А. Н. Дудин
Г. А. Медведев
Ю. В. Меленец

ПРАКТИКУМ НА ЭВМ ПО ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

УЧЕБНОЕ ПОСОБИЕ

Дудин А.Н., Медведев Г.А., Меленец Ю.В. Практикум на ЭВМ по теории массового обслуживания [Электронный ресур]: Учебное пособие — Электрон. текст. дан. (953 Кб). — Мн.: “Электронная книга БГУ”, 2003. — Режим доступа: http://anubis.bsu.by/publications/elresources/AppliedMathematics/dudin.pdf . — Электрон. версия печ. публикации, 2000. — PDF формат, версия 1.4 . — Систем. требования: Adobe Acrobat 5.0 и выше.

МИНСК

«Электронная книга БГУ»

2003

МИНИСТЕРСТВО ОБРАЗОВАНИЯ БЕЛАРУСИ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

А. Н. Дудин, Г. А. Медведев, В. Меленец

Практикум на ЭВМ по теории массового обслуживания

МИНСК
2000

УДК 519.7:681.3(076.5)(078.5)
ББК 22.18я73
Д 81

Рецензенты:
Кафедра прикладной математики и экономической кибернетики Белорусского государственного университета; Г.И.Лебедева,
Кандидат технических наук, доцент

Дудин А.Н., Медведев Г.А., Меленец Ю.В.
Д 81 Практикум на ЭВМ по теории массового обслуживания: Учеб.пособие. Мн.: Университетское, 2000. – 109 с.
ISBN 985-09-0300-7.

Пособие подготовлено по программам курсов “Теория случайных процессов” и “Теория массового обслуживания” для студентов, обучающихся по специальности “Прикладная математика”. Будет полезно студентам физико-математических, технических и экономических специальностей вузов по курсам “Теория вероятностей и математическая статистика”, “Проектирование информационно-вычислительных сетей”, “Системы управления базами данных” и т.д.

УДК 519.7:681.3(076.5)(078.5)
ББК 22.18я73

ISBN 985-09-0300-7

ПРЕДИСЛОВИЕ

Процессы массового обслуживания являются адекватными математическими моделями функционирования многих технических объектов, а также эволюций в экономике, природе и обществе. Довольно эффективны подходы ТМО при описании процессов в ЭВМ, их отдельные устройств или, наоборот, в сетях, составленных из них. Традиционно при помощи ТМО описываются и решаются задачи для систем и сетей передачи информации.

Настоящее учебное пособие имеет целью ознакомить с основными моделями теории массового обслуживания (ТМО) и метода их анализа. Следует сказать, что достаточно просто формулируемые математические модели ТМО не всегда поддаются простому математическому анализу. Как раз аналитические решения для систем массового обслуживания (СМО) довольно редки. Поэтому большое внимание уделяется численному анализу. Вместе с тем и уравнения для численного анализа иногда оказываются довольно сложными. В связи с этим приходится прибегать к имитационному моделированию для получения численных результатов. В этом случае играет роль не только возможность получения достаточно точных результатов вычисления характеристик, но и умение правильно оценивать статистическую значимость результатов имитационного моделирования.
СМО — структуры со случайным поведением, поэтому язык их описания — это язык теории вероятностей.

Учебное пособие предназначено для студентов, изучивших теорию вероятностей и математическую статистику или соответствующие разделы курса высшей математики. Поэтому разъяснению фактов из теории вероятностей и математической статистики в пособии внимания не уделяется. По ходу изложения результатов приводятся основные факты без доказательств. Иногда эти факты сопровождаются краткими объяснениями, если в этом имеется необходимость. В конце пособия приводится краткий список литературы, в которой содержится более подробное и строгое изложение общих результатов. В списке литературы приводится два учебника по курсу теории вероятностей и математической статистики, по мнению авторов, наиболее подходящих для понимания и описания результатов, изложенных в пособии. Приводится также несколько учебников по теории массового обслуживания, в которых подробно и строго описаны результаты пособия и по которым это пособие составлено. В первую очередь авторы при описании результатов ТМО следовали книге Л. Клейнрока.

Предполагается, что учебное пособие должно служить руководством к выполнению лабораторных работ. Студент, ознакомившись с результатами в теории, должен выполнить лабораторную работу, описание которой составлено по следующему принципу. Формируется цель работы и порядок ее выполнения, в котором сформулирована последовательность знаний. Выполнение лабораторных работ ориентировано на использование ЭВМ с графическим дисплеем, так как многие результаты очень полезно (а иногда необходимо) иллюстрировать графиками. Наиболее подходящим средством выполнения лабораторных работ являются персональные ЭВМ типа IBM PC с использованием соответствующей переферии и программного обеспечения. Основная направленность лабораторных работ — не реализация методов и процедур в виде программ, а проведения исследования эффективности этих методов и процедур. Поэтому отчет по лабораторным работам должен содержать не столько описания методов, процедур и реализующих их программ, сколько анализ численных результатов с особым креном на сравнение эффективности разных методов и процедур.

Пособие состоит из трех глав, посвященных основным направлениям в численном анализе ТМО. В первой главе (автор Г. А. Медведев) рассматриваются марковские СМО, т. е. такие, поведение которых описывается марковскими процессами или может быть описано ими при некотором расширении пространства состояний. Во второй главе (автор А. Н. Дудин) рассматриваются СМО, поведение которых только в некоторые моменты времени описывается марковскими процессами. В третьей главе (автор Ю. В. Меленец) используется метод имитационного моделирования, который пригоден для исследования произвольных СМО.

Глава 1

МАРКОВСКИЕ СМО

1.1 Цепи Маркова и СМО

СМО определяют следующие компоненты.

1. Входящий поток требований.
2. Характеристики времени обслуживания каждого требования.
3. Количество обслуживающих приборов.
4. Правила обслуживания на каждом приборе.

Несколько СМО могут быть объеденены в сеть массового обслуживания (СеМО), когда после обслуживания в одной СМО требование нуждается в обслуживании другой СМО. В этом случае имеют значение и свойства выходящего потока из рассматриваемой СМО, поскольку этот поток может оказаться входящим потоком (или частью входящего потока) в другую СМО сети. Удобно пользоваться обозначениями Кендалла, компактно описывающими тип СМО: $A|B|m|n$, где A — код функции распределения вероятностей длительности интервалов между последовательными требованиями входящего потока; B — код функции распределения вероятностей длительностей обслуживания требований; m — число обслуживающих приборов; n — число мест для ожидания. В случае, когда число мест для ожидания неограниченное, последний символ в обозначениях Кендалла обычно опускается. Так что обозначение $A|B|m|\infty$ эквивалентно обозначению $A|B|m$.

Символы A и B наиболее часто принимают значения из множества $\{M, E_k, H_k, G, GI, D\}$.

M обозначает экспоненциальное распределение

$$A(x) = 1 - \mathrm{e}^{-\lambda x}, \quad B(x) = 1 - \mathrm{e}^{-\mu x},\ x \geq 0. \tag{1.1}$$

E_k — распределение Эрланга порядка k с плотностью вероятностей:

$$a(x) = \frac{\partial A(x)}{\partial x} = \frac{k\lambda\,(k\lambda x)^{k-1}\mathrm{e}^{-k\lambda x}}{(k-1)!},\ x \geq 0. \tag{1.2}$$

H_k — гиперэкспоненциальное распределение порядка k:

$$A(x) = 1 - \sum_{i=1}^{k} \alpha_i \mathrm{e}^{-\lambda_i x},\ x \geq 0,\ \alpha_i \geq 0, \quad \sum_{i=1}^{k} \alpha_i = 1. \tag{1.3}$$

G, GI — произвольное распределение (I означает, что последовательные интервалы независимы).

D — распределение регулярного потока, когда плотность вероятностей интервалов вырождается в дельта-функцию:

$$a(x) = \delta(x - 1/\lambda), \quad b(x) = \delta(x - 1/\mu). \tag{1.4}$$

Вышеприведенные обозначения Кендалла используются в том случае, когда требования входящего потока порождаются источником неограниченного объема. Имеет смысл рассматривать случай, когда требования порождаются ограниченным числом источников, но каждый из них порождает только одно требование. Причем, если требование некоторого источника находится в СМО, то других требований этот источник не порождает. Для обозначения такой системы используются также обозначения Кендалла, но на первом месте указывается количество источников, а сами обозначения ограничиваются скобками, например, $< N|A|B|m|n >$. Здесь N — число источников, A — закон, порождающий требования входящего потока, B — распределение времени обслуживания, m — число приборов обслуживания, n — число мест ожидания.

В наиболее простых случаях под состоянием СМО понимают число требований, находящихся в данный момент в СМО.
В свою очередь, число требований в СМО является случайным, и в случайные моменты поступления требований оно увеличивается скачком на некоторую целочисленную величину (на единицу в случае ординарного входящего потока) и в случайные моменты окончания обслуживания оно уменьшается также на некоторое целое число.

Таким образом, адекватной математической моделью поведения СМО (то есть изменения ее состояний со временем) является случайный процесс. Поведение широкого круга СМО можно описать марковскими случайными процессами. Марковские случайные процессы с конечным или счетным множеством возможных состояний обычно называют *цепями Маркова*. В простых случаях цепью Маркова описывается поведение таких СМО, у которых функции распределения A и B относятся к классу M.

В большинстве случаев для исследования представляет интерес определение характеристик эргодических СМО в стационарном режиме, то есть в такие моменты времени, в которые влияние начального состояния уже не ощущается. Наиболее интересными характеристиками СМО являются следующие.

1. Условие существования стационарного режима.
2. Стационарное распределение вероятностей состояний СМО.
3. Среднее число требований, находящихся в СМО в стационарном режиме (или находящихся в очереди на обслуживание).
4. Вероятность потери требования (в СМО с потерями), когда в момент поступления требования в СМО нет свободных мест.
5. Вероятность того, что поступившее в СМО требование будет обслужено после некоторого ожидания.
6. Функция распределения времени ожидания.
7. Среднее время ожидания обслуживания.
8. Среднее время пребывания требования в СМО.

Рассмотрим проблему нахождения этих характеристик. Марковские процессы подобно решениям дифференциальных уравнений имеют следующее основное свойство: если состояние процесса в момент времени t зафиксировано, то дальнейшее течение процесса определяется только этим состоянием и не зависит от того, каким образом процесс оказался в этом состоянии, иначе говоря, изменение марковского процесса после момента времени t не зависит от того, как этот процесс вел себя до момента t. Основываясь на этом свойстве и формуле полной вероятности, можно получить важную формулу Чепмена—Колмогорова для цепей Маркова

$$p_{ij}(t+\tau) = \sum_{k\in\Omega} p_{ik}(t)p_{kj}(\tau),\ i,j\in\Omega, \tag{1.5}$$

где Ω — множество возможных состояний (конечное или счетное), i, j, k — состояние цепи Маркова (в простых случаях под состоянием можно понимать количество требований в СМО, описываемой цепью Маркова), $p_{ik}(t)$ — вероятность того, что цепь Маркова за время t перейдет из состояния i в состояние k.

Соотношение (1.5) можно рассматривать как уравнение, связывающее элементы некоторых матриц. Пусть $\omega = |\Omega|$ — число элементов множества Ω $(1 < \omega \leq +\infty)$. Определим $P(t)$ — $(\omega\times\omega)$−матрицу с элементами $p_{\alpha\beta}(t),\ 1 \leq \alpha,\beta \leq \omega$. Тогда (1.5) можно записать в матричной форме

$$P(t+\tau) = P(t)P(\tau). \tag{1.6}$$

Цепь Маркова называется *эргодической*, если

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j,\ i,j\in\Omega,\ \sum_{j\in\Omega}\pi_i = 1. \tag{1.7}$$

Вероятности π_i называются *финальными* вероятностями. Финальные вероятности образуют стационарное распределение, если для всех $t > 0$ выполняется равенство

$$\pi_j = \sum_{i\in\Omega}\pi_i p_{ij}(t),\ j\in\Omega.$$

В дальнейшем различия между финальными и стационарными вероятностями делаться не будут, так как в большинстве случаев эргодические цепи Маркова имеют единственное стационарное распределение, образуемое финальными вероятностями.
Обозначим
$\pi = (\pi_1, \pi_2, \ldots, \pi_\omega)$ — вектор стационарных вероятностей. Таким образом, проблема существования стационарного распределения вероятностей связана с установлением эргодичности цепи Маркова. Ниже будут сформулированы некоторые условия эргодичности.

В случае дискретного времени, когда состояния цепи Маркова могут изменяться только в моменты времени, кратные некоторой единице времени δ, то есть в моменты $t = k\delta,\ k = 1, 2, \ldots$, удобно $p_{ij}(k\delta)$ обозначить $p_{ij}(k),\ k \geq 1$ и далее ввести $p_{ij}(1) = p_{ij}$ — вероятности переходов за 1 шаг, то есть за единицу дискретного времени δ. В этом случае матричное соотношение (1.6) можно переписать следующим образом (при $t = k\delta,\ \tau = \delta$):

$$P(k+1) = P(k)P = P^{k+1}, \tag{1.8}$$

где P — матрица вероятностей одношаговых переходов.

В случае, когда стационарный режим СМО существует, предельная матрица

$$P(\infty) = \lim_{k\to\infty} P(k) = \lim_{k\to\infty} P^k$$

должна обладать следующим свойством: значения элементов матрицы $P(\infty)$ не должны зависеть от номера строки. Матрица $P(\infty)$ состоит из одинаковых строк, которыми являются π. Отсюда и из (1.7) получаем уравнение для определения стационарных вероятностей

$$\pi = \pi P, \quad \text{или} \quad \pi(I-P) = 0. \tag{1.9}$$

Это линейное алгебраическое уравнение относительно π. Оно является однородным. Матрица $(I - P)$ является вырожденной (так как сумма ее столбцов равна нулю). Поэтому решение уравнения

(1.9) получается с точностью до постоянного множителя, который может быть найден из условия нормировки

$$\sum_{j\in\Omega}\pi_j = 1, \quad \text{или} \quad \pi\mathbf{E} = 1. \tag{1.10}$$

Здесь $\mathbf{E}$ — вектор—столбец, составленный из единиц. Таким образом, в случае дискретного времени определение стационарного распределения вероятностей состояний сводится к решению системы (1.9)–(1.10).

В случае непрерывного времени в (1.6) $t \in (0, +\infty)$. Перепишем (1.6) в виде

$$\frac{P(t+\tau)-P(t)}{\tau} = P(t)\frac{P(\tau)-I}{\tau}$$

и найдем предельные выражения этого равенства при $\tau \to 0$ (соответствующие производные существуют), известное под названием*прямое уравнение Колмогорова*:

$$\frac{\partial P(t)}{\partial t} = P(t)Q, \quad Q = \lim_{\tau\to 0}\frac{P(\tau)-I}{\tau}. \tag{1.11}$$

Элементами матрицы Q являются

$$\nu_{ij} = \lim_{\tau\to 0}\frac{P_{ij}(\tau)}{\tau},\ i \neq j, \quad \nu_{ii} = \lim_{\tau\to 0}\frac{P_{ii}(\tau)-I}{\tau}. \tag{1.12}$$

Величины ν_{ij} называют *инфинитезимальными коэффициентами*, а Q — *инфинитезимальной матрицей*. В нашем случае удобнее другие названия: если $i \neq j$, то ν_{ij} — интенсивность перехода цепи Маркова из состояния i в j; если $i = j$, то $-\nu_{ii}$ — интенсивность выхода из состояния i. По определению $\nu_{ij} \geq 0$ для $i \neq j$ и $\nu_{ii} < 0$. Поскольку для любых $\tau \quad (P(\tau) - I)\mathbf{E} = 0$, то

$$\sum_{j\in\Omega}\nu_{ij} = 0, \quad \text{или} \quad \nu_{ii} = -\sum_{\substack{j\neq i\\ j\in\Omega}}\nu_{ij},\ i \in \Omega. \tag{1.13}$$

Как и для дискретного времени, $P(\infty)$ — матрица, состоящая из одинаковых строк $\pi = (\pi_1, \ldots, \pi_\omega)$, которые не зависят от времени t. Поэтому из первого равенства (1.11) получаем так называемое уравнение равновесия для стационарных вероятностей в случае непрерывного времени

$$\pi Q = 0, \tag{1.14}$$

которое совместно с (1.10) однозначно определяет вероятности π. Таким образом, для определения стационарных вероятностей в случае непрерывного времени приходится решать систему (1.10), (1.14).

Трудности определения стационарных вероятностей π возникают тогда, когда порядок систем (1.9), (1.14) очень большой, т. е. когда $\omega \gg 1$.

Как уже было сказано выше, существование стационарных распределений связано со свойством эргодичности цепи Маркова. Приведем некоторые условия, при выполнении которых цепь Маркова обладает эргодическими свойствами и имеет единственное стационарное распределение.

Для дискретного времени:

1. *Достаточные условия Маркова*: если $\omega < +\infty$ и существует такое k, что

$$\min_{i,j\in\Omega} p_{ij}(k) > 0 \ \text{ для всех } i, j \in \Omega.$$

2. *Достаточные условия Маркова—Бернштейна:* если $\omega < +\infty$ и существуют такие $j \in \Omega$ и $k \geq 1$, что для любого $i \in \Omega \quad p_{ij}(k) > \qquad\qquad > 0$.

3. *Необходимое и достаточное условие Фостера:* система уравнений (1.9) имеет ненулевое решение, такое, что $\sum_{i\in\Omega}|\pi_i| < +\infty$.

4. *Достаточное условие Мустафы:* существует $\varepsilon > 0$, натуральное число c и набор неотрицательных чисел $x_j,\ j \in \Omega$, таких, что

$$\sum_{j\in\Omega} p_{ij}x_j \le x_i - \varepsilon \quad \text{для} \quad i > c,$$

$$\sum_{j\in\Omega} p_{ij}x_j \le +\infty \quad \text{для} \quad i \le c.$$

Для непрерывного времени:

5. *Достаточные условия Маркова–Берштейна:* если $\omega < +\infty$ и существует $j \in \Omega$ и $t > 0$ такие, что $p_{i,j}(t) > 0$ для всех $i \in \Omega$.

6. *Необходимое и достаточное условие Фостера:* система уравнений равновесия (1.14) имеет ненулевое решение такое, что $\sum_{i\in\Omega}|\pi_i| < +\infty$.

С практической точки зрения наиболее удобным критерием эргодичности является критерий Фостера, поскольку фактически он предполагает выяснение того, имеют ли решения системы уравнений (1.9), (1.10) или (1.10), (1.14). Но так как решение этих систем необходимо для получения стационарных вероятностей, то, решая их, одновременно устанавливают эргодичность СМО.

После того, как установлена эргодичность СМО и найдено стационарное распределение вероятностей ее состояний, среднее число $\overline{N}$ требований, находящихся в СМО в стационарном режиме, и среднее число $\overline{N}_q$ требований, находящихся в очереди, определяются по формулам

$$\overline{N} = \sum_{j\in\Omega} j\pi_j, \quad \overline{N}_q = \sum_{j>m}(j-m)\pi_j. \tag{1.15}$$

Для систем общего вида $G|G|m|\infty$ справедлива также формула $\overline{N}_q = \overline{N} - \rho_m$, где $\rho_m = \lambda/\mu m$, λ — интенсивность входящего потока требований; $(1/\mu)$ — средняя длительность обслуживания каждого требования; m — число приборов обслуживания в СМО.

Если входящий поток требований является простейшим (это означает, что $A(x)$ имеет код M), то каждое требование, поступающее в СМО, застает в нем в стационарном режиме k требований с вероятностью $\pi_k,\ k \in \Omega$. Время ожидания обслуживания при $k < m$ равно нулю, а в случае $k \ge m$ оно равно времени обслуживания $(k-m+1)$ требований m приборами.

Потеря требования входящего потока происходит тогда, когда это требование застает в СМО $(m+n)$ требований. Это означает, что все приборы обслуживания и все места для ожидания заняты. Вероятность этого равна p_{m+n}.

Вероятность того, что требование будет обслужено после некоторого ожидания

$$p_w = \sum_{k=m}^{m+n-1} \pi_k. \tag{1.16}$$

Наконец, вероятность того, что поступившее требование начнет обслуживаться без ожидания сразу же по прибытии

$$p_0 = \sum_{k=0}^{m-1} \pi_k.$$

Для марковских СМО $M|M|m|n$, когда время обслуживания распределено экспоненциально, время обслуживания $(k-m+1)$ требований m приборами имеет функцию распределения

$$W_k(x) = 1 - \mathrm{e}^{-m\mu x}\sum_{j=0}^{k-m}\frac{(m\mu x)^j}{j!},\ k \ge m.$$

Поэтому функция распределения времени ожидания имеет вид

$$W(x)=\sum_{k=0}^{m-1}\pi_k+\sum_{k=m}^{m+n-1}\pi_k W_k(x)=1-\mathrm{e}^{-m\mu x}\sum_{j=0}^{n-1}\frac{(m\mu x)^j}{j!}\sum_{k=j}^{n-1}\pi_{k+m},\ x>0. \tag{1.17}$$

Среднее время ожидания вычисляется по формуле

$$W=\frac{1}{m\mu}\sum_{k=0}^{n-1}(k+1)\pi_{k+m}. \tag{1.18}$$

Время пребывания в СМО определяется как сумма времени ожидания требования и времени его обслуживания. Поэтому его плотность вероятностей

$$s(y)=\int\limits_0^y \mu\mathrm{e}^{-\mu(y-x)}\,dW(x)$$

, где $W(x)$ определяется формулой (1.17).

Среднее время пребывания требования в СМО $M|M|m|n$

$$T=\frac{1}{\mu}+W=\frac{1}{\mu}\left(1+\frac{1}{m}\sum_{k=0}^{n-1}(k+1)\pi_{k+m}\right). \tag{1.19}$$

Между введенными средними характеристиками в общем случае СМО $G|G|m|\infty$, когда число мест для ожидания неограничено ($n=\infty$), имеют место простые соотношения, известные под названием *формулы Литтла*:

$$\overline{N}=\lambda T,\quad \overline{N}_q=\lambda W. \tag{1.20}$$

Итак, основные характеристики СМО $M|M|m|n$ можно определить, решая системы уравнений (1.9), (1.10) или (1.10), (1.14) и используя формулы (1.15)–(1.19). Как видим, основную трудность при этом представляет решение систем уравнений, которые могут иметь неограниченный порядок (при $\omega=\infty$).

Решения для некоторых простейших СМО имеют следующую структуру.

СМО $\mathbf{M|M|\infty}$. В этой системе неограниченное число приборов обслуживания ($m=\infty$). Стационарный режим существует для любых λ,μ, определенных в (1.1).

$$\pi_k=\mathrm{e}^{-\rho}\frac{\rho^k}{k!},\ 0\le k\le+\infty.$$

Здесь $\rho=\lambda/\mu$.

$$\overline{N}=\rho,\quad \overline{N}_q=0,\quad p_0=1-p_w=1,\quad W=0,\quad T=1/\mu.$$

СМО $\mathbf{M|M|m|n}$. Стационарный режим существует при любых λ,μ.
$\Omega=\{0,1,2,\ldots,m+n\}$, $\rho_m=\lambda/\mu m,\ \rho_1=\rho$.

$$\pi_k=\pi_0\frac{m^{(\min\{k,m\})}}{(\min\{k,m\})!}\rho_m^k,\ 0\le k\le m+n.$$

$$\pi_0=\left(\sum_{k=0}^{m-1}\frac{\rho_1^k}{k!}+\frac{\rho_1^m}{m!}\sum_{k=0}^{n}\rho_m^k\right)^{-1}.$$

Среднее число требований в СМО

$$\overline{N}=\rho\pi_0\pi(m-1)+\frac{\rho_m\pi_m}{(1-\rho_m)^2}(1+(1-\rho_m)m-(1+(1-\rho_m)(m+n)\rho_m^n).$$

Среднее число требований в очереди

$$\overline{N}_q = \frac{\rho_m \pi_m}{(1-\rho_m)^2}(1-(1+(1-\rho_m)n)\rho_m^n).$$

В случае, когда $\rho_m = 1$ ($\rho = m$),

$$\overline{N} = m\pi_0\pi(m-1) + n\pi_m(m+(n+1)/2), \quad \overline{N}_q = n\pi_m(n+1)/2.$$

Здесь обозначено $\pi(k) = \sum\limits_{i=0}^{k} \rho^i/i!$. Вероятность потери требования

$$\pi_{m+n} = \pi_0 \frac{\rho^{m+n}}{m!m^n} = \pi_m(\rho_m)^n.$$

Вероятность ожидания

$$p_w = \pi_m \left(1-(\rho_m)^n\right) / \left(1-\rho_m\right).$$

Вероятность обслуживания без ожидания

$$p_0 = 1 - p_w - \pi_{m+n} = 1 - \pi_m \left(1-(\rho_m)^{n+1}\right) / \left(1-\rho_m\right).$$

Функция распределения времени ожидания

$$W(x) = 1 - \frac{\pi_m}{1-\rho_m} \mathrm{e}^{-m\mu x} \sum_{k=0}^{n-1} \frac{(m\mu x)^k}{k!} \left((\rho_m)^k - (\rho_m)^n\right), \ x > 0.$$

Среднее время ожидания

$$W = \frac{m\pi_m}{\mu(m-\rho)^2} \left(1-(n+1)(\rho_m)^n + n(\rho_m)^{n+1}\right).$$

Среднее время пребывания требования в СМО

$$T = W + 1/\mu.$$

В частном случае СМО $M|M|1|n$ имеем

$$\pi_k = \rho^k(1-\rho)/(1-\rho^{n+2}), \ 0 \le k \le n+1.$$

$$\overline{N} = \rho \frac{1-(1+(1-\rho)(n+1))\rho^{n+1}}{(1-\rho)(1-\rho^{n+2})}, \quad \overline{N}_q = \rho^2 \frac{1-(1+(1-\rho)n)\rho^{n+1}}{(1-\rho)(1-\rho^{n+2})}.$$

СМО $\mathbf{M|M|m|0}$. Это система с потерями без ожидания. Стационарное распределение существует всегда. $\rho = \lambda/\mu$.

$$\pi_k = \rho^k/k!\,\pi(m), \ 0 \le k \le m, \quad \pi(m) = \sum_{i=0}^{m} \rho^i/i!.$$

$$\overline{N} = \rho\pi(m-1)/\pi(m), \quad \overline{N}_q = 0.$$

Вероятность потери требования (формула потерь Эрланга)

$$\pi_m = \rho^m/m!\,\pi(m).$$

Другие характеристики:

$$p_w = 0, \quad W = 0, \quad T = 1/\mu.$$

СМО M|M|m|∞. Это так называемая система с ожиданием. Стационарное распределение существует, если $\rho_m = \lambda/\mu m < 1,\ \rho_1 = \rho$.

$$\pi_k = \pi_0 \frac{m^{(\min\{k,m\})}}{(\min\{k,m\})!} \rho_m^k, \quad 0 \le k \le +\infty.$$

$$\pi_0 = \left(\sum_{k=0}^{m-1} \frac{\rho^k}{k!} + \frac{\rho^m}{m!} \frac{1}{1-\rho_m} \right)^{-1}.$$

Средние значения числа требований в СМО и в очереди

$$\overline{N} = \rho(1 + \frac{\rho_w}{m-\rho}), \quad \overline{N}_q = \frac{\rho \rho_w}{m-\rho},$$

где вероятность ожидания

$$p_w = \pi_0 \rho^m/(m-1)!(m-\rho).$$

Потерь требований в этой системе не происходит.

