и задания для студентов заочной формы обучения. Самара : СамГАПС, 2002. 38 с.

12. Дудин А. Н., Клименок В. И. Системы массового обслуживания с коррелированными потоками. Минск : Изд-во Белорус. ун-та, 2000.

13. Suurballe J. W. Disjoint Paths in a Network // Networks. 1974. № 4. P. 125–145.

14. Кремер Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов. 2-е изд., перераб. и доп. М. : ЮНИТИДАНА, 2004. 573 с.

15. URL: http://www.stgau.ru/company/personal/user/7912/files/lib/%D0%9F%D1%80%D0%B8%D0%BA%D0%BB%D0%B0%D0%B4%D0%BD%D1%8B%D0%B5%20%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D1%8B.pdf

Учебное издание

Черушева Татьяна Вячеславовна,
Зверовщикова Наталья Васильевна

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ

Редактор *А. Г. Темникова*, *В. В. Устинская*
Технический редактор *Ю. В. Анурова*
Компьютерная верстка *Ю. В. Ануровой*
Дизайн обложки *А. А. Стаценко*

Подписано в печать 30.11.2021.
Формат $60\times84^{1}/_{16}$. Усл. печ. л. 13,02.
Заказ № 645. Тираж 28.

Издательство ПГУ.
440026, Пенза, Красная, 40.
Тел. 66-60-49, 66-67-77; e-mail: iic@pnzgu.ru

Вниманию авторов!

Издательство ПГУ выпускает учебную, научную и художественную литературу, презентационную и акцидентную продукцию, а также полноцветные юбилейные и мемориальные издания в соответствии с ГОСТ 7.60–2003.

Издательство ПГУ принимает к изданию рукописи, подготовленные с использованием текстового редактора Microsoft Word for Windows версий **2003 и выше**. Формат – А4, основной шрифт – Times New Roman, 14–16 pt через одинарный интервал (минимальный размер шрифта в таблицах и сносках – 12,5 pt). Тип файла в электронном виде – doc, docx.

Работа должна содержать индекс УДК, аннотацию.

Аннотация (ГОСТ 7.86–2003, ГОСТ 7.9–1995) включает характеристику основной темы, проблемы объекта, цели работы и ее результаты. В аннотации указывают, что нового несет в себе данный документ в сравнении с другими, родственными по тематике и целевому назначению. Аннотация может включать сведения о достоинствах произведения. Текст аннотации начинают фразой, в которой сформулирована главная тема документа. Заканчивается аннотация читательским адресом.

Рисунки и таблицы должны быть размещены в тексте после ссылки на них (растровые рисунки представляются в виде отдельных файлов в формате jpg, BMP с разрешением 300 dpi, векторные рисунки в формате Corel Draw с минимальной толщиной линии 0,75 pt. Рисунки должны быть доступны для правки!). Они должны сопровождаться подрисуночными подписями, на все рисунки и таблицы в тексте должны быть ссылки.

Формулы в тексте выполняются только в редакторе формул **MathType версий 5.0** и выше. Символы греческого и русского алфавита должны быть набраны прямо, нежирно; латинского – курсивом, нежирно; обозначения векторов и матриц – прямо, жирно; цифры – прямо, нежирно. Наименования химических элементов набираются прямо, нежирно. Эти же требования необходимо соблюдать и в рисунках.

В списке литературы **нумерация источников** должна соответствовать очередности ссылок на них в тексте ([1], [2], …). Номер источника указывается в квадратных скобках. Требования к оформлению списка литературы, русских и иностранных источников (ГОСТ 7.0.5–2008): для книг – фамилия и инициалы автора, название, город, издательство, год издания, том, количество страниц; для журнальных статей, сборников трудов – фамилия и инициалы автора, название статьи, полное название журнала или сборника, серия, год, том, номер, страницы; для материалов конференций – фамилия и инициалы автора, название статьи, название конференции, город, издательство, год, страницы.

К материалам **должна** прилагаться следующая информация: фамилия, имя, отчество, контактные телефоны.

Контакты Издательства ПГУ: (8412) 66-60-49, 66-67-77. E-mail: iic@pnzgu.ru