Функция распределения времени ожидания

$$W(x) = 1 - p_w \mathrm{e}^{-(m\mu - \lambda)x},\ x > 0.$$

Среднее значение времен ожидания и пребывания

$$W = p_w/\mu(m-\rho), \quad T = (p_w + m - \rho)/\,\mu(m-\rho).$$

В частном случае СМО $M|M|1|\infty$ имеем

$$\pi_k = (1-\rho)\rho^k,\ 0 \le k < \infty,$$

$$p_w = \rho, \quad \overline{N} = \rho/(1-\rho), \quad \overline{N}_q = \rho^2/(1-\rho),$$

$$W(x) = 1 - \rho \mathrm{e}^{-(\mu - \lambda)x},\ x > 0,$$

$$W = \rho/\mu(1-\rho), \quad T = 1/\,\mu(1-\rho).$$

СМО $<$ N|M|M|m|n $>$. Это так называемая *замкнутая СМО*, в которой функционирует всего N требований. Здесь естественно считать, что $n + m \le N$. Стационарный режим в такой СМО существовать может при любых λ и μ.

Стационарные вероятности определяются формулами

$$\pi_k = N!\,\pi_0 \frac{m^{(\min\{k,m\})}}{(N-k)!(\min\{k,m\})!} \rho_m^k,\ 0 \le k \le m+n.$$

$$\pi_0 = \frac{1}{N!} \left(\sum_{k=0}^{m-1} \frac{1}{(N-k)!} \frac{\rho^k}{k!} + \frac{\rho^m}{m!} \sum_{k=0}^{n} \frac{1}{(N-m-k)!} \rho_m \right)^{-1}.$$

Аналитические выражения для остальных характеристик СМО в этом случае являются достаточно громоздкими, поэтому они не приводятся, а их вычисление удобнее осуществить при помощи формул (1.15)–(1.19) после вычисления стационарных вероятностей π_k.

Лабораторная работа: "Исследование простых марковских СМО".

Цель работы. Исследовать зависимость основных характеристик СМО от условий функционирования СМО.

Задание 1. Выбрать тип СМО и вычислить характеристики (1.15)–(1.19)для выбранной СМО. Найти зависимость этих характеристик от коэффициента загрузки ρ. Зависимость получить в виде таблиц и графиков.

Задание 2. Провести следующий сравнительный анализ СМО $M|M|m|n$ и $< N|M|M|m|n >$. Понятно, что СМО с ограниченным количеством требований $< N|M|M|m|n >$ при $N \to \infty$ должна приближаться по своим показателям к СМО $M|M|m|n$ при одинаковой интенсивности входящего потока. Поэтому, если задаться некоторой точностью ε, то интересно выяснить, при каком значении N указанные СМО не различаются на уровне выбранной точности.

Иначе говоря, необходимо положить в СМО $< N|M|M|m|n >$ интенсивность входящего потока от одиночного требования равной λ/N. Тогда для этой системы в формулах для π_k необходимо использовать вместо ρ величину ρ/N. Имея это ввиду, установить, при каком значении $N_{\min}$ удовлетворится неравенство

$$\sum_{k=o}^{m+n} |\pi_k - < \pi_k > | \leq \varepsilon.$$

Здесь π_k — стационарные вероятности СМО $M|M|m|n$, а $< \pi_k >$ — стационарные вероятности СМО $< N|M|M|m|n >$. Установить зависимость $N_{\min}(\rho)$ для различных m и n.

Задание 3. Провести следующий сравнительный анализ СМО 1) $M|M|m|n$ и СМО 2) $M|M|m|\infty$. Понятно, что первая СМО переходит во вторую при $n \to \infty$. Если задаться некоторой точностью ε, то интересно выяснить, при каком значении $n_{\min}$ характеристики указанных СМО не различаются на уровне выбранной точности. Как в предыдущем случае, это будет при

$$\sum_{k=0}^{m+n} |\pi_k^{(1)} - \pi_k^{(2)}| \leq \varepsilon$$

и, кроме того, должны выполняться неравенства

$$\pi_{m+n}^{(1)} \leq \varepsilon, \quad |W^{(1)} - W^{(2)}| \leq \varepsilon, \quad |\overline{N}_q^{(1)} - \overline{N}_q^{(2)}| \leq \varepsilon.$$

Установить зависимость $n_{\min}$ для различных m.

Задание 4. Выяснить, насколько точно формулы Литтла (1.20) характеризуют $\overline{N}$ и $\overline{N}_q$ для СМО $M|M|m|n$ и $< N|M|M|m|n >$. Для этого рассмотреть зависимость точности формул Литтла от загрузки системы, то есть определить зависимость

$$\overline{N}(1.15) - \overline{N}(1.20) = f_1(\rho), \quad \overline{N}_q(1.15) - \overline{N}_q(1.20) = f_2(\rho).$$

Здесь $\overline{N}(1.15)$ означает значение $\overline{N}$, вычисленное по формуле (1.15) и т. д. Определить также подобные зависимости от m и n.

Задание 5. Предположим, что за обслуживание каждого требования СМО получает доход c_0. Вместе с тем на содержание каждого ожидающего требования в течение одной единицы времени СМО несет затраты c_1. Кроме этого, на содержание одного прибора обслуживания в течение одной единицы времени требуются затраты c_2, а на содержание одного места для ожидания в течение единицы времени требуются затраты c_3. Тогда в стационарном режиме СМО в течение времени T получает в среднем доход

$$I = c_0 \lambda T (1 - \pi_{m+n})$$

и несет средние суммарные затраты

$$L = \overline{N}_q c_1 T + m c_2 T + n c_3 T.$$

Таким образом, прибыль для такой СМО за работу в течение времени T равна

$$P = I - L = c_0 \lambda T \left(1 - \pi_{m+n} - \frac{1}{\rho}\overline{N}_q l_1 - \frac{m}{\rho} l_2 - \frac{n}{\rho} l_3\right),$$

где через l_k обозначены отношения $c_k/\mu c_0$. Понятно, что применение описанной СМО целесообразно в том случае, когда $P > 0$, то есть при выполнении неравенства

$$RL = \rho\pi_{m+n} + \overline{N}_q l_1 + ml_2 + nl_3 < \rho. \tag{1.21}$$

Необходимым условием также является неравенство

$$l_1 + l_2 + l_3 < 1, \; l_i \geq 0. \tag{1.22}$$

Представляет интерес выяснить, при каких значениях m и n относительные потери RL оказываются минимальными, то есть определить оптимальные значения m_{opt} и n_{opt} при заданном коэффициенте загрузки ρ. Поэтому необходимо решить следующую задачу для СМО $M|M|m|n$:

$$RL = \rho\pi_{m+n} + \overline{N}_q l_1 + ml_2 + nl_3 \to \min_{m,n} \tag{1.23}$$

при выполнении условий (1.21) и (1.22). Установить также зависимости $m_{opt}(\rho)$ и $n_{opt}(\rho)$.

Задание 6. Выполнить задание 5 со следующим изменением условий. $c_1 = 0$, но СМО несет потери c_4, если время ожидания требования превысит некоторую величину T_0. В этом случае относительные потери определятся формулой

$$RL = \rho\pi_{m+n} + ml_2 + nl_3 + l_4(1 - W(T_0)),$$

где $W(T_0)$ — вероятность того, что время ожидания не превышает T_0, и вычисляется при помощи (1.17).

$$1 - W(T_0) = \mathrm{e}^{-m\mu T_0} \sum_{j=0}^{n-1} \frac{(m\mu T_0)^j}{j!} \sum_{k=j}^{n-1} \pi_{k+m}.$$

Так что задачей является определение m и n, минимизирующих модифицированные RL при аналогичных условиях, что и в задании 5. Кроме зависимости $m_{opt}(\rho)$ и $n_{opt}(\rho)$, рассмотреть зависимость $m_{opt}(T_0)$ и $n_{opt}(T_0)$. Провести сравнение с результатами задания 5.

1.2 Процессы размножения и гибели.

Скалярные процессы размножения и гибели. Марковские цепи с непрерывным временем, у которых инфинитезимальные коэффициенты (1.12) обладают свойствами

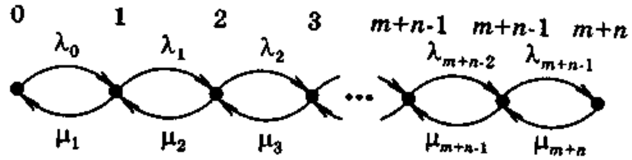
$$\nu_{ij} = 0 \quad \text{для всех } |i - j| > 1, \; i, j \in \Omega, \tag{1.24}$$

называются *процессами размножения и гибели*. При этом обычно обозначают

$$\nu_{i,i+1} = \lambda_i, \quad \nu_{i,i-1} = \mu_i, \tag{1.25}$$

так что $\nu_{ii} = -\lambda_i - \mu_i$. Все рассмотренные в предыдущем параграфе конкретные СМО описываются именно такими процессами. Поведение этих СМО удобно иллюстрировать при помощи стохастических графов, у которых вершины обозначают состояния СМО, а дуги — возможные переходы из состояния в состояние. При этом каждой дуге приписывается некоторый вес, равный интенсивности соответствующего перехода. Такой граф для СМО $M|M|m|n$ представлен на рис. 1.1, коэффициенты λ_i определяются формулами

$$\lambda_i = \lambda, \quad \mu_i = \mu \, \min\{i, m\}.$$



Для СМО $< N|M|M|m|n >$ граф переходов будет иметь такой же вид, как и на рис. 1.1, однако коэффициенты λ_i определяются по—другому:

$$\lambda_i = (N - i)\lambda.$$

Все сказанное будет также справедливо и для цепей Маркова с дискретным временем, если вместо инфинитезимальных коэффициентов ν_{ij}, $i \neq j$, использовать вероятности одношаговых переходов p_{ij}, а вместо ν_{ii} — величину $(p_{ii} - 1)$.

Произведение $\pi_i \nu_{ij}$ принято называть *потоком вероятности из состояния i в состояние j*. Уравнение равновесия (1.14) может быть записано в виде

$$\sum_{i \in \Omega(j)} \pi_i \nu_{ij} = \pi_j \sum_{k \in \Omega(j)} \nu_{jk},\ j \in \Omega. \tag{1.26}$$

Здесь через $\Omega(j)$ обозначено множество $\Omega \setminus \{j\}$. Это можно сформулировать словами: в стационарном режиме поток вероятности из любого состояния j цепи Маркова равен суммарному потоку вероятностей из других состояний цепи в это состояние j.

Для процессов размножения и гибели с учетом свойства (1.24) это уравнение записывается в виде

$$\pi_j(\lambda_j + \mu_j) = \pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1}. \tag{1.27}$$

Используя эту формулу как рекуррентное соотношение мож-но облегчить процесс вычисления стационарных вероятностей, связанных с решением системы уравнений (1.10), (1.14).

Справедливо и другое, более общее свойство стационарных потоков вероятности в цепи Маркова.

Пусть множества Ω' и Ω'' определены соотношениями

$$\Omega' \cup \Omega'' = \Omega, \quad \Omega' \cap \Omega'' = \varnothing. \tag{1.28}$$

Тогда для любых Ω' и Ω'', удовлетворяющих (1.28), суммарный поток вероятностей из состояний множества Ω' в состояния множества Ω'' равен суммарному потоку вероятностей из состояний множества Ω'' в состояния множества Ω'. То есть

$$\sum_{i \in \Omega'} \sum_{j \in \Omega''} \pi_i \nu_{ij} = \sum_{i \in \Omega'} \sum_{j \in \Omega''} \pi_j \nu_{ji}. \tag{1.29}$$

Соотношение (1.29) может оказаться очень полезным при определении стационарных вероятностей π_k, когда можно найти такое разбиение множества состояний Ω, при котором ненулевых коэффициентов ν_{ij} или ν_{ji}, $i \in \Omega'$, $j \in \Omega''$, достаточно мало. Например, если при анализе процессов размножения и гибели, для которых справедливы (1.24), (1.25), выбрать $\Omega' = \{k \,|\, 0 \leq k \leq i\}$, $\Omega'' = \{k \,|\, k > i\}$, то соотношение (1.29) принимает наиболее простой вид:

$$\lambda_i \pi_i = \mu_{i+1} \pi_{i+1}, \ 0 \le i < \omega. \tag{1.30}$$

Используя это рекуррентное соотношение и условие нормировки (1.10) получаем

$$\pi_i = \pi_0 \prod_{j=0}^{i-1} \left(\frac{\lambda_j}{\mu_{j+1}} \right), \quad \pi_0 = \left(1 + \sum_{i=1}^{\omega-1} \prod_{j=0}^{i-1} \left(\frac{\lambda_j}{\mu_{j+1}} \right) \right)^{-1}. \tag{1.31}$$

Стационарные вероятности для СМО, описанных в предыдущем параграфе, как раз и вычисляются по таким формулам, конкретизированным в соответствие с тем, как зависят от индексов λ_j и μ_j.

Заметим, что при конечных ω вероятности π_k всегда положительны, что обеспечивает существование стационарного режима. При $\omega \to \infty$ сумма ряда в скобках формулы, определяющей π_0, может расходиться. Тогда $\pi_0 \to 0$ при $\omega \to \infty$. В этом случае для всякого $k < +\infty$ при $\omega \to \infty$ $\pi_k \to 0$ и стационарный режим не существует. Отсюда следует, что условием существования стационарного режима при $\omega \to \infty$ является выполнение неравенства

$$\sum_{i=1}^{\infty} \prod_{j=0}^{i-1} \left(\frac{\lambda_j}{\mu_{j+1}} \right) < +\infty. \tag{1.32}$$

Для выполнения неравенства (1.32) достаточно, чтобы существовало такое i_0, для которого $\lambda_j < \mu_{j+1}$ при всех $j > i_0$.

Из изложенного видно, что определение стационарных вероятностей состояния СМО, описываемых процессами размножения и гибели, является достаточно простым. Эта простота обеспечивается свойством (1.24). Можно расширить понятия процессов размножения и гибели с тем, чтобы использовать рассмотренную процедуру определения стационарных вероятностей для более широкого класса цепей Маркова.

Векторные процессы размножения и гибели. До сих пор под состоянием $j \in \Omega$ СМО понималось число требований в СМО. В дальнейшем будем понимать под j номер состояния СМО, а под Ω— множество возможных состояний СМО. Как и прежде, будем считать, что $\Omega = \{0, 1, 2, \ldots, \omega\}, \ \omega \le +\infty$. Предположим, что состояния СМО можно перенумеровать таким образом, чтобы выполнились следующие свойства.

1) множество Ω представимо в виде совокупности непустых подмножеств $\Omega_i, \ 0 \le i \le n$;

$$\Omega = \Omega_0 \cup \Omega_1 \cup \ldots \cup \Omega_n, \quad \Omega_i \cap \Omega_j = \varnothing \quad \text{для всех } i \ne j, \tag{1.33}$$

причем $\alpha < \beta$, если $\alpha \in \Omega_i, \ \beta \in \Omega_j$ и $i < j$;

2) инфинитезимальные коэффициенты (1.12) обладают свойством

$$\nu_{\alpha,\beta} = 0, \ \ \alpha \in \Omega_i, \ \beta \in \Omega_j, \ |i - j| > 1. \tag{1.34}$$

Цепи Маркова с непрерывным временем, у которых инфинитезимальные коэффициенты обладают свойством (1.34) будем называть *векторными процессами размножения и гибели*. Подмножества $\Omega_i, \ 0 \le i \le n$ будем называть *уровнями*. На уровне Ω_i находится ω_i состояний, $\omega_i > 0, \ \omega_0 + \omega_1 + \ldots + \omega_n = \omega$. Свойство инфинитезимальных коэффициентов (1.34) означает, что цепь Маркова может переходить из состояния некоторого уровня Ω_i только в состояние соседнего уровня (Ω_{i-1} или Ω_{i+1}).

Для таких процессов уравнения равновесия (1.29) будут выглядить так:

$$\sum_{\alpha \in \Omega_{i-1}} \pi_\alpha \nu_{\alpha j} + \sum_{\beta \in \Omega_i} \pi_\beta \nu_{\beta j} + \sum_{\gamma \in \Omega_{i+1}} \pi_\gamma \nu_{\gamma j} = 0, \quad j \in \Omega_i. \tag{1.35}$$

Пусть $\pi = (\pi_0 \pi_1 \ldots \pi_\omega) = (\pi(0)\pi(1) \ldots \pi(n))$, где мы обозначим через $\pi(\mathbf{i})$ вектор—строку, составленную из стационарных вероятностей $\pi_j, \ j \in \Omega_i$. Через Q_{ij} обозначим матрицу, элементами которой являются инфинитезимальные коэффициенты $\nu_{\alpha,\beta}$, $\alpha \in \Omega_i,\ \beta \in \Omega_j$. Так что Q_{ij} — прямоугольная $(\omega_i \times \omega_j)$−матрица, у которой число строк совпадает с числом номеров в подмножестве Ω_i, а число столбцов совпадает с числом номеров в подмножестве Ω_j. Тогда уравнения равновесия (1.35) примут вид

$$\pi(0)Q_{00} + \pi(1)Q_{10} = 0, \tag{1.36}$$

$$\pi(i-1)Q_{i-1,i} + \pi(i)Q_{i,i} + \pi(i+1)Q_{i+1,i} = 0, \quad 1 \le i < n, \tag{1.37}$$

$$\pi(n-1)Q_{n-1,n} + \pi(n)Q_{n,n} = 0. \tag{1.38}$$

Эти уравнения обобщают рекуррентное соотношение (1.30), полученное ранее для процесса размножения и гибели, и являются эквивалентными (1.14).

Использование (1.36)–(1.38) вместо (1.14) может оказаться более предпочтительным по следующим причинам. Решение системы (1.10), (1.14) связано с обращением матрицы порядка ω. И если ω велико, то это может привести к существенным вычислительным погрешностям. При использовании (1.36)–(1.38) обращать приходится матрицы в среднем порядка ω/n, что можно делать точнее и быстрее. Короче говоря, удобнее пользоваться (1.36)–(1.38), а не (1.14), если проблема обращения n матриц порядка ω/n решается точнее и быстрее, чем обращение одной матрицы порядка ω. Если это так, то поступают следующим образом. Заметим, что для эргодических СМО матрицы $Q_{ii},\ 1 \le i \le n$, являются невырожденными. Пользуясь уравнениями (1.37) как рекуррентными соотношениями, можно получить формулы

$$\pi(i+1) = \pi(i)D_{i+1}, \quad 0 \le i \le n-1, \tag{1.39}$$

где матрица D_{i+1} находится из рекуррентного соотношения

$$D_{i+1} = -D_{i,i+1}(Q_{i+1,i+1} + D_{i+2}Q_{i+2,i+1})^{-1},\ D_{n+1} = 0. \tag{1.40}$$

Вектор−строка $\pi(\mathbf{0})$ находится из системы уравнений

$$\pi(0)(Q_{0,0} + D_1Q_{1,0}) = 0, \quad \pi(0)\left(I + \sum_{k=1}^{n}\prod_{j=1}^{k} D_j\right)\mathbf{E} = 1 \tag{1.41}$$

Алгоритм вычислений при определении стационарных вероятностей сводится к следующему.

1. По формулам (1.40) находят $D_j,\ 1 \le j \le n$.

2. Вычисляют $\pi(0)$ из системы уравнений (1.41).

3. Находят $\pi(i),\ 1 \le i \le n$ по формулам (1.39).

4. Проверяют условия
$$\pi_i \ge 0,\ i \in \Omega, \quad \sum_{i\in\Omega}\pi_i = 1. \tag{1.42}$$

Если условия (1.42) выполняются с достаточной точностью, процесс вычислений стационарных вероятностей завершается. Если же условия не выполняются с достаточной точностью, то это означает, что алгоритм решения системы (1.41) либо алгоритм обращения матриц (1.40) недостаточно точны. В этом случае необходимо использовать более точные алгоритмы.

Понятие векторного процесса размножения и гибели может быть использовано для численного анализа СМО, не являющихся марковскими, но такими, что, образно говоря, немарковские состояния СМО могут быть представлены в виде совокупности марковских состояний. Это означает, что путем увеличения размерности пространства состояний СМО можно добиться того, чтобы поведение СМО в новом пространстве состояний описывалось цепью Маркова. Такое преобразование можно осуществить, в частности, для СМО $A|B|m|n$, у которой распределения A и B относятся к классу эрланговских E_k распределений (см. (1.2)) или гиперэкспоненциальных H_k (см. (1.3)) распределений. Проиллюстрируем это на примерах.

Из (1.1) и (1.2) следует, что распределение E_k — это свертка k показательных распределений M.

Поэтому если входящий поток (*процесс обслуживания*) характеризуется распределением E_k, то интервал между поступлениями требований (*время обслуживания*) является суммой k одинаково распределенных по показательному закону случайных интервалов времени, называемых *этапами*. В связи с этим будем говорить, что входящий поток (процесс обслуживания) находится в состоянии i, если в данный момент времени продолжается i—й этап. Нумерацию этапов удобно осуществить для входящего потока и процесса обслуживания по разному следующим образом. Для входящего потока начало первого этапа совпадает с поступлением в СМО предыдущего требования. Окончание i—го этапа совпадает с началом $(i+1)$—го этапа. Этап k заканчивается поступлением требования. Для процесса обслуживания начало k—го этапа совпадает с поступлением требования на обслуживание. Окончание i—го этапа совпадает с началом $(i-1)$—го этапа. Окончание первого этапа совпадает с окончанием обслуживания.

Таким образом, для СМО $E_k|E_l|m|n$ событие, заключающееся в том, что в СМО находится r требований (которые мы ранее в простых случаях называли состоянием СМО) расчленяется на kl частных событий, каждое из которых и образует состояние СМО. Эти события будем обозначать тройкой (i,j,r), где i — номер этапа для входящего потока, $1 \le i \le k$; j — номер этапа для процесса обслуживания, $1 \le j \le l$; r — число требований в СМО, $0 \le r \le m+n$.

На рис. 1.2—1.4 для иллюстрации приведены графы переходов для СМО $M|E_4|1|5$, $E_4|M|1|5$, $E_3|E_2|1|1$ (напомним, что E_1 является показательным распределением, то есть эквивалентным M). Графы переходов показывают, что разбиение пространства возможных состояний Ω на уровни следует производить следующим образом: на уровень Ω_s помещают все состояния (i,j,r), для которых $r=s$. Таким образом, для СМО $E_k|E_l|m|n$ образуется $(m+n+1)$ уровней $\Omega_r,\ 0 \le r \le m+n$. Причем на уровень Ω_0 попадает k состояний, $\Omega_0 = \{(i,l,r)\,|\,1 \le i \le k\}$, а на уровени Ω_r — по kl состояний, $\Omega_r = \{(i,l,r)\,|\,1 \le i \le k,\ 1 \le j \le l\},\ 1 \le r \le m+n$.

Для того чтобы сформулировать конкретные результаты, необходимо еще установить способ упорядочения состояний. Ниже принят следующий способ упорядочения. Пусть $N(i,j,r)$ обозначает номер состояния (i,j,r). Тогда:

1) если $r < r'$, то для любых i,j,i',j' $N(i,j,r) < N(i',j',r')$.

2) если $i < i'$, то для любых j,j',r $N(i,j,r) < N(i',j',r')$.

3) если $j < j'$, то для любых i,r $N(i,j,r) < N(i,j',r)$.

Примеры немарковских СМО, описываемых векторными процессами размножения и гибели. Определим конкретный вид матричных блоков Q_{rs}, определяющих уравнения равновесия (1.36)—(1.38) для некоторых СМО, не являющихся марковскими.

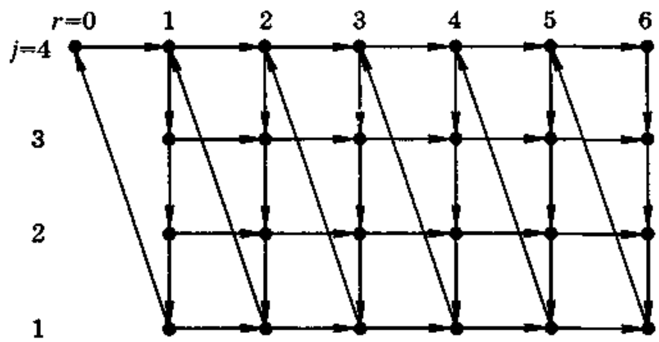
СМО $\mathbf{M|E_l|1|n}$ (см рис. 1.2)

Поскольку на уровне $Omega_0$ всего одно состояние, Q_{00} вырождается в скалярную величину $Q_{00} = -\lambda$.

$\mathbf{Q_{10}}$ является вектор—столбцом с l компонентами, которые определяются формулой

$$(\mathbf{Q_{10}})_j = l\mu\delta_{j1},\ 1 \le j \le l. \tag{1.43}$$

$\mathbf{Q_{01}}$ — это вектор—строка с l компонентами



$$(\mathbf{Q_{01}})_j = \lambda\delta_{jl},\ 1 \le j \le l. \tag{1.44}$$

Матрицы $Q_{ii},\ 1 \le i \le l$ одинаковы. Их элементы определяются равенствами

$$(Q_{ii})_{\alpha,\beta} = -(l\mu + \lambda)\delta_{\alpha,\beta} + l\mu\delta_{\alpha,\beta+1}, \quad 1 \le \alpha, \beta \le l. \tag{1.45}$$

$$l\mu(\delta_{\alpha,\beta+1} - \delta_{\alpha,\beta})(Q_{n+1,n+1})_{\alpha,\beta} = -l\mu\delta_{\alpha,\beta} + l\mu\delta_{\alpha,\beta+1}, \quad 1 \le \alpha, \beta \le l.$$

Матричные блоки $Q_{i-1,i} = \lambda I_l$, где I_l — единичная матрица порядка l, для всех $i > 1$ $(2 \le i \le n+1)$.

Матричные блоки $Q_{i+1,i}$ также одинаковы для всех $i \le 1, \quad (1 \le \le i \le n)$ с элементами

$$(Q_{i+1,i})_{\alpha,\beta} = l\mu\delta_{1\alpha}\delta_{l\beta}, \quad 1 \le \alpha, \beta \le l. \tag{1.46}$$

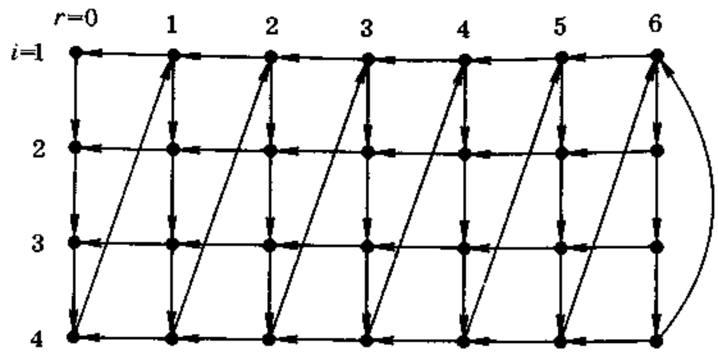
Таким образом, решение в виде (1.39) получается вычислением $(n+1)$ матриц D_i по формулам (1.40), предполагающим обращение матриц порядка l каждая. Заметим, что D_1 вырождается в вектор–строку. Поэтому система (1.41) превращается в скалярное соотношение для вычисления вероятности $\pi(0)$:

$$\pi(0) = \left(1 + \sum_{k=1}^{n+1} D_1 \left(\prod_{j=2}^{k} D_j\right) \mathbf{E}\right)^{-1}. \tag{1.47}$$

При использовании общих соотношений (1.10), (1.14) приходится сталкиваться с проблемой обращения матрицы порядка $(n+1)l+1$. В (1.14) матрица Q, составленая из матричных блоков $Q_{i,j},\ 0 \le i,j \le n+1$, имеет трехдиагональную структуру: у нее отличными от нуля являются элементы, стоящие на главной диагонали, на диагонали под главной и на диагонали, находящейся над главной через i элементов. В таб. 1.1 для примера представлена матрица Q для системы $M|E_3|1|2$.

Таблица 1.1

Матрица Q для СМО $\mathbf{M|E_3|1|2}$.



$$
\begin{array}{c@{\;}c}
 & \begin{matrix} 30 & 11 & 21 & 31 & 12 & 22 & 32 & 13 & 23 & 33 \end{matrix} \\
\begin{matrix} 30 \\ 11 \\ 21 \\ 31 \\ 12 \\ 22 \\ 32 \\ 13 \\ 23 \\ 33 \end{matrix} &
\begin{pmatrix}
-\lambda & & & \lambda & & & & & & \\
\mu & -(\mu+\lambda) & & & \lambda & & & & & \\
 & \mu & -(\mu+\lambda) & & & \lambda & & & & \\
 & & \mu & -(\mu+\lambda) & & & \lambda & & & \\
 & & & \mu & -(\mu+\lambda) & & & \lambda & & \\
 & & & & \mu & -(\mu+\lambda) & & & \lambda & \\
 & & & & & \mu & -(\mu+\lambda) & & & \lambda \\
 & & & & & & \mu & -\mu & & \\
 & & & & & & & \mu & -\mu & \\
 & & & & & & & & \mu & -\mu
\end{pmatrix}
\end{array}
$$

СМО $\mathbf{E_k|M|1|n}$ (рис. 1.3)

На каждом $\Omega_r,\ 0 \le r \le n+1$, находится по k состояний, поэтому
все матрицы Q_{ij} квадратные порядка k.
Q_{00} имеет элементы

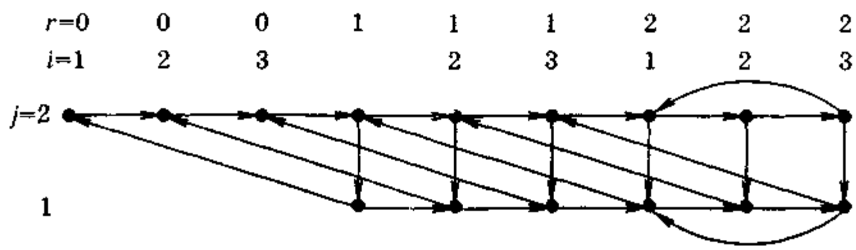
$$k\lambda(\delta_{\alpha+1,\beta} - \delta_{\alpha,\beta})(Q_{00})_{\alpha,\beta} = k\lambda(\delta_{\alpha,\beta} + \lambda\delta_{\alpha,\beta-1}),\ 1 \le \alpha, \beta \le k. \tag{1.48}$$

Матрицы $Q_{ii},\ 1 \le i \le n$, одинаковы. Их элементы определяются равенствами

$$(Q_{ii})_{\alpha,\beta} = k\lambda(\delta_{\alpha,\beta-1} - (k\lambda+\mu)\delta_{\alpha,\beta}),\ 1 \le \alpha, \beta \le k. \tag{1.49}$$

Матрицы $Q_{i+1,i},\ 0 \le i \le n+1$, тоже одинаковы и являются диагональными: $Q_{i+1,i} = \mu I_k$, где I_k — единичная матрица порядка k.

Одинаковыми являются и матрицы $Q_{i-1,i},\ 0 \le i \le n$. У этих матриц только один элемент отличен от нуля:



$$(Q_{i-1,i})_{k,1} = k\lambda. \tag{1.50}$$

Наконец, матрица $Q_{n+1,n+1} = Q_{n,n} + Q_{n-1,n}$.

В этом случае решение в виде (1.39) предполагает вычисление $(n+1)$ матриц D_i по формулам (1.40), которые предусматривают обращение матриц порядка k. Вектор $\pi(\mathbf{0})$ определяется из системы (1.41), решение которой тоже предполагает обращение матриц порядка k.

При использовании общей системы (1.10), (1.14) необходимо обращать матрицу порядка $k(n+2)$. Матрица Q уравнения (1.14) представлена в таб.1.2. для частного случая $E_3|M|1|1$.

Таблица 1.2

Матрица Q для СМО $|\mathbf{E_3}|\mathbf{M}|\mathbf{1}|\mathbf{1}$

$$
\begin{array}{cc}
 & \begin{array}{ccccccccc} 10 & 20 & 30 & 11 & 21 & 31 & 12 & 22 & 32 \end{array} \\
\begin{array}{c} 10 \\ 20 \\ 30 \\ 11 \\ 21 \\ 31 \\ 12 \\ 22 \\ 32 \end{array} &
\begin{pmatrix}
-k\lambda & k\lambda & & & & & & & \\
 & -k\lambda & k\lambda & & & & & & \\
 & & -k\lambda & k\lambda & & & & & \\
\mu & & & -\mu-k\lambda & k\lambda & & & & \\
 & \mu & & & -\mu-k\lambda & k\lambda & & & \\
 & & \mu & & & -\mu-k\lambda & k\lambda & & \\
 & & & \mu & & & -\mu-k\lambda & k\lambda & \\
 & & & & \mu & & & -\mu-k\lambda & k\lambda \\
 & & & & & \mu & k\lambda & & -\mu-k\lambda
\end{pmatrix}
\end{array}
$$

СМО $\mathbf{E_k}|\mathbf{E_l}|\mathbf{1}|\mathbf{n}$ (рис. 1.4)

Структуру матриц в этом случае удобно описать, отталкиваясь от предыдущего, так как отличие состоит лишь в том, что состояние предыдущей СМО расщепляется в рассматриваемом случае на l различных состояний в зависимости от того, какой этап обслуживания реализуется. Формально это означает, что ненулевые элементы матриц Q предыдущей СМО превращаются в ненулевые квадратные матричные блоки порядка l. Нулевые элементы соответственно превращаются в нулевые матричные блоки такого же размера. Для удобства введем следующие обозначения

$$U = (U_{\alpha\beta}), \quad U_{\alpha\beta} = l\mu\delta_{1\alpha}\delta_{l\beta},\ 1 \le \alpha\beta \le l. \tag{1.51}$$

$$V = (V_{\alpha\beta}), \quad V_{\alpha\beta} = l\mu\delta_{\alpha,\beta+1} - (l\mu + k\lambda)\delta_{\alpha\beta},\ 1 \le \alpha\beta \le l. \tag{1.52}$$

$$W = k\lambda I_l. \tag{1.53}$$

U_0 — $((kl) \times k)$—матрица с элементами

$$(U_0)_{\alpha\beta} = l\mu\delta_{\alpha,(\beta-1)l+1},\ 1 \le \alpha \le kl,\ 1 \le \beta \le k. \tag{1.54}$$

V_0 — $(k \times k)$—матрица с элементами

$$(V_0)_{\alpha,\beta} = k\lambda(\delta_{\alpha,\beta-1} - \delta_{\alpha,\beta}),\ 1 \le \alpha\beta \le k. \tag{1.55}$$

W_0 — $((kl) \times (kl))$—матрица с элементами

$$(W_0)_{\alpha\beta} = k\lambda\delta_{\alpha k}\delta_{\beta l},\ 1 \le \alpha \le k,\ 1 \le \beta \le kl. \tag{1.56}$$

В СМО $E_k|E_l|1|n$ одновременно может находиться от 0 до $(n+1)$ требований. Поэтому пространство состояний этой СМО расчленяется на $(n+2)$ уровня, причем на уровне Ω_0 находится k состояний, а на остальных $(n+1)$ уровнях Ω_r находятся по kl состояний. Пользуясь введенными обозначениями, матричные блоки Q_{ij}, определяющие уравнения равновесия (1.36)—(1.38), можно представить в виде

$$Q_{00} = V_0, \quad Q_{01} = W_0, \quad Q_{10} = U_0. \tag{1.57}$$

$Q_{r,r}$, $1 \le r \le n$, — $(k \times k)$—матрица, элементами которой являются $(l \times l)$—матричные блоки, определяемые по формуле

$$\left(Q_{r,r}\right)_{\alpha,\beta} = \delta_{\alpha,\beta}V + \delta_{\alpha,\beta-1}W,\ 1 \le \alpha,\beta \le k. \tag{1.58}$$

$(Q_{r,r+1})$, $1 \le r \le n$, — $(k \times k)$—матрица, элементами которой являются $(l \times l)$—матричные блоки, определяемые по формуле

$$\left(Q_{r,r+1}\right)_{\alpha,\beta} = \delta_{k\alpha}\delta_{1\beta l}W,\ 1 \le \alpha,\beta \le k. \tag{1.59}$$

$(Q_{r,r-1})$, $1 \le r \le n+1$, — $(k \times k)$—матрица, элементами которой являются $(l \times l)$—матричные блоки, определяемые по формуле

$$\left(Q_{r,r-1}\right)_{\alpha,\beta} = \delta_{\alpha,\beta}U,\ 1 \le \alpha,\beta \le k. \tag{1.60}$$

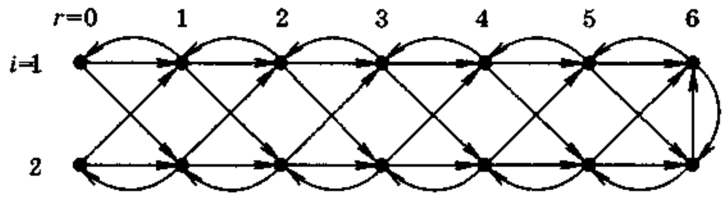
Наконец,

$$Q_{n+1,n+1_{\alpha,\beta}} = Q_{n,n} + Q_{n,n+1}. \tag{1.61}$$

Таким образом, матричные блоки Q_{ij}, $1 \le i,j \le n+1$, являются квадратными размера $(kl \times kl)$.

Получение решения в виде (1.39) будет предпочтительнее, чем решение системы (1.14), если обращение $(n+1)$ матрицы порядка kl можно выполнить точнее и быстрее, чем обращение одной матрицы порядка $(k + kl(n+1))$.

СМО $\mathbf{M|H_l|1|n}$ (рис. 1.5)



В этом случае процесс обслуживания характерезуется гиперэкспоненциальным распределением (см. (1.3)). Такое обслуживание можно интерпретировать как экспоненциальное обслуживание, интенсивность которого с вероятностью β_j имеет значение μ_j, $1 \le j \le l$.

Нулевой уровень Ω_0 в этом случае содержит только одно состояние, а остальные $(n+1)$ уровни Ω_r — по l состояний. Матричный блок Q_{00} поэтому вырождается в скалярную величину $Q_{00} = -\lambda$.

Q_{01} является строкой:

$$Q_{01} = \lambda(\beta_1 \beta_2 \dots \beta_l). \tag{1.62}$$

$\mathbf{Q_{10}}$ — вектор–столбец размера l:

$$Q^*_{10} = (\mu_1 \mu_2 \dots \mu_l). \tag{1.63}$$

Для удобства введем обозначения: μ — вектор–столбец, составленный из интенсивностей обслуживания μ_j, $1 \le j \le l$. β — вектор–строка, составленная из вероятностей β_j, $1 \le j \le l$. Тогда можно записать, что

$$(Q_{01}) = \lambda\beta, \quad \mathbf{Q_{10}} = \mu. \tag{1.64}$$

Остальные матрицы определяются равенствами

$$Q_{i-1,1} = \lambda I_l, \;\; 2 \le i \le n+1, \tag{1.65}$$

$$Q_{i,i} = -\mathrm{diag}\{(\mu_j + \lambda), \, 1 \le i \le l\}, \;\; 1 \le i \le n+1, \tag{1.66}$$

$$Q_{i+1,i} = \mu\beta, \;\; 1 \le i \le n. \tag{1.67}$$

Таким образом, чтобы получить решение в виде (1.39) нужно воспользоваться (1.64)–(1.67) для вычисления последовательности матриц D_i при помощи рекуррентного соотношения (1.40); $\pi(0)$ вычисляется по формуле (1.47).

СМО $\mathbf{H_k|M|1|n}$ (рис. 1.6)

Состояния этой СМО группируются на $(n+2)$ уровнях по k состояний в каждом. Размерность задачи (1.10), (1.14) в этом случае равна $k(n+2)$. Обозначим α — вектор–строку с компонентами α_i, $1 \le i \le k$; λ — вектор–столбец с компонентами λ_i, $1 \le i \le k$. Матричные блоки Q_{ij} определяются равенствами

$$Q_{00} = -\mathrm{diag}\{\lambda_i, \, 1 \le i \le k\},$$

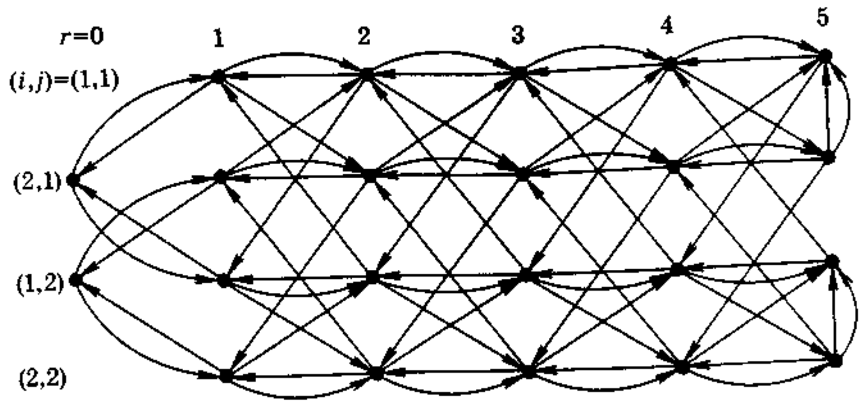
$$Q_{01} = -Q_{00}.$$

$$Q_{jj} = -\mathrm{diag}\{(\mu + \lambda_i), \, 1 \le i \le k\}, \, 1 \le j \le n, \tag{1.68}$$

$$Q_{n+1,n+1} = Q_{n,n} + \lambda\alpha,$$

$$Q_{j+1,j} = \mu I_k, \, 0 \le j \le n, \tag{1.69}$$

$$Q_{j,j+1} = \lambda\alpha, \, 1 \le i \le n. \tag{1.70}$$



Стационарные вероятности определяются аналогично тому, как это делалось в предыдущих случаях.

СМО $\mathbf{H_k|H_l|1|n}$ (рис. 1.7)

Множество возможных состояний из $(k+kl(n+1))$ элементов, которые естественным образом группируются на $(n+2)$ уровнях. На нулевом уровне Ω_0 содержится k состояний, а на остальных по kl состояний.

Матричные блоки Q_{ij} здесь удобно записать с использованием кронекерова произведения матриц. Если $A = (a_{ij})$ — матрица размера $(m \times n)$, а $B = (b_{ij})$ — матрица размера $(p \times q)$, то кронекеровым произведением $C = A \otimes B$ этих матриц называется матрица размера $(mp \times nq)$, имеющая структуру

$$C = A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \dots & \dots & \dots & \dots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}.$$

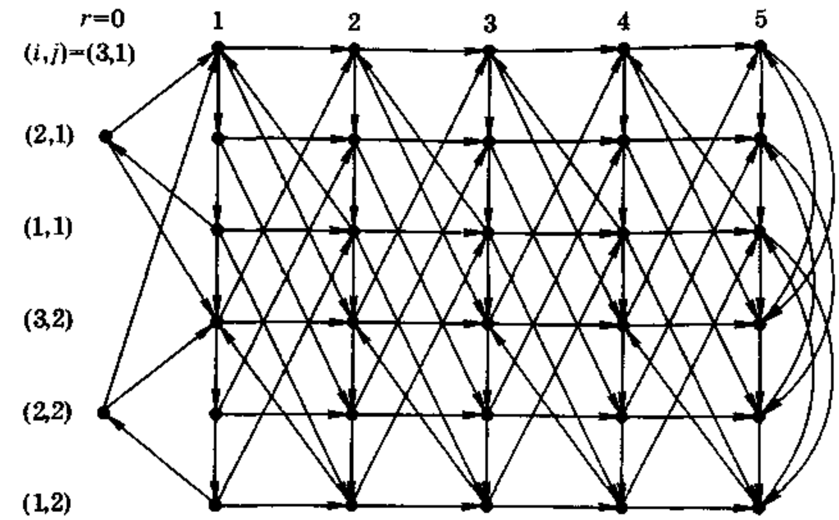
Пусть, как и прежде α и β — вектора–строки: $\alpha = = (\alpha_1\alpha_2 \dots \alpha_k)$, $\beta = (\beta_{\mathbf{1}}\beta_{\mathbf{2}} \dots \beta_{\mathbf{l}})$, составленные из вероятностей, определяющих распределения H_k и H_l входящего потока и процесса обслуживания, соответственно; λ и μ — вектора–столбцы $\lambda^* = (\lambda_1\lambda_2 \dots \lambda_k)$, $\mu^* = (\mu_1\mu_2 \dots \mu_l)$, составленные из интенсивностей, определяющих H_k и H_l. Тогда матричные блоки Q_{ij} СМО $H_k|H_l|1|n$ определяются равенствами

$$Q_{00} = -\mathrm{diag}\{\lambda_i,\ 1 \le i \le k\},$$

$$Q_{01} = \alpha \otimes (\lambda\beta),$$

$$Q_{10} = I_k \otimes \mu,$$

$$Q_{r,r} = -\mathrm{diag}\{(\mu_j + \lambda_i),\ 1 \le (i-1)l + j \le kl,\ 1 \le i \le k,\ 1 \le j \le l\}, \qquad (1.71)$$



$$1 \le r \le n,$$

$$Q_{r+1,r} = I_k \otimes (\mu\beta), 1 \le r \le n, \tag{1.72}$$

$$Q_{r,r+1} = (\lambda\alpha) \otimes I_l,\ 1 \le r \le n, \tag{1.73}$$

$$Q_{n+1,n+1} = Q_{n,n} + Q_{n,n+1}.$$

Знание матричных блоков Q_{ij} позволяет наиболее рациональным образом построить процесс вычисления стационарных вероятностей в соответствии с описанными выше подходами.

СМО $\mathbf{E_k|H_l|1|n}$ (рис. 1.8)

Число возможных состояний такой системы равно $(k + kl(n+1))$.
На нулевом уровне Ω_0 содержится k состояний, на остальных уровнях Ω_r содержится по kl состояний.

При описании матричных блоков рассматриваемой СМО под μ и β будем понимать вектора с l компонентами, которые использовались при анализе предыдущей СМО; λ — скалярный параметр. Матричные блоки Q_{ij} определяются следующим образом:

Q_{00} — $(k \times k)$—матрица с элементами

$$(Q_{00})_{uv} = \lambda(\delta_{u,v+1} - \delta_{u,v}),\ \ 1 \le u, v \le k.$$

Q_{01} — $(k \times (kl))$—матрица с элементами

$$(Q_{01})_{uv} = \lambda(\delta_{1,u} \sum_{j=1}^{l} \beta_j \delta_{v,(k-1)l+j}),\ \ 1 \le u \le k,\ 1 \le v \le (kl).$$

Q_{10} — $((kl) \times k)$—матрица, определяемая равенством

$$Q_{10} = I_k \otimes \mu.$$

Остальные матрицы Q_{ij} являются квадратными $((kl) \times (kl))$—матрицами с элементами

$$(Q_{r-1,r})_{u,v} = \lambda \sum_{j=1}^{i} \delta_{v,(k-1)l+j},\ 1 \le u, v \le (kl),\ 2 \le r \le n+1. \tag{1.74}$$

$$(Q_{r,r})_{u,v} = \lambda \delta_{u,v+l} - \delta_{u,v} \sum_{i=1}^{k} \sum_{j=1}^{l} (\lambda + \mu_j) \delta_{u,(i-1)l+j},\ 1 \le u, v \le (kl),\ 1 \le r \le n. \tag{1.75}$$

$$Q_{r+1,r} = I_k \otimes (\mu\beta),\ 1 \le r \le n, \quad Q_{n+1,n+1} = Q_{n,n} + Q_{n,n+1}. \tag{1.76}$$

Использование матриц Q_{ij} позволяет построить процедуру вычисления стационарных вероятностей по формулам (1.39)–(1.41).

СМО $\mathbf{H_k|E_l|1|n}$ (рис. 1.9)

Число возможных состояний в этом случае, как и в предыдущем, равно $(k+kl(n+1))$. Распределение этих состояний по уровням в количественном соотношении тоже совпадает с предыдущим.

Под λ и α здесь понимается вектор—столбец и вектор—строка порядка k, соответственно; μ — скалярный параметр. Матричные блоки Q_{ij} определяются следующим образом

Q_{00} — $(k \times k)$—матрица с элементами

$$(Q_{00})_{uv} = -\lambda_u \delta_{u,v},\ 1 \le u, v \le k.$$

Q_{01} — $(k \times (kl))$—матрица с элементами

$$(Q_{01})_{uv} = \lambda_u \sum_{i=1}^{k} \alpha_i \delta_{v,il}.$$

Q_{10} — $((kl) \times k)$—матрица с элементами

$$(Q_{10})_{u,v} = \mu \sum_{i=1}^{k} \delta_{v,i} \delta_{u,1+l(i-1)}.$$

Остальные матричные блоки Q_{ij} являются квадратными матрицами порядка kl с элементами

$$(Q_{r+1,r})_{u,v} = \mu \sum_{i=1}^{k} \delta_{u,(i-1)l+1} \delta_{v,il},\ 1 \le u, v \le (kl),\ 1 \le r \le n. \tag{1.77}$$

$$(Q_{r,r})_{u,v} = \mu \delta_{u,v+l} - \delta_{u,v} \sum_{i=1}^{k} \sum_{j=1}^{l} (\lambda_i + \mu) \delta_{u,(i-1)l+j},\ 1 \le u, v \le (kl),\ 1 \le r \le n. \tag{1.78}$$

$$Q_{r,r+1} = (\lambda \alpha) \otimes I_l,\ 1 \le r \le n, \quad Q_{n+1,n+1} = Q_{n,n} + Q_{n,n+1}. \tag{1.79}$$

На основании этих соотношений строится вычислительная процедура определения стационарных вероятностей состояний рассматриваемой СМО.

Системы с несколькими приборами. Итак, мы рассмотрели весь набор СМО типа $A|B|1|n$ с одним обслуживающим прибором, где распределения A и B выбираются из множества $\{M, E_k, H_k\}$. (Напомним, что M эквивалентно распределениям E_1 и H_1.) Изучение СМО с несколькими приборами обслуживания принципиальных затруднений не вызывает, но при этом увеличивается размерность задачи. Это происходит потому, что различные приборы в некоторый фиксированный момент времени могут находиться в различных состояниях обслуживания. Для E_l — это различные этапы обслуживания, для H_l — это различные интенсивности обслуживания. Предположим, что в СМО имеется m неразличающихся приборов обслуживания. Тогда состояние совокупности этих приборов можно описать многомерной переменной $j = (j_0, j_1, \ldots, j_l)$, где j_0 — число свободных приборов; j_r — число приборов, находящихся в состоянии r, $1 \le r \le l$. Понятно, что каждая из переменных j_r может принимать $(m+1)$ значение $0 \le j_r \le m$, но таким образом, чтобы $j_0 + j_1 + \ldots + j_l = m$. Количество состояний, в которых может находиться рассматриваемая совокупность приборов, достаточно быстро растет с ростом m. Оно равно C_{m+l-1}^{m}. С увеличением числа состояний усложняется и проблема вычисления элементов в матричных блоках Q_{ij}.

В таб. 1.3 для ориентировки указаны значения величины C_{m+l-1}^{m} для нескольких значений l и m.

Таблица 1.3

l	m							
	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	2	3	4	5	6	7	8	9
3	3	6	10	15	21	28	36	45
4	4	10	20	35	56	84	120	165
5	5	15	35	70	126	210	285	450
6	6	21	56	126	252	462	747	1197
7	7	28	84	210	462	924	1671	2868
8	8	36	120	285	747	1671	3342	6210

Вместе с тем, когда $l = 1$, то есть E_l и H_l вырождаются в M $E_k|M|m|n$ и $H_k|M|m|n$ описываются теми же матрицами Q_{ij}, что и в случае $m = 1$, за тем лишь исключением, что в выражениях для элементов матрицы Q_{ij}, вместо параметра μ необходимо подставить значение $\mu' = \mu \min\{i, m\}$.

В общем случае $l > 1$ структура матриц Q_{ij} достаточно сложная. Общее число состояний СМО определяется суммой $\omega = \omega_0 + \omega_1 + \ldots + \omega_{m+n}$, где ω_r — число состояний на уровне Ω_r. Приведем значения этих чисел

$$\omega_0 = \begin{cases} 1 & \text{для входящего потока } H_k, \\ k & \text{для входящего потока } E_k, \end{cases}$$

$$\omega_r = \begin{cases} k\, C_{r+l-1}^{r} & \text{для } 1 \le r < m, \\ k\, C_{m+l-1}^{m} & \text{для } m \le r \le m+n. \end{cases}$$

Таким образом, размер матрицы Q_{ij}, определяющей интенсивности переходов из состояний уровня ω_i в состояние уровня ω_j, характеризуется парой $(\omega_i \times \omega_j)$ и быстро растет с ростом m. Правда, часто матрицы Q_{ij} содержат достаточно много нулевых элементов. Приведем сведения, позволяющие определить элементы матриц Q_{ij}.

Элементы матричных блоков $Q_{r,r+1}$ характеризуют интенсивности переходов из состояния (i, j, r) уровня Ω_r в состояние (i', j, r) уровня Ω_{r+1}. При входящем потоке E_k ненулевые элементы равны $k\lambda$, если $i = 1$, а $i' = k$. Остальные элементы — нулевые. При входящем потоке H_k ненулевые элементы равны $\lambda_i \alpha_{i'}$, $1 \le i, i' \le k$. Остальные элементы — нулевые.

Элементы матричных блоков $Q_{r,r-1}$ характеризуют интенсивности переходов из состояния (i, j, r) уровня Ω_r в состояние $(i, j', r-1)$ уровня Ω_{r-1}. Напомним, что $j = (j_0, j_1, \ldots, j_l)$, $j' = (j'_0, j'_1, \ldots, j'_l)$. Когда обслуживание происходит в соответствии с E_l, интенсивность перехода из (i, j', r) в $(i, j', r-1)$ ненулевая при выполнении следующих условий:

$$j_1 - j'_1 = 1, \;\; i_u = i'_u, \;\; 2 \le u \le l-1;$$

$$j'_0 = \begin{cases} j_0 + 1, & \text{если } r \le m, \\ j_0, & \text{если } r > m; \end{cases}$$

$$j'_l = \begin{cases} j_l + 1, & \text{если } r > m, \\ j_l, & \text{если } r \le m. \end{cases}$$

Если эти условия выполнены, то соответствующий элемент матрицы $Q_{r,r-1}$ равен $(j_1 l \mu)$. Когда обслуживание происходит в соответствии с H_l, интенсивность перехода из (i, j, r) в $(i, j', r-1)$ ненулевая при выполнении условий

$$|j_u - j'_u| \le 1, \; 0 \le u \le l;$$

$$\sum_{u=0}^{l} (j_u - j'_u) = 0, \; \sum_{u=0}^{l} |j_u - j'_u| = 2.$$

Пусть эти условия выполняются в следующем варианте: $j_u - j'_u = 1$; $j_v - j'_v = -1$; $j_w - j'_w = 0$ для всех $w \neq u$ или $w \neq v$. Тогда интенсивность соответствующего перехода, то есть соответствующий элемент матриц $Q_{r,r-1}$ равен (j_u, μ_u, β_v). При невыполнении указанных условий соответствующие элементы матриц $Q_{r,r-1}$ равны нулю.

Матричные блоки $Q_{r,r}$ являются квадратными для любых $r,\ 0 \leq r \leq m+n$. Элементы главной диагонали этих матричных блоков являются отрицательными и определяются условием: сумма столбцов матриц $Q_{r,r-1}, Q_{r,r}$ и $Q_{r,r+1}$ равна нулевому столбцу. При определении элементов матриц $Q_{r,r-1}, Q_{r,r}, Q_{r,r+1}$ до сих пор было неважно, каким образом упорядочиваются состояния на уровнях Ω_r. Для определения недиагональных элементов матрицы $Q_{r,r}$ удобно упорядочить состояния. Ранее было сформулировано правило нумерации для случая, когда j является скалярным параметром (см. стр.28). Его можно не менять, полагая, что неравенство $j < j'$ означает, что

$$\sum_{k=0}^{l} j_k m^k < \sum_{k=0}^{l} j'_k m^k.$$

Такой способ нумерации обеспечивает следующее свойство матрицам $Q_{r,r}$ и вообще матрице Q: элементы, стоящие над главной диагональю, зависят только от свойств входящего потока, а элементы, стоящие под главной диагональю, зависят только от свойств процесса обслуживания. Исключение составляют лишь элементы матричного блока $Q_{m+n,m+n}$, который определяется как сумма $Q_{m+n-1,n+m-1} + Q_{n+m-1,n+m}$. В этом блоке элементы, стоящие под главной диагональю, могут зависеть от свойств входящего потока.

Недиагональные элементы матричных блоков $Q_{r,r}$, $1 \leq r \leq \leq n+m-1$, определяются следующим образом. Для систем $H_k|H_l|m|n$ и $H_k|E_l|m|n$ элементы, стоящие под главной диагональю, равны нулю, то есть для этих СМО $(Q_{r,r})_{u,v} = 0$ для всех $u < v$, $1 \leq r < n+m$.

Для систем $H_k|H_l|m|n$ и $E_k|H_l|m|n$ элементы, стоящие над главной диагональю равны нулю, то есть $(Q_{r,r})_{u,v} = 0$ для всех $u > v,\ 1 \leq r < n+m$.

Для систем $E_k|E_l|m|n$ и $H_k|E_l|m|n$ элементы, стоящие под главной диагональю и характеризующие интенсивность перехода из состояния (i, j, r) в состояние (i, j', r), будут ненулевыми только в том случае, если переменные j и j' обладают свойством: существует единственный индекс $u,\ 0 \leq u \leq l$, такой что $j_u - j'_u > 0$; для остальных индексов $j_u - j'_u \leq 0$. В этом случае соответствующий элемент матрицы равен $(j_u l \mu)$. Остальные элементы, стоящие над главной диагональю, нулевые.

Для систем $E_k|E_l|m|n$ и $E_k|H_l|m|n$ элементы, стоящие над главной диагональю и характеризующие интенсивность перехода из состояния (i, j, r) в состояние $(i', j, r),\ j > j'$, будут ненулевыми только в том случае, если $i' = i + 1$. В этом случае соответствующий элемент матрицы равен $k\lambda$. Остальные элементы, стоящие над главной диагональю, нулевые.

Из описанного следует, что матрицы $Q_{r,r},\ 1 \leq r \leq n+m$, для систем $H_k|H_l|m|n$ являются диагональными. Наконец, как уже выше говорилось,

$$Q_{m+n,n+m} = Q_{m+n-1,n+m-1} + Q_{n+m-1,n+m}.$$

Этим и завершается определение элементов всех матриц Q_{ij}, определяющих уравнение равновесия.

Итак, можно сформулировать алгоритм вычисления стационарных вероятностей состояния СМО типа $A|B|m|n$, описываемых векторным процессом размножения и гибели, следующим образом.

1. На основе задания типа СМО определяется множество возможных состояний $\Omega = \{(i, j, r)\}$.

2. Множество Ω разбивается на уровни $\Omega_r,\ 0 \leq r \leq m+n$.

3. Определяются матрицы Q_{ij}, характеризующие интенсивности переходов из состояний уровня Ω_i в состояния уровня Ω_j. Матрицы Q_{ij} являются блоками, составляющими матрицу Q — матрицу инфинитезимальных коэффициентов.

4. Определяются стационарные вероятности состояния СМО при помощи уравнений (1.10), (1.14) с использованием матрицы Q или при помощи рекуррентных соотношений (1.39)–(1.41) с использованием матричных блоков $Q_{ij},\ 0 \le i, j \le m + n$.

Вычисление характеристик СМО. После того, как стационарные вероятности состояний СМО определены, можно определить среднее число $\overline{N}$ требований в СМО, среднее число $\overline{N}_q$ требований в очереди, вероятность потери требования p_l, вероятность обслуживания p_w, вероятность обслуживания без ожидания p_o. Для этого формулы (1.15)–(1.16) следует преобразовать к виду (в случае СМО $A|B|1|n$)

$$\overline{N} = \sum_{i=1}^{k} \sum_{j=1}^{l} \sum_{r=1}^{n+1} r\pi(i, j, r), \tag{1.80}$$

$$\overline{N}_q = \sum_{r=1}^{n} r \sum_{i=1}^{k} \sum_{j=1}^{l} \pi(i, j, r + 1). \tag{1.81}$$

Вероятности для СМО $E_k|B|1|n$

$$p_l = \sum_{j=1}^{l} \pi(1, j, n + 1) \,/\, \sum_{j=1}^{l} \sum_{r=0}^{n+1} \pi(1, j, r), \tag{1.82}$$

$$p_o = \pi(1, j, 0) \,/\, \sum_{j=1}^{l} \sum_{r=0}^{n+1} \pi(1, j, r). \tag{1.83}$$

(Заметим, что $\pi(1, j, 0)$ не зависит от j.)

Вероятности для СМО $H_k|B|1|n$

$$p_l = \sum_{i=1}^{k} \sum_{j=1}^{l} \pi(i, j, n + 1), \tag{1.84}$$

$$p_o = \sum_{i=1}^{k} \pi(i, j, 0). \tag{1.85}$$

(Вероятности $\pi(i, j, 0)$ не зависят от j.)

$$p_w = 1 - p_o - p_l. \tag{1.86}$$

В случае, когда используется наиболее распространенная дисциплина обслуживания в порядке очередности прибывания, распределения времени ожидания обслуживания определяется из следующих соображений. Если в системе $A|B|1|n$ поступившее требование застало СМО в состоянии $(i, j, r),\ r > 1$, то время ожидания складывается из времени дообслуживания требования, находящегося на обслуживании по типу j и суммарного времени обслуживания $(r - 1)$ требований в соответствии с распределением B. Таким образом, в каждом конкретном случае прибытия требования в СМО распределение времени ожидания характеризуется соответствующей сверткой распределения времени обслуживания. Введем для компактности изложения следующие обозначения

$$M_j(x) = \mu_j \mathrm{e}^{-\mu_j x}, \tag{1.87}$$

$$E_j(x) = \frac{l\mu(l\mu x)^{j-1}}{(j-1)!} \mathrm{e}^{-l\mu x}, \tag{1.88}$$

$$H(x) = \sum_{j=1}^{l} \beta_j \mu_j \mathrm{e}^{-\mu x}. \tag{1.89}$$

Свертку распределений $A(x)$ и $B(x)$ обозначим $A(x)_{\square}B(x)$; $r-$кратную свертку распределения $B(x)$ обозначим $\boxed{\mathrm{r}}B(x)$. Свертку распределения $A(x)$ с $r-$кратной сверткой распределения $B(x)$ обозначим

$$A(x)\boxed{\mathrm{r}}B(x), \quad A(x)\boxed{\mathrm{o}}B(x) \equiv A(x).$$

Тогда условная плотность вероятностей времени ожидания при условии, что поступившее требование застало СМО в состоянии (i, j, r), имеет вид для системы $A|H_l|1|n$

$$W(x\,|\,i, j, r) = M_j(x)\boxed{\mathrm{r\text{-}1}}H(x).$$

Для системы $A|E_l|1|n$

$$W(x\,|\,i, j, r) = E_{j+(r-1)l}(x).$$

Здесь использовано следующее свойство эрланговских плотностей

$$E_n(x)_{\square}E_\nu(x) = E_{n+\nu}(x),\ \boxed{\mathrm{r}}E_u(x) = E_{ru}(x).$$

К сожалению, свертка гиперэкспоненциальных плотностей выражается довольно сложно, и громоздкость ее выражения возрастает с увеличением r.

Таким образом, плотность вероятностей времени ожидания вычисляется для систем $E_k|E_l|1|n$ по формуле

$$W(x) = \frac{\pi(1, j, 0)\delta(x) + \sum\limits_{j=1}^{l}\sum\limits_{r=0}^{n}\pi(1, j, r+1)E_{j+rl}(x)}{\pi(1, j, 0) + \sum\limits_{j=1}^{l}\sum\limits_{r=0}^{n}\pi(1, j, r+1)}; \tag{1.90}$$

для систем $E_k|H_l|1|n$

$$W(x) = \frac{\pi(1, j, 0)\delta(x) + \sum\limits_{j=1}^{l}\sum\limits_{r=0}^{n}\pi(1, j, r+1)(M_j(x)\boxed{\mathrm{r}}H(x))}{\pi(1, j, 0) + \sum\limits_{j=1}^{l}\sum\limits_{r=0}^{n}\pi(1, j, r+1)}; \tag{1.91}$$

для систем $H_k|E_l|1|n$

$$W(x) = \sum_{i=1}^{k}\pi(i, j, 0)\delta(x) + \sum_{i=1}^{k}\sum_{j=1}^{l}\sum_{r=0}^{n}\pi(i, j, r+1)E_{j+rl}(x); \tag{1.92}$$

для систем $H_k|H_l|1|n$

$$W(x) = \left(\sum_{i=1}^{k}\pi(i, j, 0)\right)\delta(x) + \sum_{i=1}^{k}\sum_{j=1}^{l}\sum_{r=0}^{n}\pi(i, j, r+1)(M_j(x)\boxed{\mathrm{r}}H(x)). \tag{1.93}$$

Эти формулы довольно громоздки и поэтому их использование является ограниченным. Обычно их приходится применять при определении вероятности $p(T)$ того, что время ожидания будет больше, чем T:

$$p(T) = \int\limits_{T}^{\infty} W(x)\,dx. \tag{1.94}$$

Эта вероятность является одной из важных характеристик СМО. Чаще интересуются средним временем ожидания W, для которого получаются более простые формулы.

Для систем $E_k|E_l|1|n$

$$W = \frac{\sum\limits_{j=1}^{l}\sum\limits_{r=0}^{n}\pi(1,j,r+1)(r+j/l)}{\mu\pi(1,j,0)+\sum\limits_{j=1}^{l}\sum\limits_{r=0}^{n}\pi(1,j,r+1}. \tag{1.95}$$

Для систем $E_k|H_l|1|n$ удобно ввести обозначение $\mu^{-1} =$
$= \sum\limits_{j=1}^{l}\beta_j\mu_j^{-1}$. Тогда

$$W = \frac{\sum\limits_{j=1}^{l}\sum\limits_{r=0}^{n}\pi(1,j,r+1)(r+\mu\mu_j^{-1})}{\mu(\pi(1,j,0)+\sum\limits_{j=1}^{l}\sum\limits_{r=0}^{n}\pi(1,j,r+1))}. \tag{1.96}$$

Для систем $H_k|E_l|1|n$

$$W = \frac{1}{\mu}\sum_{i=1}^{k}\sum_{j=1}^{l}\sum_{r=0}^{n}\pi(i,j,r+1)(r+\frac{j}{l}). \tag{1.97}$$

Для систем $H_k|H_l|1|n$

$$W = \frac{1}{\mu}\sum_{i=1}^{k}\sum_{j=1}^{l}\sum_{r=0}^{n}\pi(i,j,r+1)(r+\frac{\mu}{\mu_j}), \tag{1.98}$$

где снова для удобства обозначено $\mu^{-1} = \sum\limits_{j=1}^{l}\beta_j\mu_j^{-1}$.

Системы с неограниченным числом мест для ожидания. Для систем $A|B|1$ основное соотношение вычислительной процедуры определения стационарных вероятностей состояний СМО реализуемо быть не может, так как в (1.40) $n=\infty$.

Вместе с тем, если существует такое i_0, что для всяких $i \geq i_0$ матричные блоки $Q_{i-1,i}, Q_{i,i}, Q_{i+1,i}$ не зависят от i, то можно построить вычислительную процедуру, позволяющую найти решение системы (1.36)–(1.37) в так называемой матрично–геометрической форме. Заметим, что указанное свойство имеет место тогда, когда на всех уровнях $\Omega_i,\ i \geq i_0$, находится одинаковое количество состояний. Это приводит к тому, что указанные выше матричные блоки являются квадратными. Для удобства обозначим для $i \geq i_0$

$$Q_{i-1,i} = U, \quad Q_{i,i} = V, \quad Q_{i+1,i} = W. \tag{1.99}$$

Тогда уравнение (1.37) примет вид

$$\pi(i-1)U + \pi(i)V + \pi(i+1)W = 0. \tag{1.100}$$

Будем искать решение этого уравнения в виде, аналогичном (1.39),

$$\pi(i) = \pi(i-1)R, \quad i \geq i_0. \tag{1.101}$$

Подставляя (1.101) в (1.100), получим соотношение

$$\pi(i-1)(U + RV + R^2W) = 0. \tag{1.102}$$

Поскольку это соотношение должно выполняться для всех $i \geq i_0$, получим для матрицы $R = (R_{ij})$ квадратное матричное уравнение

$$R^2W + RV + U = 0, \tag{1.103}$$

которое в общем случае в аналитической форме не разрешимо. На основе (1.103) можно построить процедуры итеративного вычисления матрицы R. Например, такие:

$$R(k+1) = -(U + R^2(k)W)V^{-1}, \tag{1.104}$$

$$R(k+1) = -U(V + R(k)W)^{-1}, \tag{1.105}$$

$$R(k+1) = -\frac{1}{2}(U + R^2(k)W)V^{-1} - \frac{1}{2}U(V + R(k)W)^{-1}. \tag{1.106}$$

В качестве начального приближения удобно выбрать $R(0) = 0$ — нулевую матрицу. Последовательное применение одной из рекуррентных формул (1.104)–(1.106) определяет последовательность матриц $R(k)$, сходящихся к решению уравнения (1.103).

$$\lim_{k\to\infty} R(k) = R, \quad \lim_{k\to\infty} R_{ij}(k) = R_{ij}. \tag{1.107}$$

Процедура итеративного вычисления прекращается при таком k, при котором для заданного ε выполняется неравенство

$$\max_{i,j} |R_{i,j}(k+1) - R_{i,j}(k)| \leq \varepsilon. \tag{1.108}$$

Заметим, что из анализа примеров немарковских СМО, описываемых векторными процессами рождения и гибели, приведенных выше, следует, что матричные блоки инфинитезимальных коэффициентов становятся независимыми от номера соответствующего уровня для индексов $i \geq i_0$, где i_0 определяется таб. 1.4.

Таблица 1.4

Тип СМО	$Q_{i-1,i}$	$Q_{i,i}$	$Q_{i+1,i}$
$M\|E_l\|1$	2	0	1
$E_k\|M\|1$	1	1	0
$E_k\|E_l\|1$	2	1	1
$M\|H_l\|1$	2	1	1
$H_k\|M\|1$	2	1	0
$H_k\|H_l\|1$	2	1	1
$E_k\|H_l\|1$	2	1	1
$H_k\|E_l\|1$	2	1	1

После того, как решение (1.103) при помощи (1.104)–(1.108) найдено, процедура определения стационарных вероятностей $\pi(i)$ сводится к следующему. На основе (1.36),(1.37) составляется система уравнений

$$\pi(0)Q_{00} + \pi(1)Q_{10} = 0, \tag{1.109}$$

$$\pi(i-1)Q_{i-1,i} + \pi(i)Q_{i,i} + \pi(i+1)Q_{i+1,i} = 0, \tag{1.110}$$

$$\pi(i_0-1)U+\pi(i_0)V+\pi(i_0+1)W=0. \tag{1.111}$$

Используя (1.102), преобразовываем (1.111) к виду

$$\pi(i_0-1)U+\pi(i_0)(V+RW)=0. \tag{1.112}$$

Таким образом, получается замкнутая система линейных алгебраических уравнений (1.109), (1.110),(1.112), которая является однородной. Она полностью совпадает с (1.36)–(1.38), если принять обозначения: $i_0=n,\ Q_{n-1,n}=U,\ Q_{n,n}=(V+RW)$. Поэтому процедура определения векторов $\pi(\mathbf{i}),\ 1\le i\le n=i_0$, полностью совпадает с (1.41)–(1.40).

Условие нормировки приобретает вид

$$\sum_{i=0}^{\infty}\pi(i)\mathbf{E}=\pi(0)\Bigg[\mathbf{E}+\sum_{i=1}^{i_0}\left(\prod_{j=1}^{i}D_j\right)\mathbf{E}+$$

$$+\sum_{i=i_0+1}^{\infty}\left(\prod_{j=1}^{i}D_j\right)R^{i-i_0}\mathbf{E}\Bigg]=$$

$$=\pi(0)\left(\mathbf{E}+\sum_{i=1}^{i_0-1}\left(\prod_{j=1}^{i}D_j\right)\mathbf{E}+\left(\prod_{j=1}^{i_0}D_j\right)(I-R)^{-1}\mathbf{E}\right)=1. \tag{1.113}$$

Стационарные вероятности $\pi(0)$ определяются из системы уравнений (1.109), записанной в виде

$$\pi(0)(Q_{00}+D_1Q_{10})=0, \tag{1.114}$$

и равенства (1.113).

Условие существования стационарного распределения вероятностей состояний для исследуемой СМО определяются следующим образом. Пусть вектор α является решением системы уравнений

$$\alpha(U+V+W)=0,\ \alpha\mathbf{E}=1. \tag{1.115}$$

Тогда стационарное распределение существует, если выполняется неравенство

$$\alpha U\mathbf{E}<\alpha W\mathbf{E}. \tag{1.116}$$

Таким образом, алгоритм вычисления стационарных вероятностей состояний СМО, описываемой векторным процессом размножения и гибели, с неограниченным числом мест для ожидания может быть сформулирован следующим образом.

1. Множество состояний СМО расчленяется на уровни и определяются матричные блоки $Q_{i,j}$.
2. Определяются i_0 такое, что при $i\ge i_0$ выполняются равенства (1.99).
3. Проверяются условия стационарности (1.115),(1.116).
4. При помощи (1.104)–(1.108) определяется матрица R.
5. По формуле (1.40) определяются матрицы $D_j,\ 1\le j\le i_0$. При этом считается, что $D_{i_0+1}=0$, а $Q_{i_0,i_0}=V+RW$.
6. Из систем уравнений (1.113)–(1.114) определяется вектор $\pi(\mathbf{0})$.
7. Определяются $\pi(\mathbf{i})$, $1\le i\le i_0$, по формуле (1.39).
8. Определяется необходимое количество векторов $\pi(\mathbf{i}),\ i>i_0$ по формуле (1.101).

Характеристики СМО определяются по формулам (1.80)–(1.98), в которых следует положить $n = \infty$.

Лабораторная работа: "Исследование СМО, описываемых векторными процессами размножения и гибели"

Цель работы. Освоить технику численного анализа СМО, описываемых векторными процессами размножения и гибели. Исследовать поведение характеристик таких систем в зависимости от параметров функционирования. Результаты представить в виде таблиц и графиков.

Задание. Выбрать тип СМО и вычислить характеристики (1.80)–(1.98) для выбранной СМО. Найти зависимость этих характеристик от коэффициентов загрузки ρ и числа мест для ожидания n, включая случай $n = \infty$.

Глава 2

НЕМАРКОВСКИЕ СМО

Немарковскими системами массового обслуживания будем называть системы, у которых процесс $N(t)$ — число требований в системе в момент t — *является немарковским*. Нетрудно видеть, что СМО, у которой распределение интервалов во входном потоке или (и) распределение времени обслуживания требований отличны от показательного, является немарковской. Ниже рассмотрим кратко методы исследования немарковских систем и применение этих методов для изучения некоторых нетрадиционных СМО.

2.1 Методы исследования немарковских систем

Легче поддаются исследованию однолинейные системы, у которых не распределены по показательному закону только либо длины интервалов между требованиями во входном потоке, либо времена обслуживания требований, т. е. системы типа $GI|M|1$ и $M|G|1$.

Основной прием, используемый при исследовании таких систем, состоит в построении некоторого марковского процесса $x(t)$ по распределению которого можно будет судить о распределении интересующего нас немарковского процесса $N(t)$, нахождении его характеристик, а затем и характеристик процесса $N(t)$. При этом обычно различают два подхода — внешняя и внутренняя марковизация процесса $N(t)$.

Внешняя марковизация состоит в расширении пространства состояний системы за счет введения дополнительных переменных с тем, чтобы в результате получить марковский процесс, для которого можно вывести системы дифференциальных уравнений в частных производных или интегродифференциальных уравнений для распределения вероятностей процесса. Например, для системы $M|G|1$ источником немарковости процесса $N(t)$ является зависимость поведения процесса $N(t)$ не только от его собственного значения в момент t, но и от времени, прошедшего к моменту t с момента начала обслуживания обслуживаемого в момент t требования. Поэтому естественно в качестве марковского случайного процесса $\xi(t)$ взять двумерный процесс $\{N(t), \eta(t)\}$, где $\eta(t)$ — время, прошедшее с момента начала обслуживания обслуживаемого в момент t требования, или время, оставшееся до окончания обслуживания обслуживаемого в момент t требования. Для системы $GI|M|1$ естественно выбрать в качестве процесса $x(t)$ процесс $x(t) = \{N(t), \eta(t)\}$, где $\eta(t)$ — время, прошедшее с момента прихода последнего перед моментом t требования, или время до прихода следующего требования. Недостатком внешней марковизации является сложность исследования расширенного марковского процесса $x(t)$. Достоинством является простота получения распределения процесса $N(t)$ на основе использования распределения процесса $x(t)$.

В меньшей степени этим недостатком (но и этим достоинством) обладает подход, связанный с внутренней марковизацией, называемый *методом вложенных цепей Маркова* и разработанный Кендаллом. Сущность этого метода состоит в том, что для немарковского случайного процесса $N(t)$ ищется последовательность моментов времени $\tau_n,\ n \geq 1$, в которые процесс $N(t)$ является марковским, при этом процесс $N(\tau_n)$ является дискретной цепью Маркова (см. гл.1). Методами цепей

Маркова ищется распределение вероятностей состояний цепи $N(\tau_n)$. Затем каким–либо образом по этому распределению восстанавливается распределение вероятностей состояний процесса $N(t)$. Подробнее применение метода вложенных цепей Маркова для исследования СМО будет проиллюстрировано в следующих параграфах данной главы.

Еще одним достаточно мощным и универсальным методом получения результатов является *метод введения дополнительного события*, развитый Данцигом, Кестеном, Ранненбургом, а также Климовым. Сущность этого метода и примеры его применения для анализа СМО будут проиллюстрированы далее.

В случае, когда и распределение $A(t)$ интервалов между требованиями во входном потоке, и распределение $B(t)$ времен обслуживания требований отличны от показательного, т.е. рассматривается система $G|G|1$, задача нахождения распределения вероятностей процесса $N(t)$ к настоящему моменту времени не решена. Один из возможных подходов к решению этой задачи путем аппроксимации исходной модели СМО $G|G|1$ системой типа $PH|PH|1$ затронут в предыдущей главе. Другой подход состоит в следующем.

Обозначим через t_n длину интервала между моментами поступления $(n-1)$–го и n–го требований, а через x_n — время обслуживания n–го требования. Введем в рассмотрение случайную величину $u_n = x_n - t_{n+1}$. Используя непрерывный аналог формулы полной вероятности, можно убедиться, что функция распределения $V(t)$ этой величины (она не зависит от индекса n, поскольку мы предполагаем независимость и одинаковую распределенность времен обслуживания различных требований и интервалов между требованиями) имеет вид

$$U(\tau) = P\{u_n \le \tau\} = \int\limits_0^\infty (1 - A(y-\tau))\, dB(y) = \int\limits_0^\infty B(y+\tau)\, dA(y). \tag{2.1}$$

Обозначим через w_n длительность ожидания начала обслуживания n–м требованием. Анализируя процессы поступления и обслуживания требований, можно убедиться в справедливости соотношения

$$w_{n+1} = \begin{cases} w_n + u_n, & \text{если } w_n + u_n \ge 0, \\ 0, & \text{если } w_n + u_n \le 0, \end{cases} \tag{2.2}$$

впервые полученного Линдли.

Обозначая через $W_n(y)$ функцию распределения случайной величины w_n, на основе (2.2) нетрудно убедиться в справедливости следующего соотношения

$$W_{n+1}(y) = \int\limits_{0-0}^\infty U(y-\omega)\, dW(\omega), \quad y \ge 0. \tag{2.3}$$

Далее будем предполагать, что выполняется условие $\rho = \overline{x}\,/\,\overline{t} < 1$, где $\overline{x}$ — среднее время обслуживания требования; $\overline{t}$ — средняя длина интервала между моментами поступления требований. Можно убедиться, что при выполнении этого неравенства существует стационарное распределение вероятностей величин w_n:

$$W(y) = \lim_{n\to\infty} P\{w_n \le y\}.$$

и на основе использования (2.3) можно получить интегральное уравнение для функции $W(y)$:

$$W(y) = \int\limits_{0-0}^\infty U(y-\omega)\, dW(\omega), \quad y \ge 0. \tag{2.4}$$

Очевидно, что $W(y) = 0$ для $y < 0$.

Уравнение (2.4) в литературе называют *интегральным уравнением Линдли*. Используя интегрирование по частям, можно получить две эквивалентные формы этого уравнения:

$$W(y) = \int\limits_{0-0}^{\infty} W(\omega)\, dU(y-\omega), \quad y \geq 0, \tag{2.5}$$

и

$$W(y) = \int\limits_{-\infty}^{y} W(y-u)\, dU(u), \quad y \geq 0. \tag{2.6}$$

Решив уравнение Линдли (которое является уравнением типа уравнения Винера—Хопфа), в принципе, мы получаем решение задачи нахождения распределения вероятностей времени ожидания в системе. А с использованием соотношений типа формулы Литтла (см. ниже) можем получить и моменты распределения процесса $N(t)$.

Лабораторная работа: "Нахождение распределения времени ожидания требований в системе G|G|1"

Цель работы. Овладение навыком нахождения распределения времени ожидания требований в однолинейной системе массового обслуживания с ожиданием на основе использования интегрального уравнения Линдли.

Порядок выполнения работы. По заданному виду распределений интервалов входного потока $A(t)$, времен обслуживания $B(t)$ и формуле (2.1) ищется вид распределения $U(t)$ величин u_n. Далее решается уравнение Линдли в форме (2.4), или (2.5) или (2.6) с заданным видом функций $U(t)$. Как правило, это производится путем угадывания вида функции $W(y)$ с точностью до неизвестных констант, которые затем находятся с использованием уравнения (2.4), либо (2.5), либо (2.6).

Задание 1. Найти распределение времени ожидания требования в системе $M|H_k|1$.

Задание 2. Найти распределение времени ожидания требования в системе $M|E_k|1$.

Задание 3. Найти распределение времени ожидания требования в системе $M|D|1$.

Задание 4. Найти распределение времени ожидания требования в системе $E_l|E_k|1$.

Задание 5. Найти распределение времени ожидания требования в системе $D|E_k|1$.

Задание 6. Найти распределение времени ожидания требования в системе $GI|M|1$.

На основе полученных аналитических формул построить с использованием ЭВМ графики зависимости функций $W(y)$ от аргумента y при различных значениях интенсивностей входного потока и обслуживания.

2.2 Метод введения дополнительной переменной

Рассмотрим применение этого метода к решению задачи нахождения распределения процесса $N(t)$ — числа требований в системе $M|G|1$ в момент t. Как отмечалось выше, процесс $N(t)$ является немарковским. Вводя в рассмотрение дополнительную переменную $\eta(t)$ — время, оставшееся до окончания обслуживания обслуживаемого в момент t требования, получаем двумерный процесс $\{N(t), \eta(t)\}$, являющийся марковским. Попытаемся найти стационарное распределение вероятностей этого процесса.

Введем обозначения:

$$F_t(i,x) = \mathbf{P}\{N(t) = i,\ \eta(t) < x\}, \quad i \geq 1,\ x > 0,$$
$$F_t(0) = \mathbf{P}\{N(t) = 0\}.$$

Используя обычные рассуждения, основанные на марковости процесса $\{N(t), \eta(t)\}$ и формуле полной вероятности, получаем следующую систему разностных уравнений для функций $F_t(0)$, $F_t(i,x)$:

$$F_{t+\Delta t}(0) = F_t(0)(1 - \lambda\Delta t) + F_t(1, \Delta t) + o(\Delta t),$$

$$F_{t+\Delta t}(1, x) = (F_t(1, x + \Delta t) - F_t(1, \Delta t))(1 - \lambda\Delta t) + F_t(2, \Delta t)B(x)+$$

$$+ F_t(0)\lambda\Delta t B(x) + o(\Delta t), \tag{2.7}$$

$$F_{t+\Delta t}(i, x) = (F_t(i, x + \Delta t) - F_t(i, \Delta t))(1 - \lambda\Delta t) + F_t(i + 1, \Delta t)B(x)+$$

$$+F_t(i - 1, x + \Delta t)\lambda\Delta t + o(\Delta t), \quad i \le 2.$$

Отсюда обычным образом получаем следующую систему уравнений в частных производных для функций $F_t(0), F_t(i, x)$:

$$\frac{\partial F_t(0)}{\partial t} = -\lambda F_t(0) + \frac{\partial F_t(1, 0)}{\partial x},$$

$$\frac{\partial F_t(1, x)}{\partial t} - \frac{\partial F_t(1, x)}{\partial x} = -\lambda F_t(1, x) - \frac{\partial F_k(1, 0)}{\partial x}+$$

$$+ \frac{\partial F_t(2, 0)}{\partial x} B(x) + \lambda F_t(0)B(x), \tag{2.8}$$

$$\frac{\partial F_t(i, x)}{\partial t} - \frac{\partial F_t(i, x)}{\partial x} = -\lambda F_t(i, x) - \frac{\partial F_t(i, 0)}{\partial x}+$$

$$+\frac{\partial F_t(i + 1, 0)}{\partial x} B(x) + \lambda F_t(i - 1, x), \quad i \ge 2.$$

Как правило, задача нахождения распределения значений процесса $N(t)$ в ТМО успешно решается только в случае рассмотрения стационарного, не зависящего от t, распределения вероятностей. Поэтому введем в рассмотрение такие стационарные распределения:

$$F(0) = \lim_{t\to\infty} F_t(0),$$

$$F(i, x) = \lim_{t\to\infty} F_t(i, x), \quad i \le 1. \tag{2.9}$$

Условием существования пределов (2.9) в системе типа $G|G|1$ является выполнение неравенства

$$\rho = \lambda \,/\, \mu = \lambda b_1 < 1. \tag{2.10}$$

Будем считать это условие выполненным. Переходя в (2.8) к пределу при $t \to \infty$, получаем следующую систему уравнений для функций $F(i, x)$:

$$\lambda F(0) = \frac{\partial F(1, 0)}{\partial x}, \tag{2.11}$$

$$\frac{\partial F(1, x)}{\partial x} - \frac{\partial F(1, 0)}{\partial x} - \lambda F(1, x) + \frac{\partial F(2, 0)}{\partial x} B(x)+$$

$$+\lambda F(0)B(x) = 0,$$

$$\frac{\partial F(i, x)}{\partial x} - \frac{\partial F(i, 0)}{\partial x} - \lambda F(i, x) + \frac{\partial F(i + 1, x)}{\partial x} B(x)+$$

$$+\lambda F(i-1,x)=0, \quad i\geq 2. \tag{2.12}$$

Удобным аппаратом для решения многих систем уравнений является аппарат производящих функций.

Производящей функцией последовательности величин $p_0, p_1, p_2\ldots$ назовем, вообще говоря, комплексную функцию $P(z)=\sum\limits_{k=0}^{\infty}p_k z^k,\ |z|<1.$

Если $p_0, p_1, p_2\ldots$ — есть распределение некоторой неотрицательной дискретной случайной величины $\xi\ :\ \ p_k=P\{x=k\}$, то $P(z)=Mz^{\xi}$. При этом, очевидно, $P(0)=p_0,\ P(1)=1$. Используя вид функции $P(z)$, в принципе можно получить (произвести) вид всех вероятностей $p_i:\ p_i=\frac{1}{i!}\frac{d^iP(z)}{dz^i}\Big|_{z=0}$.

Кроме того, используя производящую функцию, можно находить моменты распределения $p_k,\ k\geq 0$. Например:

$$M\xi=P'(1),$$
$$M\xi^2=P''(1)+P'(1),$$
$$D\xi=M(\xi-M\xi)^2=P''(1)+P'(1)-(P'(1))^2$$

и т. д.

Применение производящих функций позволяет свести задачу решения системы уравнений к задаче решения одного уравнения (правда, из класса более сложных уравнений).

Введем в рассмотрение производящую функцию

$$\Phi(z,x)=F(0)+\sum_{i=1}^{\infty}F(i,x)z^i, \quad |z|<1.$$

Умножая уравнения системы (2.12) на соответствующие степени z и суммируя, с учетом (2.11), получаем следующее уравнение:

$$\frac{\partial\Phi(z,x)}{\partial x}-\lambda(1-z)\Phi(z,x)-\frac{\partial\Phi(z,0)}{\partial x}\left[1-\frac{B(x)}{z}\right]+$$
$$+\lambda F(0)(1-z)(1-B(x))=0. \tag{2.13}$$

Уравнение (2.13) относится к классу дифференциально-функциональных уравнений. Методики решения таких уравнений нет. Для дальнейшего продвижения в решении задачи нахождения распределения процесса $N(t)$ от функций $\Phi(z,x)$ перейдем к их преобразованиям Лапласа. *Преобразованием Лапласа* функции $G(x)$ называется *комплексная функция*

$$g(s)=\int_0^{\infty}\mathrm{e}^{-sx}G(x)\,dx.$$

Функция $\gamma(s)=\int\limits_0^{\infty}\mathrm{e}^{-sx}\,dG(x)$ называется *преобразованием Лапласа–Стилтьеса функции* $G(x)$. Если обе функции $g(x)$ и $\gamma(s)$ существуют (соответствующий интеграл является сходящимся), то они связаны следующим образом: $\gamma(s)=\ sg(s)$. Несложно видеть, что если в качестве переменной s взять $-it$, а $G(x)$ есть вероятностное распределение случайной величины ξ, то $\gamma(s)=M\mathrm{e}^{it\xi}$ есть характеристическая функция случайной величины ξ. Характеристические функции подробно изучаются в университетском курсе "Теория вероятностей и математическая статистика", что является, в первую очередь, следствием применения их при доказательстве центральной предельной теоремы теории вероятностей. Поэтому подробно останавливаться на свойствах преобразования

Лапласа и Лапласа–Стилтьеса не будем. Отметим только важнейшие из них, которые будут использоваться нами ниже: 1) преобразование Лапласа–Стилтьеса производной функции $G(x)$ есть $s\gamma(s) - sG(+0)$; 2) преобразование Лапласа–Стилтьеса свертки распределений есть произведение преобразований Лапласа–Стилтьеса с этих распределений; 3) теорема о начальном значении:

$$\lim_{s\to\infty} sg(s) = \lim_{x\to 0} G(x);$$

4) теорема о конечном значении:

$$\lim_{s\to 0} sg(s) = \lim_{x\to\infty} G(x),$$

если функция $sg(s)$ аналитична для $Re\, s \geq 0$; 5) если $G(x)$ есть распределение случайной величины ξ, то начальные моменты g_i этого распределения определяются следующим образом:

$$g_i = \int_0^\infty x^i\, dG(x) = (-1)^i \frac{d^i\gamma(s)}{ds^i}\bigg|_{s=0}, \quad i \geq 1.$$

Итак, берем преобразование Лапласа от обоих частей уравнения (2.13), используя при этом свойство 1 и связь между преобразованиями Лапласа и Лапласа–Стилтьеса распределения $B(x)$. В результате для функции

$$\varphi(z,s) = \int_0^\infty e^{-sx}\Phi(z,x)\, dx$$

имеем уравнение

$$\varphi(z,s)(s - \lambda(1-z)) = \frac{1}{s}\Big[\frac{\partial\Phi(z,0)}{dx}\left(1 - \frac{\beta(s)}{z}\right) -$$

$$- \lambda F(0)(1-z)(1-\beta(s)) + sF(0)\Big], \tag{2.14}$$

где $\beta(s) = \int_0^\infty e^{-sx}\, dB(x)$ — преобразование Лапласа–Стилтьеса времени обслуживания требований. Несложно видеть, что функция $f(z,s)$ является аналитической в области $Re\, s > 0$. Из этого следует, что при $s = \lambda(1-z) > 0$ левая часть уравнения (2.14) обращается в нуль. Следовательно, при этом s равна нулю и правая часть (2.14). Исключая из этого условия неизвестную функцию $\dfrac{\partial\Phi(z,0)}{\partial x}$, имеем:

$$\frac{\partial\Phi(z,0)}{\partial x} = -\lambda z F(0)\frac{(1-z)\beta(\lambda(1-z))}{z - \beta(\lambda(1-z))}.$$

Подставим это выражение в (2.14):

$$\varphi(z,s) = \frac{F(0)}{s(s-\lambda(1-z))}\Big\{\frac{\lambda(1-z)\beta(\lambda(1-z))}{z-\beta(\lambda(1-z))}(z - \beta(s)) -$$

$$- \lambda(1-z)(1-\beta(s)) + s\Big\}. \tag{2.15}$$

Теперь вспомним, что процесс $\eta(t)$ был введен нами в рассмотрение с целью получения марковского процесса и, собственно, нас интересует распределение вероятностей процесса $N(t)$ — числа требований в системе в момент t.

Обозначим

$$p_i = \lim_{t\to\infty}\mathbf{P}\{N(t) = i\}.$$

Очевидно, $p_i = \lim_{x\to\infty}\Phi(i,x), \quad i \geq 0$.

Введя в рассмотрение производящую функцию $R(z) = \sum_{i=1}^\infty p_i z^i,\ |z| < 1$, устремляя в (2.15) величину s к нулю и используя свойство 4 преобразования Лапласа, получаем:

$$R(z) = \frac{F(0)(1-z)\beta(\lambda(1-z))}{\beta(\lambda(1-z)) - z}. \tag{2.16}$$

Неизвестную пока константу $F(0) = p_0$ легко найти из условия нормировки: $R(1) = 1$ с использованием очевидного разложения

$$\beta(\lambda(1-z)) - z =$$

$$= (1-z)\left(1 - \lambda b_1 + \frac{\lambda^2 b_2}{2}(1-z) + \frac{\lambda^3 b_3}{6}(1-z)^2 + o(1-z)^2\right). \tag{2.17}$$

В результате имеем: $F(0) = 1 - \lambda b_1 = 1 - \rho$ и, окончательно,

$$R(z) = \frac{(1-\rho)(1-z)\beta(\lambda(1-z))}{\beta(\lambda(1-z)) - z}. \tag{2.18}$$

Формула (2.18) называется *формулой Поллячека–Хинчина* для производящей функции распределения числа требований в системе $M|G|1$.

Лабораторная работа: "Применение метода введения дополнительной переменной для нахождения распределения числа требований в системе M|G|1“

Цель работы. Овладение методом введения дополнительной переменной и использование формулы Поллячека–Хинчина для нахождения характеристик системы $M|G|1$.

Порядок выполнения работы. Задания выполняются на основе использования формулы (2.18).

Задание 1. Найти вид производящей функции $R(z)$ в случае, когда время обслуживания требований имеет экспоненциальное распределение, явный вид вероятностей $p_i,\ i \geq 1$.

Задание 2. Проделать это же для детерминированного времени обслуживания требований.

Задание 3. Используя (2.18) и разложение (2.17), получить выражения для среднего числа $\overline{N}$ требований в системе и дисперсии $\overline{D}$ распределения числа требований в системе.

Задание 4. Используя полученные выражения для величин $\overline{N}$ и $\overline{D}$, найти среднее число требований в системе и дисперсию для систем, у которых время обслуживания имеет следующие распределения:

— детерминированное $\left(B(t) = \begin{cases} 0, & t < b_1 \\ 1, & t \geq b_1 \end{cases}\right)$;

— экспоненциальное ($B(t) = 1 - e^{-\mu t}$);

— эрланговское порядка k

$$\left(B(t) = \int\limits_0^t \frac{\mu(\mu\tau)^{k-1}}{(k-1)!} e^{-\mu\tau}\, d\tau\right);$$

— гиперэкспоненциальное порядка k

$$\left(B(t) = \sum_{r=1}^{k} q_r(1 - e^{-\mu_r t}),\ q_r \geq 0,\ \sum_{r=1}^{k} q_r = 1\right);$$

— гиперэрланговское

$$\left(B(t) = \sum_{r=1}^{k} q_r \int_0^t \frac{\mu_r(\mu_r\tau)^{l_r-1}}{(l_r-1)!} e^{-\mu_r\tau}\, d\tau\right);$$

— равномерное в отрезке $[t_1, t_2]$

$$\left(\frac{dB(t)}{dt} = \begin{cases} \frac{1}{t_2 - t_1}, & t \in [t_1, t_2], \\ 0, & t \notin [t_1, t_2] \end{cases} \right).$$

Задание 5. Выяснить скорость изменения величины $\overline{N}$ при изменении величины загрузки ρ от 0 до 1 для этих распределений.

2.3 Метод вложенных цепей Маркова. Система $M|G|1$

Как отмечалось в п. 2.1, другим методом марковизации процессов, описывающих поведение СМО, является метод внутренней марковизации или метод вложенных цепей Маркова. Идея метода там же вкратце изложена. В данном параграфе проиллюстрируем его применение для нахождения распределения числа требований в системе $M|G|1$.

Итак, процесс $N(t)$ не является марковским. Причиной его немарковости является необходимость учета времени, прошедшего к моменту t с момента начала обслуживания обслуживаемого требования, или времени с момента t до момента окончания обслуживания (времени дообслуживания), для предсказания поведения процесса $N(t)$ после момента t. Из сказанного очевидно, что, если мы будем рассматривать поведение процесса $N(t)$ только для значений t, равных $\tau_1, \tau_2, \ldots, \tau_n$, где τ_i — момент окончания обслуживания i—го требования, то проблема необходимости учета времени дообслуживания снимается и получающийся процесс — цепь Маркова $N(\tau_n),\ n \geq 1$, с дискретным временем. Представление о цепях Маркова можно составить из 1.1. В данном случае пространством состояний цепи Маркова $N(\tau_n),\ n \geq 1$, (под значением $N(\tau_n)$ будем понимать величину $N(\tau_n + 0)$) является множество неотрицательных целых чисел. Начальное распределение этой цепи значения не имеет, поскольку нас будут интересовать только стационарные распределения эргодических цепей. Таким образом, для полного задания цепи Маркова нам осталось найти матрицу одношаговых переходных вероятностей:

$$p_{i,j} = \mathbf{P}\{N(\tau_{n+1}) = j \,|N(\tau_n) = i\}.$$

Пусть $i > 0$. Тогда после момента τ_n происходит следующее. В течение времени до момента τ_{n+1} обслуживается требование, время обслуживания его имеет функцию распределения $B(x)$. За это время в систему может прийти η требований, которые встанут в очередь на обслуживание. В момент τ_{n+1} требование находящееся на обслуживании, уходит из системы. Таким образом $j = i + \eta - 1$. Найдем распределение случайной величины η. Очевидно, если время обслуживания примет значение t, то вероятность того, что η примет значение l есть вероятность прихода за время t в систему l требований стационарного пуассоновского потока с параметром λ, т. е. $P\{\eta = l\,|$ время обслуживания равно $t\} = \dfrac{(\lambda t)^l}{l!}\mathrm{e}^{-\lambda t}, \quad l \geq 0$. Усредняя теперь по распределению времени обслуживания, получаем:

$$P\{\eta = l\} = f_l = \int_0^\infty \frac{(\lambda t)^l}{l!}\mathrm{e}^{-\lambda t}\, dB(t), \quad l \geq 0.$$

Обратим внимание читателя на то, что в курсе теории вероятностей обычно рассматривают формулу полной вероятности при наличии конечного или счетного числа альтернатив-событий из полной группы. Здесь и далее мы будем иметь дело с формулой полной вероятности, но уже с непрерывным множеством альтернатив. Соответственно операция суммирования, имеющая место в классической формуле полной вероятности, заменяется интегрированием по распределению вероятностей событий—альтернатив.

Если же $i = 0$, то после момента τ_n система ждет поступления требования, а с момента его поступления она ведет себя также, как в случае, когда $i = 1$.

Поэтому

$$p_{i,j} = f_{j-i+1} = \int\limits_0^\infty \frac{(\lambda t)^{j-i+1}}{(j-i+1)!} \mathrm{e}^{-\lambda t}\, dB(t), \quad j \geq i-1,\ i \geq 1,$$

$$p_{0,j} = f_j. \tag{2.19}$$

Итак, цепь Маркова $N(\tau_n),\ n \geq 1$, нами полностью определена. Найдем ее стационарное распределение вероятностей состояний

$$\pi_k = \lim_{n\to\infty} P\{N(\tau_n) = k\}, \quad k \geq 0. \tag{2.20}$$

С использованием известных критериев существования стационарного распределения цепей Маркова, можно убедиться, что пределы (2.20) существуют при выполнении условия (2.10). Как следует из п.1.1. и (2.19), вероятности $\pi_k,\ k \geq 1$, удовлетворяют системе линейных алгебраических уравнений вида

$$\pi_k = \sum_{l=1}^{k+1} \pi_l f_{k-l+1} + \pi_0 f_k. \tag{2.21}$$

Для решения системы (2.21) используем аппарат производящих функций. Вводим в рассмотрение производящие функции

$$\prod(z) = \sum_{k=0}^\infty \pi_k z^k \text{и } F(z) = \sum_{k=0}^\infty f_k z^k, \quad |z| < 1.$$

Умножая уравнения системы (2.21) на соответствующие степени z и суммируя, получаем

$$\prod(z) = \sum_{k=0}^\infty z^k \sum_{l=1}^{k+1} \pi_l f_{k-l+1} + \pi_0 F(z) =$$

$$= \sum_{l=1}^\infty z^{l-1}\pi_l \sum_{k=l-1}^\infty z^{k-l+1} f_{k-l+1} + \pi_0 F(z) =$$

$$= \frac{\prod(z) - \pi_0}{z} F(z) + \pi_0 F(z).$$

Отсюда имеем

$$\prod(z) = \pi_0 \frac{(1-z)F(z)}{F(z) - z}.$$

Подсчитывая значения производящей функции $F(z)$, с учетом явного вида вероятностей f_l, получаем $F(z) = \beta(\lambda(1-z))$. Таким образом, окончательно имеем

$$\prod(z) = \pi_0 \frac{(1-z)\beta(\lambda(1-z))}{\beta(\lambda(1-z)) - z}, \tag{2.22}$$

где вероятность p_0 находится так же, как и в предыдущем параграфе и также равна $1-\rho$. Из (2.18) и (2.22) делаем вывод, что у системы $M|G|1$ стационарное распределение вероятностей состояний системы в произвольный момент времени совпадает с соответствующим распределением в моменты окончания обслуживания требований. Поэтому обычно присутствующий при применении метода вложенных цепей Маркова этап нахождения распределения вероятностей состояний системы в произвольный момент времени по распределению вероятностей состояний системы во вложенные моменты времени здесь можем опустить. Такой этап будет присутствовать в следующем параграфе при рассмотрении системы $GI|M|1$.

Таким образом, использование производящей функции позволило нам быстро и эффективно решить систему линейных алгебраических уравнений (2.21).

Отметим, что используя формулу Поллячека–Хинчина (2.22) для производящей функции числа требований в системе, легко получить и формулу Поллячека–Хинчина для преобразования Лапласа-Стилтьеса распределения времени ожидания требования в системе.

Пусть $W(t)$ — функция распределения времени ожидания требования в системе; $V(t)$ — функция распределения времени пребывания требования в системе,

$$w(s) = \int\limits_0^\infty \mathrm{e}^{-st}\, dW(t), \quad v(s) = \int\limits_0^\infty \mathrm{e}^{-st}\, dV(t)$$

— их преобразования Лапласа–Стилтьеса. Очевидно, функция распределения $V(t)$ времени пребывания требования в системе является сверткой распределений $W(t)$ времени ожидания и $B(t)$ времени обслуживания требований. Используя свойство 2 преобразований Лапласа–Стилтьеса из п. 2.2, имеем

$$v(s) = w(s)\beta(s). \tag{2.23}$$

Далее, при обслуживании требований в порядке поступления число требований, остающихся в системе в момент окончания обслуживания некоторого требования, равно числу требований, поступивших в систему за время пребывания в ней этого уходящего требования. Используя формулу полной вероятности, имеем

$$\pi_j = \int\limits_0^\infty \frac{(\lambda t)^j}{j!}\mathrm{e}^{-\lambda t}\, dV(t),\ j \geq 0. \tag{2.24}$$

Умножая соотношения (2.24) на соответствующие степени z и суммируя, получаем соотношение:

$$\prod(z) = \sum_{j=0}^{\infty} \pi_j z^j = \sum_{j=0}^{\infty} \int\limits_0^\infty \frac{(\lambda t)^j}{j!} z^j \mathrm{e}^{-\lambda t}\, dV(t) = v(\lambda(1-z)).$$

Отсюда с использованием (2.22) и (2.23) имеем:

$$w(\lambda(1-z)) = \frac{(1-\rho)(1-z)}{\beta(\lambda(1-z)) - z}. \tag{2.25}$$

Вводя обозначение $s = \lambda(1-z)$, из (2.25) окончательно имеем:

$$w(s) = \frac{1-\rho}{1 - \lambda \dfrac{1-\beta(s)}{s}}. \tag{2.26}$$

Формула (2.26) известна в литературе как формула Поллячека–Хинчина для преобразования Лапласа–Стилтьеса распределения времени ожидания требования в системе $M|G|1$, чаще ее получают на основе уравнения Такача.

Как видно, метод вложенных цепей Маркова является очень эффективным при исследовании систем типа $M|G|1$. Интересным вариантом систем $M|G|1$ являются системы с удаляющимся на отдых прибором, адекватно моделирующие поведение многих реальных объектов и процессов. Одна из возможных моделей системы следующая.

Имеется система, отличающаяся от классической СМО $M|G|1$ поведением при попадании системы в свободное состояние. Если в некоторый момент окончания обслуживания требования система оказалась свободной от требований, с вероятностью $1-p$ система будет вести себя как обычная СМО типа $M|G|1$ — ожидать поступления требования, после чего начнет его обслуживание. А с вероятностью p происходит следующее. Начинается период отдыха системы. Период отдыха состоит из медленного цикла отдыха, длительность которого имеет функцию распределения $H(t)$ с первым моментом $\tau_1 = \int_0^\infty (1 - H(t))\,dt$, и, возможно, некоторого числа быстрых циклов отдыха, длительность каждого из которых имеет экспоненциальное распределение с первым моментом τ_2. Период отдыха заканчивается (и начинается обслуживание требований), если за время медленного цикла в систему поступило не менее $n,\ n \geq 1$, требований. Если же число требований оказалось меньше, чем n, то период отдыха закончится с окончанием быстрого цикла, во время которого число требований в системе достигло величины $k,\ k \geq n$. Необходимо при заданных параметрах системы (p, n, k) найти среднее время пребывания требования в системе. Очевидно, в качестве вложенных моментов времени здесь также берем моменты окончания обслуживания требований. Соответствующая цепь Маркова во всех состояниях $i,\ i > 0$, ведет себя так же, как и цепь в рассмотренном выше классическом случае.

Рассмотрим поведение системы после попадания ее в состояние 0. С вероятностью $1-p$ отдых в системе не начинается, и она ведет себя как классическая СМО $M|G|1$. Очевидно, в состояния с номерами, меньшими $n-1$, мы можем попасть только в этом варианте, поэтому

$$p_{0,l} = (1-p)f_l, \quad l = \overline{0, n-2}. \tag{2.27}$$

С вероятностью p начинается отдых. При этом в состояния с номерами $\overline{n-1, k-2}$ мы можем попасть только в случае, когда отдых закончится медленным циклом. Таким образом

$$p_{0,n+l} = (1-p)f_{n+l} + p\sum_{m=0}^{l+1} g_{n+m} f_{l-m+1}, \quad l = \overline{-1, k-n-2}, \tag{2.28}$$

где

$$g_i = \int_0^\infty \frac{(t\lambda)^i}{i!} \mathrm{e}^{-\lambda t}\, dH(t),\ i \geq 0,$$

есть вероятность поступления i требований за время медленного цикла.

И, наконец, в состояния с номерами от $k-1$ и выше мы можем попасть как в случае, когда отдых не начался, так и в случаях, когда он начался и кончился медленным циклом либо быстрым циклом. Учитывая, что вероятность того, что в период отдыха будут быстрые циклы, равна $G = \sum_{m=0}^{n-1} g_m$, а также тот факт, что длительность интервала от момента достижения в каком–либо быстром цикле числом требований уровня k до момента окончания этого цикла имеет показательное распределение с первым моментом τ_2, получаем выражение

$$p_{0,k+l} = (1-p)f_{k+l} + p\sum_{m=0}^{k-n+l+1} g_{n+m} f_{k+l-m-n+1} +$$
$$+ pG\sum_{m=0}^{l+1} \frac{\rho_2^m}{(1+\rho_2)^{m+1}} f_{l-m+1}, \quad l \geq -1, \tag{2.29}$$

где $\rho_2 = \lambda\tau_2$.

При этом отметим, что если для медленного цикла его длительность может иметь произвольное распределение, то для длительности быстрого цикла мы предполагаем показательное распределение, в противном случае третье слагаемое в (2.29) имеет чрезвычайно сложный вид. Кроме того, поясним, что величина

$$\frac{\rho_2^m}{(1+\rho_2)^{m+1}} = \int_0^\infty \frac{(\lambda t)^m}{m!} \mathrm{e}^{-\lambda t}\, d\left(1 - \mathrm{e}^{-\tau_2^{-1} t}\right)$$

— вероятность поступления в систему m требований за остав- шуюся часть быстрого цикла после момента достижения уровня k.

Таким образом, переходные вероятности $p_{i,j}$ вложенной цепи Маркова для рассматриваемой системы имеют вид (2.19) для $i > 0$ и вид (2.27)–(2.29) для $i > 0$. Подставляя явный вид этих вероятностей в систему линейных алгебраических уравнений

$$\pi_i = \sum_{l=0}^{i+1} \pi_l p_{l,i}$$

для стационарных вероятностей цепи Маркова, умножая уравнения этой системы на соответствующие степени z и суммируя их, получаем следующее выражение для производящей функции $\prod(z)$ этих вероятностей:

$$\prod(z) = \pi_0 \frac{(1-z)\beta(\lambda(1-z))}{\beta(\lambda(1-z))-z} + p\pi_0 \frac{\beta(\lambda(1-z))}{\beta(\lambda(1-z))-z} \times$$
$$\times \left[z - \sum_{n=m}^{\infty} g_m z^m - \frac{Gz^k}{1+\rho_2(1-z)} \right], \tag{2.30}$$

где вероятность π_0 находится обычным образом из условия нормировки $\prod(1) = 1$ и имеет вид

$$\pi_0 = (1-\rho)\left[1 + p\left(G(\rho_2 + k) - 1 + \sum_{n=m}^{\infty} m g_m\right)\right]^{-1}. \tag{2.31}$$

Отметим, что для данной системы остается верным соотношение $\prod(z) = v(\lambda(1-z))$, связывающее производящую функцию $\prod(z)$ числа требований в системе в моменты окончания обслуживания требований и преобразование $v(s)$ Лапласа–Стилтьеса распределения времени пребывания требования в системе. Дифференцируя это соотношение по z, устремляя z к единице и учитывая свойства производящей функции и преобразования Лапласа–Стилтьеса, получаем следующую формулу

$$\lambda \overline{V} = \overline{N}, \tag{2.32}$$

связывающую среднее время $\overline{V}$ пребывания требования в системе и среднее число требований в системе. Отметим, что формула (2.32) справедлива и в более общем классе СМО, а именно для СМО $G|G|1$, и известна в литературе как формула Литтла. Используя формулу (2.32), дифференцируем (2.30) по z, устремляем z к единице и получаем следующее выражение для величины $\overline{V}$ среднего времени пребывания требования в системе:

$$\overline{V} = b_1 + \frac{\lambda b_2}{2(1-\rho)} +$$
$$+ p\frac{\pi_0}{1-\rho}\lambda^{-1}\left[\left(\rho_2^2 + k\rho_2 + \frac{k(k-1)}{2}\right)G - \frac{1}{2}\sum_{m=n}^{\infty} m(1-m)g_m\right]. \tag{2.33}$$

Несложно видеть, что при $p = 0$ (отдых никогда не начинается) выражения (2.30), (2.31), (2.33) переходят в известные выражения для классической СМО $M|G|1$.

Лабораторная работа: "Применение метода вложенных цепей Маркова для нахождения распределения числа требований в системе M|G|1 и среднего времени пребывания требования в системе M|G|1 с прибором, уходящим на отдых“

Цель работы. Овладение методом вложенных цепей Маркова и использование его для расчета характеристик системы $M|G|1$.

Порядок выполнения работы. Задания выполняются с использованием дополнительных указаний.

Задание 1. Найти распределение требований в моменты окончания обслуживания требований в системе $M_x|G|1$, отличающейся от классической системы $M|G|1$ тем, что требования приходят группами, поток групп является простейшим с параметром λ, число требований в группе имеет распределение $\theta_k,\ k \geq 1$, с производящей функцией $\theta(z) = \sum_{k=1}^{\infty} \theta_k z^k$.

Указание. В классической системе производящая функция вероятностей числа запросов, пришедших в систему за время t есть $e^{-\lambda(1-z)t}$. В системе с вышеописанным потоком такая производящая функция имеет вид: $e^{-\lambda(1-\theta(z))t}$.

Задание 2. Для модели, описанной в задании 1, найти выражение для среднего числа требований в системе в моменты окончания обслуживания требований. Отдельно рассмотреть случай, когда распределение θ_k является геометрическим, т. е. $\theta_k = \theta^{k-1}(1-\theta),\ k \geq 1$.

Задание 3. Для модели СМО $M|G|1$ с удаляющимся на отдых прибором найти вид производящей функции $\prod(z)$ числа требований, вероятности π_0 и среднего времени пребывания $\overline{V}$ требования в системе в случае, когда распределение длительности медленного цикла является показательным.

Указание. При этом $g_i = \dfrac{\rho_1^i}{(1+\rho_1)^{i+1}},\ i \geq 0$, где $\rho_1 = \lambda\tau_1$.

Задание 4. Для модели СМО с удаляющимся на отдых прибором найти вид производящей функции $\prod(z)$ в случае, когда входной поток идентичен описанному в задании 1.

Задание 5. Для модели СМО с удаляющимся на отдых прибором найти выражение для среднего числа периодов отдыха в единицу времени.

Указание. Попутно показать, что среднее число быстрых циклов в одном периоде отдыха равно

$$G + \sum_{m=0}^{n-1} g_m \frac{k-m}{\rho_2}.$$

Задание 6. Для различных видов распределения времени обслуживания требований найти вид вероятностей $f_l,\ l \geq 1$. С помощью полученных формул и рекуррентной процедуры

$$\pi_{l+1} = \frac{1}{f_0}\left(\pi_l - \pi_0 f_l - \sum_{m=1}^{l} \pi_m f_{l-m+1}\right), \quad l \geq 0,$$

реализовать на ЭВМ расчет коэффициентов α_l, определяемых соотношением $\pi_l = \alpha_l \pi_0$, для системы $M|G|1$.

Задание 7. Для системы $M|E_l|1$ найти явный вид вероятностей $p_k,\ k \geq 1$.

Указание. Функция в правой части формулы Поллячека—Хинчина для данной системы является отношением двух полиномов. Разложить ее на простые дроби и воспользоваться известными формулами для их обращения.

2.4 Метод вложенных цепей Маркова. Система $GI|M|m$.

Несложно видеть, что у данной СМО процесс $N(t)$ не является марковским, поскольку для предсказания поведения этого процесса после момента t необходима дополнительная информация о времени, прошедшем к моменту t с момента поступления в систему последнего требования. Из сказанного ясно, что в качестве моментов $\tau_n,\ n \geq 1$, в которые процесс $N(\tau_n),\ n \geq 1$, является марковским, логично выбрать моменты поступления требований в систему. Под значением процесса $N(\tau_n)$ в момент τ_n будем понимать величину $N(\tau_n - 0)$. Пространством состояний этой цепи Маркова является множество всех неотрицательных целых чисел. Найдем переходные вероятности $p_{i,j}$ этой цепи Маркова. При этом необходимо различать следующие четыре случая.

1. $i + 1 < j$. В этом случае, очевидно, $p_{i,j} = 0$.

2. $j \le i+1 \le m$. В этом случае, очевидно, между двумя моментами поступления требований все требования получали обслуживание, причем какие-то j требований обслуживаться не закончили, а $i+1-j$ — закончили. Применяя формулу полной вероятности, получаем:

$$\pi_{i,j} = \int\limits_0^\infty C_{i+1}^j \mathrm{e}^{-\mu t j}\left(1-\mathrm{e}^{-\mu t}\right)^{i+1-j} dA(t), \quad j \le i+1 \le m. \tag{2.34}$$

3. $m \le j \le i+1$. В этом случае между моментами поступления требований все приборы непрерывно заняты обслуживанием требований. Поэтому поток уходящих из системы требований является стационарным пуассоновским с параметром $m\mu$. Вероятность того, что за время t будем обслужено $i+1-j$ требований, при этом есть

$$\frac{(m\mu t)^{i+1-j}}{(i+1-j)!}\mathrm{e}^{-m\mu t}.$$

Усредняя по распределению длины интервала между моментами поступления требований, получаем:

$$\pi_{i,j} = \int\limits_0^\infty \frac{(m\mu t)^{i+1-j}}{(i+1-j)!}\mathrm{e}^{-m\mu t}\, dA(t), \quad m \le j \le i+1. \tag{2.35}$$

4. $j < m < i+1$. В этом случае между моментами поступления требований происходит следующее. Через время y, в течение которого непрерывно шло обслуживание требований, обслужилось $i-m+1$ требований, за оставшееся время $t-y$ из m присутствующих в системе в момент y требований какие—то j требований не окончили обслуживания, а $m-j$ — окончили. Учитывая, что сумма времен обслуживания $i-m+1$ требований имеет эрланговское распределение и усредняя по распределению длины интервала между моментами поступления, получаем:

$$p_{i,j} = \int\limits_0^\infty \int\limits_0^t m\mu \frac{(m\mu y)^{i-m}}{(i-m)!}\mathrm{e}^{-m\mu y} C_m^j \mathrm{e}^{-j\mu(t-y)} \times$$

$$\times \left(1-\mathrm{e}^{-\mu(t-y)}\right)^{m-j} dy\, dA(t) = \int\limits_0^\infty C_m^j \mathrm{e}^{-j\mu t} \int\limits_0^t \frac{(m\mu y)^{i-m}}{(i-m)!} \times$$

$$\times \left(\mathrm{e}^{-\mu y} - \mathrm{e}^{-\mu t}\right)^{m-j} m\mu\, dy\, dA(t), \quad j < m < i+1. \tag{2.36}$$

Таким образом, переходные вероятности $p_{i,j}$ нами полностью определены.

Используя результаты теории цепей Маркова, можно убедиться, что при выполнении условия $\lambda < m\mu$, существует стационарное распределение $r_k,\ k \ge 0$, вероятностей вложенной цепи Маркова:

$$r_k = \lim_{n\to\infty} \mathbf{P}\{N(\tau_n) = k\}, \quad k \ge 0.$$

Опишем процедуру нахождения вероятностей $r_k,\ k \ge 1$. Система линейных алгебраических уравнений для вероятностей $r_k,\ k \ge 0$, имеет вид

$$r_k = \sum_{i=k-1}^\infty r_l p_{l,k}, \quad k \ge 1. \tag{2.37}$$

В данном случае аппарат производящих функций не является достаточно эффективным средством решения системы. Обратим внимание на то, что при $k \ge m$ все переходные вероятности имеют вид (2.35) и из (2.37) имеем:

$$r_k = \sum_{i=k-1}^{\infty} r_l \int_0^{\infty} \frac{(m\mu t)^{i+1-k}}{(i+1-k)!} \mathrm{e}^{-m\mu t}\, dA(t), \quad k \geq m. \tag{2.38}$$

Вероятности r_k будем искать в виде:

$$r_k = C\delta^k, \quad k \geq m-1. \tag{2.39}$$

Подставляя (2.39) в (2.38) и сокращая обе части (2.38) на $C\delta^{k-1}$, получаем соотношение

$$\delta = \sum_{l=k-1}^{\infty} \delta^{l-k+1} \int_0^{\infty} \frac{(m\mu t)^{l+1-k}}{(l+1-k)!} \mathrm{e}^{-m\mu t}\, dA(t) =$$

$$= \sum_{n=0}^{\infty} \delta^n \int_0^{\infty} \frac{(m\mu t)^n}{n!} \mathrm{e}^{-m\mu t}\, dA(t) =$$

$$= \int_0^{\infty} \mathrm{e}^{-(m\mu - m\mu\delta)t}\, dA(t) = A^*(m\mu(1-\delta)).$$

где $A^*(\,)$ есть преобразование Лапласа—Стилтьеса распределения $A(t)$. Таким образом, константа δ является решением функционального уравнения

$$\delta = A^*(m\mu(1-\delta)). \tag{2.40}$$

Несложно показать, что если $\rho < 1$, то уравнение (2.40) имеет единственное действительное решение в области $0 < d < 1$. Таким образом, неизвестными пока остались константа C и вероятности $r_i,\ i = \overline{0, m-2}$. С учетом (2.39) система (2.37) переписывается в виде

$$r_k = \sum_{i=k-1}^{m-2} r_i p_{i,k} + \sum_{i=m-1}^{\infty} C\delta^i p_{i,k}, \quad k \geq 1. \tag{2.41}$$

Введя обозначение $R_k = r_k \,/\, G$, где $G = C\delta^{m-1}$, систему (2.41) перепишем в виде

$$R_k = \sum_{i=k-1}^{m-2} R_i p_{i,k} + \sum_{i=m-1}^{\infty} \delta^{i-m+1} p_{i,k},$$

откуда

$$R_{k-1} = \frac{R_k - \sum_{i=k}^{m-2} R_i p_{i,k} - \sum_{i=m-1}^{\infty} \delta^{i+1-m} p_{i,k}}{p_{k-1,k}}, \quad k = \overline{1, m-1}. \tag{2.42}$$

Можно убедиться, что бесконечная сумма в (2.42) равна

$$C_m^k \sum_{l=0}^{m-k} C_{m-k}^l (-1)^l \Lambda^*(\mu(k+l)) \frac{m-k-l}{m-k-l-\delta m} + \delta^2 m \frac{m!}{k!} \prod_{i=0}^{m-k} (i+k+m(\delta-1))^{-1}.$$

Формула (2.42) в сочетании с начальным условием

$$R_{m-1} = 1 \tag{2.43}$$

задает рекуррентную процедуру нахождения вида величин $R_k,\ k = \overline{0, m-2}$. Неизвестная пока константа C находится из условия нормировки $\sum\limits_{k=0}^{\infty} r_k = 1$ и имеет вид

$$C = \frac{\delta^{1-m}}{\dfrac{1}{1-\delta} + \sum\limits_{k=0}^{m-2} R_k}. \tag{2.44}$$

Формулы (2.39), (2.40), (2.42)–(2.44) полностью задают процедуру нахождения стационарного распределения вероятностей $r_k,\ k \geq 0$, вложенной по моментам поступления требований цепи Маркова для системы $GI|M|m$.

Очевидно, вероятность того, что поступающее требование не будет ждать обслуживания, определяется следующим образом:

$$W(0) = \sum_{k=0}^{m-1} r_k. \tag{2.45}$$

Требование, пришедшее в систему, когда в ней находится $k,\ k \geq m$, требований, будет ждать, пока $k - m + 1$ требований, пришедших в систему раньше его, окончат обслуживаться. Учитывая, что время их обслуживания имеет распределение Эрланга с параметрами $(m\mu, k - m + 1)$, и применяя формулу полной вероятности, получаем следующее выражение для функции распределения $W(y)$ времени ожидания требования в системе:

$$W(y) = W(0) + C \sum_{k=m}^{\infty} \delta^k \int\limits_0^y \frac{(m\mu)(m\mu x)^{k-m}}{(k-m)!} e^{-m\mu x}\, dx =$$

$$= W(0) + C\delta^m \int\limits_0^y m\mu e^{-m\mu x(1-\delta)}\, dx. \tag{2.46}$$

Учитывая (2.44) и (2.45), из (2.46) окончательно имеем

$$W(y) = 1 - \frac{\delta}{1 + (1-\delta)\sum\limits_{k=0}^{m-2} R_k} e^{-m\mu(1-\delta)y}, \quad y \geq 0. \tag{2.47}$$

Среднее время ожидания

$$\overline{W} = \frac{C\delta^m}{m\mu(1-\delta)^2}. \tag{2.48}$$

В случае, когда рассматриваемая система является однолинейной ($m = 1$), все вероятности состояний цепи Маркова определяются формулой (2.39), где константа $C = 1 - \delta$. Напомним, что $r_k,\ k \geq 1$, есть стационарное распределение вероятностей вложенной цепи Маркова $N(\tau_n)$. Покажем теперь, как найти вероятности $p_l,\ l \geq 1$, наличия l требований в системе в произвольный момент времени для однолинейной СМО.

Сначала пусть $l > 0$. Известно, что время с момента поступления последнего требования до произвольного момента времени имеет функцию распределения $G(t) = \lambda \int_0^t (1 - A(\tau))\, dt$. В этот момент поступления в системе могло быть k, k не менее $l-1$, требований. За время до данного произвольного момента должно обслужиться $l-k+1$ требований, причем поток обслуженных требований является простейшим с параметром μ. Применяя формулу полной вероятности (по распределению числа требований в системе в момент поступления требования и по распределению $G(t)$), получаем:

$$p_l = \sum_{k=l-1}^{\infty} r_k \lambda \int\limits_0^{\infty} \frac{(\mu t)^{l-k+1}}{(l-k+1)!} e^{-\mu t} (1 - A(t))\, dt =$$

$$= \lambda(1-\delta)\delta^{l-1} \int\limits_0^{\infty} e^{-\mu t(1-\delta)} (1 - A(t))\, dt = (1-\delta)\delta^{l-1}\rho, \quad l \geq 1. \tag{2.49}$$

Пусть теперь $l = 0$. Аналогично, время с момента поступления последнего требования до данного произвольного момента времени имеет распределение $G(t)$. В последний момент поступления требования в системе было k, $k \geq 0$, требований, но до данного момента все они окончили обслуживаться. Учитывая, что время обслуживания $k+1$ требований имеет распределение Эрланга соответствующего порядка, и применяя формулу полной вероятности, получаем:

$$p_0 = \sum_{k=0}^{\infty} r_k \int\limits_0^{\infty} \lambda(1 - A(t)) \int\limits_0^t \frac{\mu(\mu y)^k}{m!} \mathrm{e}^{-\mu y}\, dy\, dt =$$

$$= (1-\delta)\lambda\mu \int_0^{\infty} (1 - A(t)) \int\limits_0^t \mathrm{e}^{-\mu(1-\delta)y}\, dy\, dt = 1 - \rho.$$

Таким образом, стационарное распределение числа требований в системе в произвольный момент времени задается следующим образом:

$$p_0 = 1 - \rho,$$

$$p_l = \rho(1-\delta)\delta^{l-1}, \quad l \geq 1.$$

Лабораторная работа: "Применение метода вложенных цепей Маркова для нахождения распределения числа требований и времени ожидания в системе
$GI|M|m$"

Цель работы. Овладение методом вложенных цепей Маркова и использование его для расчета характеристик системы $GI|M|m$.

Задание 1. Реализовать на ЭВМ подсчет переходных вероятностей $p_{i,j}$ вложенной цепи Маркова для различных видов распределения $A(t)$ интервалов между моментами поступления требований (см. задание 4 в п. 2.2).

Задание 2. Для различных видов распределения $A(t)$ реализовать на ЭВМ подсчет величин R_k, задаваемых рекуррентными формулами (2.42) с начальным условием (2.43).

Задание 3. Доказать существование единственного действительного корня уравнения (2.40) и реализовать на ЭВМ процедуру его нахождения для различных видов распределения $A(t)$.

Задание 4. Найти явный вид вероятностей r_k, $k \geq 0$, и p_l, $l \geq 0$, для системы вида $E_2|M|1$, а также распределение времени ожидания требования в этой системе.

Задание 5. Найти явный вид вероятностей r_k, $k \geq 0$, p_l, $l \geq 0$, и распределения времени ожидания требований в системе $M|M|2$.

2.5 Метод введения дополнительного события

Этот метод часто дает простое и изящное решение задач нахождения характеристик СМО. Он разработан Данцигом, Ранненбургом, а также Климовым. Метод основывается на вероятностной интерпретации производящих функций и преобразований Лапласа—Стилтьеса. Производящей функцией $P(z)$ распределения случайной величины ξ можно придать вероятностный смысл следующим образом. Будем интерпретировать ξ как число каких—то требований, поступивших на некотором интервале времени. Каждое из этих требований независимо от других с вероятностью z окрасим в красный цвет, а с дополнительной вероятностью — в синий. Тогда из формулы полной вероятности следует, что $P(z)$ является вероятностью того, что на данном интервале времени поступали только красные требования.

Для преобразования Лапласа—Стилтьеса распределения $F(x)$ вероятностная интерпретация дается следующим образом. Случайную величину η с функцией распределения $F(x)$ будем трактовать как время жизни некоторого элемента. Будем считать, что во время жизни элемента поступает стационарный пуассоновский поток катастроф с интенсивностью s. Тогда, как известно, e^{-st} есть вероятность того, что за время t не поступит ни одной катастрофы. Применяя формулу полной вероятности, заключаем, что преобразование Лапласа-Стилтьеса $F^*(s) = \int_0^\infty e^{-st}\,dF(t)$ есть вероятность того, что за время жизни элемента не произошло катастрофы.

Сущность метода введения дополнительного события заключается в следующем. Используя вышеприведенную вероятностную интерпретацию преобразований Лапласа—Стилтьеса и производящих функций, вероятность интересующего нас события подсчитывается двумя различными способами. Полученные выражения приравниваются, и получается уравнение для интересующей нас характеристики системы.

Проиллюстрируем применение этого метода на примерах. Одной из важных характеристик СМО является распределение величины периода занятости системы — интервала времени с момента поступления требования в пустую систему до первого после этого момента, когда система снова станет пустой. Найдем распределение длины периода занятости системы $M|G|1$. Обозначим через $\prod(t)$ функцию распределения длины периода занятости, а через $\pi^*(s)$ — ее преобразование Лапласа—Стилтьеса: $\pi^*(s) = \int_0^\infty e^{-st}\,d\prod(t)$. Очевидно, $\pi^*(s)$ есть вероятность того, что за время периода занятости не наступит катастрофа. Очевидно, длительность периода занятости системы не зависит от того, в каком порядке будут обслуживаться требования. Поэтому без ограничения общности будем считать, что требования обслуживаются в инверсионном порядке: первым обслуживается то требование из очереди, которое пришло последним. При такой дисциплине выбора требований из очереди каждое требование порождает период занятости системы требованиями, поступившими в систему после него. Причем, очевидно, длина этого периода занятости, порожденного некоторым требованием, имеет такое же распределение, как и длина всего периода занятости. Поэтому длина периода занятости ξ имеет следующую структуру:

$$\xi = \beta + \sum_{i=0}^{\nu} \xi_i,$$

где β — время обслуживания требования, с которого начинается период занятости системы; ν — число требований, поступивших в систему за это время обслуживания; $\xi_i,\ i = \overline{1,\nu},$ — длина периода занятости, порожденного i—м требованием.

Требование назовем *плохим*, если за период занятости, порожденный им, наступает катастрофа. Очевидно, $\pi^*(s)$ есть вероятность того, что требование, породившее период занятости является неплохим. Из структуры периода занятости видно, что справедливо утверждение: для того, чтобы требование было неплохим необходимо и достаточно, чтобы за время обслуживания требования, с которого начинается период занятости, не поступило катастрофы, а также плохих требований. Поскольку каждое из требований является плохим с вероятностью $1 - \pi^*(s)$ независимо от других, то поток плохих требований, как результат применения простейшей рекуррентной процедуры просеивания к стационарному пуассоновскому потоку интенсивности λ, является стационарным пуассоновским с интенсивностью $\lambda(1 - \pi^*(s))$. Суммарный поток катастроф и плохих вызовов, как суперпозиция двух стационарных пуассоновских потоков, является стационарным пуассоновским с интенсивностью, равной сумме интенсивностей налагающихся потоков, т. е. равной $s + \lambda(1 - p^*(s))$. Вероятность того, что за время обслуживания первого требования не поступит события из этого потока, согласно вероятностной интерпретации преобразования Лапласа—Стилтьеса, равна

$$\int_0^\infty e^{(-\lambda(1 - \pi^*(s)) + s)t}\,dB(t) = B^*(s + \lambda - \lambda\pi^*(s)),$$

где $B^*(\ )$ — преобразование Лапласа—Стилтьеса распределения времени обслуживания требования.

Из сформулированного выше утверждения заключаем, что справедливо соотношение

$$\pi^*(s) = B^*(s + \lambda - \lambda\pi^*(s)). \tag{2.50}$$

которое является функциональным уравнением относительно неизвестной функции $\pi^*(s)$.

Дифференцируя (2.50) необходимое число раз и учитывая свойство 5 преобразования Лапласа—Стилтьеса, можно получать выражения для моментов распределения периода занятости. В частности, среднее значение π_1 длины периода занятости определяется следующим образом:

$$\pi_1 = \frac{b_1}{1 - \lambda b_1}. \tag{2.51}$$

Можно доказать утверждение о том, что уравнение (2.50) определяет единственную функцию $\pi^*(s)$, аналитическую в полуплоскости $Re\, s > 0$, где выполняется условие $|\pi^*(s)| \leq 1$, и представляющую собой преобразование Лапласа-Стилтьеса распределения длительности периода занятости.

Аналитическое решение уравнения (2.50) известно только для случаев, когда распределение времени обслуживания является экспоненциальным или вырожденным. В общем случае решить уравнение (2.50) можно с помощью метода итераций:

$$\pi^*(s) = \lim_{n\to\infty} \pi_n^*(s),$$

где

$$\pi_n^*(s) = B^*(s + \lambda - \lambda\pi_{n-1}^*(s)), \quad n \geq 1, \tag{2.52}$$
$$|\pi_0^*(s)| \leq 1,$$

при этом скорость сходимости оценивается по формуле

$$|\pi_n^*(s) - \pi^*(s)| \leq \frac{\rho^n}{1-\rho}|\pi_1^*(s) - \pi_0^*(s)|,$$

где $\rho = \lambda b_1,\ \ \rho < 1$.
При этом, если функция $\pi_0(s)$ является вполне монотонной и $\pi_0(0) = 1$, то каждая функция $\pi_n^*(s)$ в итерационном процессе (2.52) является преобразованием Лапласа-Стилтьеса некоторой функции распределения $\prod_n(t)$ и

$$\lim_{n\to\infty} \prod\nolimits_n(t) = \prod(t),$$

где $\prod(t)$ есть функция распределения периода занятости рассматриваемой СМО.

Рекуррентная процедура (2.52) позволяет найти значения функции $\pi^*(s)$ поточечно. Используя знание значений этой функции в точках, можно строить параметрические аппроксимации этих функций.

Удобным видом аппроксимирующей функции является дробно—рациональная функция. Одним из преимуществ таких функций является возможность их обращения путем разложения на простые дроби и использования затем таблиц, задающих прообразы этих дробей. В работах, связанных с приложением ТМО к решению реальных задач часто применяют аппроксимацию функций распределения функциями из более простых классов так, чтобы у этих распределений совпадало заданное заранее число первых начальных моментов. Поскольку, используя (2.50), можно получить выражения для какого угодно числа моментов m_r распределения $\prod(t),\ r = \overline{1, 2k-1}$, удобно аппроксимировать преобразование Лапласа—Стилтьеса $\pi^*(s)$, являющееся решением уравнения (2.50), функцией

$$\hat{\pi}^*(s) = \sum_{i=1}^{k} q_i \frac{\nu_i}{s + \nu_i}. \tag{2.53}$$

При этом, приравнивая моменты распределений, имеем следующую систему нелинейных уравнений относительно величин q_i, ν_i:

$$\sum_{i=1}^{k} q_i = 1,$$

$$r! \sum_{i=1}^{k} \frac{q_i}{\nu_i} = m_r, \quad r = \overline{0, 2k-1}, \tag{2.54}$$

решая которую численными методами с помощью ЭВМ, получаем искомые параметры аппроксимирующего преобразования Лапласа—Стилтьеса (2.53).

Интересной характеристикой СМО является также распределение числа требований, обслуженных за период занятости системы. Найдем это распределение с использованием метода введения дополнительного события для системы $M|G|1$.

Пусть q_k есть вероятность поступления k требований за период занятости системы $M|G|1$, $k \geq 1$. Каждое требование, поступающее в систему, с вероятностью $s,\ 0 < s < 1$, будем называть *красным*. Тогда $Q(s) = \sum_{k=1}^{\infty} q_k s^k$ есть вероятность того, что за период занятости поступили только красные требования. Докажем, что производящая функция $Q(s)$ удовлетворяет следующему функциональному уравнению:

$$Q(s) = sB^*(\lambda - \lambda Q(s)). \tag{2.55}$$

Как и в предыдущем примере, будем считать, что требования обслуживаются в инверсионном порядке. При этом каждое требование порождает период занятости системы требованиями, поступившими после него. Требование назовем *темно-красным*, если оно само является красным и за период занятости, порожденный им, поступали только красные требования. Таким образом, $Q(s)$ есть вероятность того, что требование, с которого начинается период занятости, является темно-красным. Легко видеть, что справедливо утверждение: для того, чтобы требование было темно-красным, необходимо и достаточно, чтобы оно было красным и все требования, поступившие в систему за время его обслуживания, были темно-красными. Вероятность того, что данное требование является красным, равна s . Вероятность того, что за время его обслуживания не поступит нетемно-красных требований равна $B^*(\lambda - \lambda Q(s))$ (см. предыдущий пример). Из сформированного утверждения следует, что $Q(s) = sB^*(\lambda - \lambda Q(s))$. Формула (2.55) доказана.

Уравнение (2.55) также можно решать с использованием процедуры последовательных приближений с последующей интерполяцией поточечно заданной функции функцией из заданного класса, а также путем аппроксимации функцией более простого вида с совпадающими начальными моментами.

Лабораторная работа: "Метод введения дополнительного события и его применение для расчета характеристик системы M|G|1"

Цель работы. Овладение основами метода введения дополнительного события и применение его для расчета характеристик СМО $M|G|1$.

Задание 1. Найти явный вид преобразования $\pi^*(s)$ как решение уравнения (2.50) для системы $M|M|1$, обратить его и реализовать на ЭВМ программу расчета значений функции распределения $\prod(t)$ при фиксированных значениях аргумента t.

Задание 2. Получить выражение для дисперсии длительности периода занятости системы $M|G|1$ и реализовать подсчет ее на ЭВМ для различных видов распределения времени обслуживания.

Задание 3. Доказать существование и единственность корня уравнения (2.55) в интервале $(0, 1)$; получить формулы для среднего значения и дисперсии распределения числа требований, обслуженных за период занятости системы $M|G|1$.

Задание 4. Реализовать на ЭВМ нахождение параметров (q_i, ν_i) для аппроксимирующего распределения $\pi^*(s)$, задаваемого формулой (2.53), для различных видов распределений времени обслуживания и различных значений порядка аппроксимации $k,\ k = \overline{1, 6}$. Проанализировать разброс значений моментов распределений с номерами, большими, чем $2k - 1$.

Задание 5. Построить аппроксимацию производящей функции $Q(s)$, являющуюся решением уравнения (2.55), гипергеометрической и реализовать на ЭВМ расчет параметров этого распределения.

2.6 Система типа $M|G|1$ с управляемым режимом функционирования

В предыдущих параграфах мы рассматривали в основном вопросы расчета характеристик классических СМО. В данном и в последующих параграфах этой главы рассмотрим вопросы расчета характеристик СМО с изменяемым режимом функционирования. Такие СМО являются адекватной математической моделью процессов, происходящих во многих фрагментах и узлах современных сетей связи и сетей ЭВМ.

Рассмотрим однолинейную СМО с ожиданием, которая может работать в $N,\ N \geq 2$, режимах. При работе системы в r—м режиме интервалы между моментами поступления требований в систему имеют экспоненциальное распределение с параметром λ_r, время обслуживания требования имеет функцию распределения $B_r(t)$ с преобразованием Лапласа—Стилтьеса $B_r^*(s)$ и конечными начальными моментами $b_m^{(r)} = \int_0^\infty t^m \, dB_r(t),\ m \geq 1,\ r = \overline{1, N}$. Изменение режима функционирования СМО возможно в моменты окончания обслуживания требований.

Имеется также экономический функционал качества функционирования СМО, имеющий вид

$$I = aV + \sum_{r=1}^{N} c_r P_r, \tag{2.56}$$

где V — среднее время пребывания требования в системе; a — штраф за пребывание требования в системе в единицу времени; P_r — средняя доля времени использования r—го режима; c — стоимость единицы времени использования r—го режима; $r = \overline{1, N}$. Предполагаем, что режимы работы перенумерованы таким образом, что

$$\rho_1 > \rho_2 > \ldots > \rho_N, \tag{2.57}$$

где $\rho_r = \lambda_r b_1^{(r)}$ — загрузка системы при работе в r—м режиме. Таким образом, первый режим является самым "тяжелым" ("медленным"),..., а N—й режим — самым "легким" ("быстрым"). Также будем предполагать, что стоимостные коэффициенты удовлетворяют условиям

$$a > 0, \quad c_1 \leq c_2 \leq \ldots \leq c_N \tag{2.58}$$

т. е. штраф за пребывание требований в системе не равен нулю, самый медленный режим — самый дешевый,..., самый быстрый режим обслуживания — самый дорогой.

При выполнении условий (2.57), (2.58) видно, что задача нахождения оптимальной стратегии выбора режимов обслуживания требований является нетривиальной. Использование более медленных режимов позволяет уменьшить плату за использование режимов, однако при этом возрастает время пребывания требований в системе. Использование более быстрых режимов позволяет уменьшить время пребывания требований в системе, но при этом возрастает штраф за использование режимов. В такой ситуации интуитивно очевидным решением является следующее - динамически изменять режим функционирования системы: при увеличении текущего значения величины очереди в системе увеличивать скорость обслуживания, при уменьшении величины очереди выбирать более медленный (и дешевый) режим. Правильность этого интуитивного решения доказана рядом авторов. Ими показано, что оптимальная стратегия управления режимами функционирования СМО является многопороговой, т. е. стратегией следующего вида. Фиксируются натуральные числа $j_1, \dots, j_{N-1}$, удовлетворяющие неравенствам $-1 = j_0 < j_1 \leq j_2 \dots \leq j_{N-1} < j_N = \infty$ и называемые *порогами*. Если число i требований в системе в данный момент окончания обслуживания требования удовлетворяет неравенству $j_{l-1} < i \leq j_l$, то следующее требование будет обслуживаться в l–м режиме, $l = \overline{1, N}$. Отметим, что если $j_{r-1} = j_r$, то r–й режим не используется в работе системы.

Разработаем алгоритм нахождения оптимального в смысле минимизации функционала (2.56) набора порогов $(j_1^*, \dots, j_{N-1}^*)$ при заданном наборе значений параметров системы и стоимостных коэффициентов. Для этого найдем стационарное распределение числа требований в данной системе в моменты окончания обслуживания требований при фиксированном наборе порогов $j_1, \dots, j_{N-1}$. Затем найдем явную зависимость функционала (2.56) от порогов $(j_1, \dots, j_{N-1})$. Тем самым задача будет сведена к нахождению минимума функции $N-1$ целочисленных переменных, которую можно будет решить с помощью ЭВМ.

Задачу нахождения стационарного распределения числа требований в системе в моменты окончания обслуживания требований будем решать с использованием уже известного нам метода вложенных цепей Маркова.

Будем обозначать, как и ранее, $N(t)$ — число требований в системе в момент t, $N(\tau_n)$ — число требований в системе в момент окончания обслуживания n-го требования, $n \geq 1$,

$$\pi_i = \lim_{n \to \infty} \mathbf{P}\{N(\tau_n) = 1\}, \quad i \geq 0, \tag{2.59}$$

является стационарной вероятностью наличия i требований в системе в моменты окончания обслуживания требований,

$$f_i^{(l)} = \int_0^\infty \frac{(\lambda_l t)^i}{i!} \mathrm{e}^{-\lambda_l t}\, dB_l(t), \quad i \geq 0,\ l = \overline{1, N},$$

является вероятностью поступления в систему i требований за время обслуживания одного требования при работе системы в l–м режиме. Можно доказать, что необходимым и достаточным условием существования пределов (2.59) является выполнение неравенства $\rho_N < 1$. Обратим внимание, что величина загрузки ρ_r системы при работе в r–м режиме, $r = \overline{1, N-1}$, может быть и больше единицы.

Используя формулу полной вероятности, можно убедиться, что стационарные вероятности $\pi_i,\ i \geq 0$, удовлетворяют следующей системе линейных алгебраических уравнений:

$$\pi_i = \pi_0 f_i^{(1)} + \sum_{k=1}^{i+1} \pi_k f_{i-k+1}^{(1)}, \quad i = \overline{0, j_1 - 1},$$

$$\pi_{j_r+l} = \pi_0 f_{j_r+l}^{(1)} + \sum_{m=0}^{r-1} \sum_{k=1}^{j_{m+1}-j_m} \pi_{j_m+k} f_{j_r+l-j_m-k+1}^{(m+1)} +$$

$$+ \sum_{k=1}^{l+1} \pi_{j_r+k} f_{l-k+1}^{(r+1)}, \quad l = \overline{0, j_{r+1} - j_r - 1},\ r = \overline{1, N},\ j_0 = 0. \tag{2.60}$$

Для решения системы (2.60) будем использовать модификацию аппарата производящих функций — аппарат частичных производящих функций. Введем в рассмотрение так называемые частичные производящие функции

$$\prod\nolimits_l(z) = \sum_{m=j_{l-1}+1}^{j_l} \pi_m z^m, \quad l = \overline{1, N},\ |z| < 1,$$

и производящую функцию

$$\prod(z)=\sum_{i=0}^{\infty}\pi_i z^i=\sum_{l=1}^{N}\prod\nolimits_l(z).$$

Умножая уравнения системы (2.60) на соответствующие степени z и суммируя их, получаем следующее соотношение для частичных производящих функций $\prod_l(z),\ l=\overline{1,N}$:

$$\pi_0 B_1^*(\lambda_1(1-z))(1-z)+\sum_{l=1}^{N}\prod\nolimits_l(z)(z-B_l^*(\lambda_l(1-z)))=0. \tag{2.61}$$

Отметим, что соотношение (2.61), предварительно переписав его в виде

$$z\sum_{l=1}^{N}\prod\nolimits_l(z)=\sum_{l=2}^{s}\prod\nolimits_l(z)B_l^*(\lambda_l(1-z))+\left(\prod\nolimits_l(z)-\pi_0\right)B_1^*(\lambda_1(1-z))+$$
$$+z\pi_0 B_l^*(\lambda_1(1-z)), \tag{2.62}$$

можно получить и с помощью метода введения дополнительного события. Каждое из требований, поступающих в систему, с вероятностью z назовем *красным*. Тогда $\prod_l(z)$ есть вероятность того, что в данный момент окончания обслуживания требования в системе от $j_{l-1}+1$ до j_l требований и все они красные. Выражение в левой части (2.62) есть вероятность события: в данный момент окончания обслуживания требования все остающиеся требования — красные и уходящее требование — тоже красное. Нетрудно видеть, что для того, чтобы данное событие произошло, необходимо и достаточно, чтобы в предыдущий момент окончания обслуживания требования — в системе были только красные требования и все требования, поступившие за время обслуживания были красными, при этом учитываем, что вероятность поступления только красных вызовов есть $B_l^*(\lambda_l(1-z))$, если обслуживание ведется в l–м режиме; так же необходимо и достаточно, чтобы в системе не оказалось требований, требование, пришедшее в систему первым, оказалось красным и за время его обслуживания в систему пришли только красные требования. Подсчитывая вероятности вышеописанных событий и получаем (2.62) и, автоматически, (2.61).

Второй способ получения уравнения (2.61) более элегантен — не требуется находить переходные вероятности цепи Маркова, выписывать уравнения для ее стационарных вероятностей и выполнять операции по суммированию и изменению порядка суммирования рядов. Однако, в этом же состоит и его недостаток. Фактически (2.61) дает нам одно уравнение для N неизвестных функций $\prod_l(z),\ l=\overline{1,N}$. И требуется предпринять еще какие-то усилия для их нахождения. Используя метод вложенных цепей Маркова и выписывая при этом систему (2.60) для стационарных вероятностей π_i, мы выясняем специфику матрицы переходных вероятностей цепи Маркова и обнаруживаем, что из i–го уравнения мы можем найти зависимость вероятности π_{i+1} от вероятностей $\pi_l,\ l=\overline{0,i}$, и, в конечном счете, от π_0. При этом получаем соотношение вида $\pi_{i+1}=\alpha_{i+1}\pi_0$, где вид коэффициентов α_{i+1} зависит только от коэффициентов первых i уравнений системы и не зависит от коэффициентов последующих уравнений. Таким образом, например, мы сразу можем заключить, что

$$\pi_i=\alpha_i^{(1)}\pi_0,\quad i=\overline{0,j_1},$$

где

$$\alpha_i^{(1)}=\frac{1}{i!}\frac{\partial^i}{\partial z^i}\frac{(1-z)B_1^*(\lambda_1(1-z))}{B_1^*(\lambda_1(1-z))-z}\bigg|_{z=0},$$

поскольку первые j_1 уравнений системы (2.60) совпадают с уравнениями (2.21) для классической СМО $M|G|1$, работающей в одном только первом режиме, а для этой системы известно выражение (2.22) для производящей функции. При нахождении коэффициентов $\alpha_i^{(2)}$ в выражении $\pi_i=\alpha_i^{(2)}\pi_o,\ i=\overline{j_1+1,j_2}$, мы, пользуясь независимостью вида этих коэффициентов от вида коэффициентов уравнений с номерами, большими, чем i, можем временно "забыть" о том, что система может переключаться на режимы с номерами большими, чем 2. При этом считаем, что у системы есть всего два режима функционирования ($N = 2$), и из (2.61) получаем:

$$\pi_0 B_1^*(\lambda_1(1-z))(1-z) + (z - B_1^*(\lambda_1(1-z)))\prod\nolimits_1(z) =$$

$$= \prod\nolimits_2(z)(B_2^*(\lambda_2(1-z)) - z). \tag{2.63}$$

Поскольку функция $\prod_1(z)$ с точностью до значения π_0 нам уже известна, из (2.63) получаем явный вид производящей функции $\prod_2(z)$. Разлагая ее в ряд Маклорена, получаем вид коэффициентов $\alpha_i^{(2)},\ i = \overline{j_1+1, j_2}$. Повторяя эту процедуру, убеждаемся в справедливости следующего утверждения.

Теорема 2.1 *Производящая функция $\prod(z)$ числа требований в рассматриваемой системе имеет вид*

$$\prod(z) = \sum_{l=1}^{N} \prod\nolimits_l(z) = \pi_0 \times$$

$$\times \Bigg(B_1^*(\lambda_1(1-z))(z-1) + \sum_{l=1}^{N-1} \sum_{m=j_{l-1}+1}^{j_l} \alpha_m^{(l)} z^m (B_l^*(\lambda_l(1-z)) -$$

$$B_N^*(\lambda_N(1-z))\Bigg) \,/\, (z - B_N^*(\lambda_N(1-z))), \tag{2.64}$$

где величины $\alpha_m^{(l)}$ подсчитываются рекуррентным образом:

$$\alpha_m^{(l)} = \sum_{r=1}^{l-1} \sum_{i=j_{r-1}+1}^{j_r} \alpha_i^{(r)} \Delta_{m-i}(r,l) + \gamma_m(l), \tag{2.65}$$

с начальным условием

$$\alpha_m^{(1)} = \gamma_m(1), \quad m = \overline{0, j_1}, \tag{2.66}$$

где

$$\gamma_m(l) = \frac{1}{m!} \frac{d^m}{dz^m} \left(\frac{(1-z)B_1^*(\lambda_1(1-z))}{B_l^*(\lambda_l(1-z)) - z} \right)\Bigg|_{z=0},$$

$$\Delta_m(r,l) = \frac{1}{m!} \frac{d^m}{dz^m} \left(\frac{z - B_r^*(\lambda_r(1-z))}{z - B_l^*(\lambda_l(1-z))} \right)\Bigg|_{z=0}. \tag{2.67}$$

При численном подсчете величин $\gamma_m(l)$ и $\Delta_m(r,l)$ более удобно использовать не формулы (2.67), а следующие рекуррентные формулы:

$$\Delta_0(r,l) = -f_o^{(r)} \,/\, f_0^{(l)}, \quad \Delta_1(r,l) = -(-\Delta_0(r,l)(1 - f_1^{(l)}) - 1 + f_1^{(r)}) \,/\, f_0^{(l)},$$

$$\Delta_m(r,l) = -(\Delta_{m-1}(r,l) + \sum_{k=0}^{m-1} \Delta_k(r,l) f_{m-k}^{(l)} + f_m^{r)}) \,/\, f_0^{(r)}, \quad m \geq 2, \tag{2.68}$$

$$\gamma_m^{(l)} = f_m^{(1)} + \sum_{i=0}^{m} f_i^{(1)(l)} \alpha_{m-i+1}, \quad m \geq 0,$$

где

$$
{}^{(l)}\alpha_0 = 1, \quad {}^{(l)}\alpha_{m+1} = \frac{1}{f_0^{(l)}} \left({}^{(l)}\alpha_m - \sum_{k=1}^{m} {}^{(l)}\alpha_k f_{m-k+1}^{(l)} - f_m^{(l)} \right), \quad m \ge 0.
$$

Величину вероятности π_0 в (2.64), как обычно, можно найти из условия нормировки $\prod(1) = 1$. Она имеет вид

$$
\pi_0 = (1 - \rho_n) \bigg/ \left(1 + \sum_{l=1}^{N-1} \sum_{m=j_{l-1}+1}^{j_l} \alpha_m^{(l)} (\rho_l - \rho_N) \right). \tag{2.69}
$$

Имея формулу (2.64) для производящей функции $\prod(z)$ и используя те же рассуждения, что и в п. 2.3, в случае $\lambda_1 = \ldots = \lambda_N = \lambda$, несложно получить и выражение для преобразования $v(s)$ распределения времени пребывания требования в рассматриваемой системе:

$$
v(s) = \pi_0 \times
$$
$$
\times \frac{B_1^*(s) - \frac{\lambda}{s} \sum_{l=1}^{N-1} \sum_{m=j_{l-1}+1}^{j_l} \alpha_m^{(l)} \left(1 - \frac{s}{\lambda}\right)^m (B_l^*(s) - B_N^*(s))}{1 - \lambda \dfrac{1 - B_N^*(s)}{s}}, \tag{2.70}
$$

а также значение для среднего времени V пребывания требования в системе

$$
V = -v'(0) = \lambda \prod{}'(1). \tag{2.71}
$$

Будем далее предполагать, что режимы функционирования СМО различаются только распределением времени обслуживания, а интенсивность входного потока является неизменной. Покажем схематично, что величины P_l средней доли времени использования l—го режима вычисляются следующим образом:

$$
P_l = \rho_l \prod{}_l(1), \quad l = \overline{1, N}. \tag{2.72}
$$

Пусть $[0, T]$ — достаточно большой промежуток времени; $P_l(T)$ — суммарное время использования l—го режима в интервале $[0, T]$.

По определению

$$
P_l = \lim_{T \to \infty} \frac{P_l(T)}{T}.
$$

Введем обозначения:

$M(T)$ — число требований, обслуженных за время T;

$N_l(T)$ — число требований, обслуженных за время T в l—м режиме;

$\xi_m^{(l)}$ — время обслуживания m-го требований из обслуженных в l—м режиме;

Очевидно,

$$
\lim_{T \to \infty} \frac{P_l(T)}{T} = \lim_{T \to \infty} \frac{M(T)}{T} \, \frac{N_l(T)}{M(T)} \, \frac{P_l(T)}{N(T)}, \tag{2.73}
$$

причем, если существуют пределы всех трех сомножителей справа, предел слева существует и равен произведению соответствующих пределов.

Предел первого из сомножителей существует в силу существования стационарного распределения вероятностей состояний СМО и эргодичности простейшего потока. Он равен λ.

Введем обозначения:

$$
\eta_m = f(i_{t_m}) = \begin{cases} 1, & \text{если } j_{l-1} < i_{t_m} \le j_l, \\ 0, & \text{в противном случае.} \end{cases}
$$

Очевидно, что $M\eta_m = Mf(i_{t_m} = \prod_l(1)$, а также что

$$N_l(T) = \sum_{m=1}^{M(T)} \eta_m.$$

Тогда с вероятностью 1 существует предел

$$\lim_{T\to\infty} \frac{N_l(T)}{M(T)} = \lim_{T\to\infty} \frac{\sum\limits_{m=1}^{M(T)} \eta_m}{M(T)} = \prod\nolimits_l(1).$$

Далее, очевидно, что случайные величины $\xi_m^{(l)}$ являются независимыми одинаково распределенными случайными величинами с конечным математическим ожиданием $b_1^{(l)} = M\xi_m^{(l)},\ m \geq 1$. Согласно теореме Колмогорова к последовательности случайных величин $\xi_m^{(l)},\ m \geq 1$, применим усиленный закон больших чисел. Следовательно, существует

$$\lim_{T\to\infty} \frac{P_l(T)}{N_l(T)} = \lim_{T\to\infty} \frac{\sum\limits_{m=1}^{N_l(T)} \xi_m^{(l)}}{N_l(T)} = b_1^{(l)}.$$

Таким образом, пределы всех трех сомножителей справа в (2.73) существуют и равны, соответственно, λ, $\prod_l(1)$, $b_1^{(l)}$, откуда и следует (2.72).

Подставляя полученные нами выражения для величин V, $P_l,\ l = \overline{1, N}$, в (2.56), получаем следующую явную зависимость значений функционала (2.56) от порогов $j_1, \dots, j_{N-1}$:

$$I = \frac{\xi_0 + \sum\limits_{l=1}^{N-1} \xi_l \sum\limits_{m=j_{l-1}+1}^{j_l} \alpha_m^{(l)} + \tilde{a} \sum\limits_{l=1}^{N-1} (\rho_l - \rho_N) \sum\limits_{m=j_{l-1}+1}^{j_l} m\alpha_m^{(l)}}{1 + \sum\limits_{l=1}^{s-1} (\rho_l - \rho_N) \sum\limits_{m=j_{l-1}+1}^{j_l} \alpha_m^{(l)}}, \tag{2.74}$$

где константы $\xi_l,\ l = \overline{0, N-1}$, и $\tilde{a}$ имеют следующий вид

$$\tilde{a} = \lambda^{-1} a, \quad \xi_0 = \tilde{a} \left(\rho_1 + \frac{\lambda^2 b_2^{(N)}}{2(1-\rho_N)} \right) + c_N \rho_N, \tag{2.75}$$

$$\xi_l = \tilde{a} \left(\frac{\lambda^2 b_2^{(l)}}{2} - \frac{\lambda^2 b_2^{(N)}}{2} \frac{1-\rho_l}{1-\rho_N} \right) + c_l \rho_l (1-\rho_N) - c_N \rho_N (1-\rho_l), \quad l = \overline{1, N-1}.$$

Для системы с двумя возможными режимами работы выражение (2.74) переходит в

$$I = I(j_1) = \frac{\xi_0 + \xi_1 \sum\limits_{i=0}^{j_1} \alpha_i^{(1)} + \tilde{a}(\rho_1 - \rho_2) \sum\limits_{i=0}^{j_1} i\alpha_i^{(1)}}{1 + (\rho_1 - \rho_2) \sum\limits_{i=0}^{j_1} \alpha_i^{(1)}}. \tag{2.76}$$

Имея выражения (2.74), (2.76), задающие зависимость значения функционала качества от порогов, с помощью ЭВМ несложно построить и реализовать процедуру нахождения оптимального набора порогов $(j_1^*, \dots, j_{N-1}^*)$.

Лабораторная работа: "Нахождение характеристик и оптимизация параметров стратегии управления системой типа M|G|1 с управляемым режимом функционирования"

Цель работы. Ознакомление с управляемой СМО, использованием аппарата частичных производящих функций для нахождения их характеристик, оптимизация функционирования СМО.

Задание 1. Реализовать на ЭВМ рекуррентную процедуру расчета коэффициентов $\alpha_m^{(l)}$ в представлениях $\pi_i = \alpha_i^{(l)}\pi_0,\ i = \overline{j_{l-1}+1, j_l}$, задаваемую соотношениями (2.65), (2.66), (2.68).

Указание. Расчет величин $f_m^{(l)}$ и ${}^{(l)}\alpha_m$ фактически уже реализован для различных видов распределения времени обслуживания требований при выполнении задания 6 в п. 2.3.

Задание 2. Реализовать на ЭВМ процедуру расчета значения критерия качества (2.56) при фиксированном наборе порогов $(j_1, \dots, j_{N-1})$, основанную на использовании формул (2.74), (2.75).

Задание 3. Реализовать на ЭВМ процедуру нахождения оптимального набора порогов $(j_1^*, \dots, j_{N-1}^*)$.

Указание. Для генерации всевозможных значений наборов порогов $(j_1, \dots, j_{N-1})$ использовать стандартную прогрумму COMBIN, текст которой содержится в приложении.

Задание 4. Упростить явную зависимость значения критерия качества функционирования СМО с двумя возможными режимами функционирования, задаваемую формулой (2.76), в случае, когда распределение времени обслуживания требований является экспоненциальным. Условно считая параметр j_1 непрерывным, получить трансцендентное уравнение для нахождения оптимального значения j_1^* этого порога. Реализовать на ЭВМ процедуру нахождения оптимального значения порога.

Задание 5. Для системы с двумя возможными режимами функционирования исследовать вопрос о зависимости величины выигрыша, даваемого использованием пороговой стратегии, по сравнению с использованием только одного из возможных режимов (при этом, очевидно, значение критерия качества равно

$$\left[b_1^{(l)} + \frac{\lambda^2 b_2^{(l)}}{2(1-\rho_l)}\right] a + c_l\rho_l),$$

от величины коэффициентов вариации распределений $B_l(t),\ \ l = \overline{1,2}$.

2.7 Система типа $GI|M|1$ с управляемым режимом функционирования.

Рассмотрим модель, близкую к рассмотренной в предыдущем параграфе, и отличающуюся от нее тем, что предполагается марковость обслуживания, а не входного потока. Как мы уже знаем из п. 2.1, система типа $G|G|1$ не поддается полному исследованию даже в классическом варианте. Поэтому для управляемых систем с входным потоком, отличным от простейшего, приходится делать упрощающее исследование предположение о том, что время обслуживания требований имеет экспоненциальное распределение.

Итак, рассмотрим однолинейную систему с ожиданием, которая может работать в $N,\ N \geq 2$, режимах. При работе системы в r—м режиме интервалы между моментами поступления требований имеют функцию распределения $A_r(t)$ с интенсивностью $\lambda_r = (\int_0^\infty (1 - A_r(t))\,dt)^{-1}$ и преобразованием Лапласа—Стилтьеса $A_r^*(s)$, а время обслуживания требования имеет экспоненциальное распределение с параметром $\mu_r,\ r = \overline{1,N}$. Изменение режима работы системы возможно в моменты поступления требований в систему. Качество функционирования системы оценивается функционалом типа (2.56).

Оптимальную стратегию также будем искать в классе многопороговых стратегий. При фиксированном наборе порогов $(j_1, \dots, j_{N-1})$ будем использовать l-й режим функционирования, если в данный момент принятия решения (момент поступления в систему требования) число i требований в системе удовлетворяет неравенству $j_{l-1} + 1 \leq i \leq j_l,\ l = \overline{1,N}$.

Требуется указать набор порогов $(j_1^*, \dots, j_{N-1}^*)$, доставляющий минимум функционалу (2.56).

Пусть $\tau_1, \tau_2, \dots, \tau_n$ — моменты поступления требований в систему,

$$r_i = \lim_{n\to\infty} \mathbf{P}\{N(\tau_n) = i\}, \quad i \geq 0. \tag{2.77}$$

Используя теорию цепей Маркова, можно доказать, что необходимым и достаточным условием существования пределов (2.77) является, как и в задаче, рассмотренной в предыдущем параграфе, выполнение неравенства $\rho_N < 1$. Считаем это условие выполненным.

Теорема 2.2 *Стационарное распределение r_i, $i \geq 0$, определяется следующим образом:*

$$r_i = C\delta^i \prod_{\nu=t}^{N-1} \sum_{r_\nu=0}^{j_\nu - i - \sum\limits_{l=t}^{\nu-1} r_l} \omega_{r_\nu}^{(\nu)} \delta^{r_\nu}, \tag{2.78}$$

$$j_{t-1} + 1 \leq i \leq j_t, \quad t = \overline{1, N-1},$$

$$r_i = C\delta^i, \quad i > j_{N-1}, \tag{2.79}$$

где символ

$$\prod_{\nu=t}^{N-1} \sum_{r_\nu=0}^{j_\nu - i - \sum\limits_{l=t}^{\nu-1} r_l} \omega_{r_\nu}^{(\nu)} \delta^{r_\nu}$$

означает в традиционных обозначениях следующее

$$\sum_{r_t=0}^{j_t - i} \omega_{r_t}^{(t)} \delta^{r_t} \sum_{r_{t+1}=0}^{j_{t+1} - i - r_t} \omega_{r_{t+1}}^{(t+1)} \delta^{r_{t+1}} \cdots \sum_{r_{N-1}=0}^{j_{N-1} - i - \sum_{l=t}^{N-1} r_l} \omega_{r_{N-1}}^{(N-1)} \delta^{r_{N-1}},$$

а константа C имеет вид

$$C = \left\{ \delta^{j_{N-1}+1}(1-\delta)^{-1} + \sum_{t=1}^{N-1} \sum_{i=j_{t-1}+1}^{j_t} \delta^i \prod_{\nu=t}^{N-1} \sum_{r_{N-1}=0}^{j_\nu - i - \sum\limits_{l=t}^{N-1} r_l} \omega_{r_{N-1}}^{(N-1)} \delta^{r_{N-1}} \right\}^{-1}, \tag{2.80}$$

где величины $\omega_m^{(\nu)}$ подсчитываются по следующим рекуррентным формулам:

$$\omega_0^{(\nu)} = \frac{\varphi^{(\nu+1)}}{\varphi_0^{(\nu)}}, \quad \omega_1^{(\nu)} = \frac{1}{\varphi_0^{(\nu)}} (\omega_0^{(\nu)} - \omega_0^{(\nu)} \varphi_1^{(\nu)} - 1 + \varphi_1^{(\nu+1)}),$$

$$\omega_m^{(\nu)} = \left(\omega_{m-1}^{(\nu)} - \sum_{l=0}^{m-1} \omega_l^{(\nu)} \varphi_{m-l}^{(\nu)} + \varphi_m^{(\nu+1)} \right) \Big/ \varphi_0^{(\nu)}, \tag{2.81}$$

$$m \geq 2, \quad \nu = \overline{1, N-1},$$

или через свои производящие функции

$$\Omega_\nu(z) = \sum_{m=0}^{\infty} \omega_m^{(\nu)} z^m = \frac{z - A_{\nu+1}^*(\mu_{\nu+1}(1-z))}{z - A_\nu^*(\mu_\nu(1-z))}, \quad \nu = \overline{1, N-1}.$$

Здесь $\varphi_l^{(\nu)}$ *— вероятность того, что за время между моментами поступления требований при работе системы в* $\nu-$*м режиме произошло* l *событий стационарного пуассоновского потока с интенсивностью* μ_ν*,*

$$\varphi_l^{(\nu)} = \int_0^\infty \frac{(\mu_\nu t)^l}{l!} e^{-\mu_\nu t}\, dA_\nu(t), \quad l \geq 0, \quad \nu = \overline{1, N-1}.$$

Величина δ *является действительным корнем уравнения*

$$\delta - A_N^*(\mu_N(1-\delta)) = 0 \tag{2.82}$$

в интервале $0 < \delta < 1$.

Доказательство. С помощью формулы полной вероятности можно убедиться, что стационарные вероятности r_i удовлетворяют системе линейных алгебраических уравнений вида:

$$r_i = \sum_{l=i-1}^{j_t} r_l \varphi_{l-i+1}^{(t)} + \sum_{\nu=t}^{N-1} \sum_{j_\nu+1}^{l=j_{\nu+1}} r_l \varphi_{l-i+1}^{(\nu+1)}, \tag{2.83}$$

$$j_{t-2} + 2 \leq i \leq j_t + 1, \quad t = \overline{1, N}, \quad i > 0.$$

Анализируя вид матрицы переходных вероятностей данной цепи Маркова, видим следующую закономерность: стационарные вероятности r_i с точностью до значения нормирующей константы C полностью определяются через вероятности состояний с номерами, большими, чем i . При этом уравнения (2.83) для $i > j_{N-1}$ полностью совпадают с соответствующими уравнениями для стационарных вероятностей классической СМО $GI|M|1$, работающей постоянно в $N-$м режиме. Из результатов 2.4 следует, что вероятности с номерами, большими j_{N-1}, для рассматриваемой системы имеют вид (2.79), где константа δ является решением уравнения (2.82). Справедливость формул (2.78) можно доказать с помощью метода математической индукции, предварительно переписав их в несколько другом виде:

$$r_i = C\delta^i \sum_{s=0}^{j_{N-1}-i} \delta^s \left(\prod_{\nu=t}^{N-2} \sum_{r_\nu=0}^{\min\{s, j_\nu - i\} - \sum\limits_{l=t}^{\nu-1} r_l} \omega_{r_\nu}^{(\nu)} \right) \omega^{(N-1)}_{s - \sum\limits_{l=t}^{N-2} r_l}, \tag{2.84}$$

$$j_{t-1} + 1 \leq i \leq j_t, \quad t = \overline{1, N-2},$$

и сравнивая коэффициенты при одинаковых степенях δ при постановке выражения (2.84) в (2.83). При этом для величин $\omega_m^{(\nu)}$ будут получены рекуррентные формулы (2.81).

Значение константы C находится из условия нормировки: $\sum\limits_{i=0}^{\infty} r_i = 1$, в результате имеем (2.80). Теорема доказана.

Если число возможных режимов равно двум, стационарное распределение $r_i,\ i \geq 0$, задается следующим образом:

$$r_i = C\delta^i \sum_{l=0}^{j_1-i} \omega_l^{(1)} \delta^l, \quad i = \overline{0, j_1},$$

$$r_i = C\delta^i, \quad i > j_1,$$

$$C = (1-\delta)\left[\sum_{l=0}^{j} \omega_l^{(1)}\delta^l + \delta^{j_1+1}\left(1 - \sum_{l=0}^{j_1}\omega_l^{(1)}\right)\right]^{-1}. \qquad (2.85)$$

Если при этом распределения интервалов между требованиями экспоненциальные: $A_l(\tau) = 1 - \exp(-\lambda_l \tau)$, $l = \overline{1,N}$, явный вид величин следующий:

$$\omega_0^{(\nu)} = \frac{\rho_{\nu+1}}{\rho_\nu}\frac{1+\rho_\nu}{(1+\rho_{\nu+1})},$$

$$\omega_i^{(\nu)} = \frac{\rho_{\nu+1}-\rho_\nu}{1+\rho_{\nu+1}-\rho_\nu}\left(\rho_\nu^{-i-1} - (1+\rho_{\nu+1})^{-i-1}\right), \quad i \geq 1. \qquad (2.86)$$

Используя полученные выражения для вероятностей r_i, можем найти явную зависимость значения критерия (2.56), сделав предварительно предположение о том, что режимы работы системы различаются только распределением интервалов между моментами поступления требований, интенсивности обслуживания во всех режимах одинаковы: $\mu_1 = \ldots = \mu_n = \mu$. При выполнении этого предположения легко видеть, что величина V подсчитывается следующим образом:

$$V = \sum_{i=0}^{\infty} r_i \frac{i+1}{\mu}. \qquad (2.87)$$

Учитывая, что вероятность того, что после произвольного момента поступления требования система будет работать в m−м режиме, равна $\sum\limits_{i=j_{m-1}+1}^{j_m} r_i$, а также что при этом среднее время до следующего момента поступления требования будет равно λ_m^{-1}, несложно убедиться, что величины , P_m, $m = \overline{1,N}$, входящие в функционал качества (2.56), определяются следующим образом:

$$P_m = \lambda_m^{-1}\sum_{i=j_{m-1}+1}^{j_m} r_i \Bigg/ \sum_{l=1}^{N}\lambda_l^{-1}\sum_{i=j_{l-1}+1}^{j_l} r_i, \quad m = \overline{1,N}. \qquad (2.88)$$

Подставляя (2.87) и (2.88) в (2.56), получаем явную зависимость значения функционала качества (2.56) от порогов $(j_1, \ldots, j_{N-1})$ и, тем самым, возможность построения процедуры нахождения оптимального набора порогов $(j_1^*, \ldots, j_{N-1}^*)$.

Лабораторная работа: "Расчет характеристик и оптимизация параметров стратегии управления режимами работы системы типа GI|M|1“

Цель работы. Овладение навыком использования метода вложенных цепей Маркова для расчета характеристик СМО с неоднородной матрицей переходных вероятностей, а также их оптимизации.

Задание 1. Реализовать на ЭВМ расчет значений вероятностей $\varphi_l^{(\nu)}$ обслуживания l требований за время между моментами приходов требований при работе системы в ν−м режиме, $\nu = \overline{1,N}$, $l \geq 0$. Для различных видов распределений $A_\nu(t)$ предварительно получить аналитические формулы для расчета.

Задание 2. Реализовать на ЭВМ процедуру рекуррентного расчета значений коэффициентов $\omega_m^{(\nu)}$ для различного вида распределений $A_\nu(t)$ $\nu = \overline{1,N}$, $m \geq 0$.

Задание 3. Используя (2.87) и (2.88), получить явную зависимость значения функционала качества (2.56) от порогов $(j_1, \ldots, j_{N-1})$. Реализовать на ЭВМ процедуру подсчета значения критерия качества при фиксированном наборе порогов.

Задание 4. Разработать и реализовать на ЭВМ процедуру нахождения оптимального набора порогов $(j_1^*, \ldots, j_{N-1}^*)$.

Задание 5. Для системы с двумя возможными режимами функционирования исследовать влияние коэффициентов вариации распределений $A_l(t)$, $l = \overline{1, N}$, на величину относительного выигрыша по сравнению со стратегией: использовать только один режим функционирования (при этом значение функционала качества равно $\frac{a}{\mu}\left(\frac{1}{1-\delta^{(l)}}\right) + c_l$, где $\delta^{(l)}$ — решение уравнения $\delta^{(l)} = A_l^*(\mu(1-\delta^{(l)})$ в области $0 < \delta^{(l)} < 1$).

Задание 6. Для системы с двумя возможными режимами функционирования и экспоненциальным распределением интервалов между требованиями найти явную зависимость значения критерия качества от порога j_1 и реализовать на ЭВМ процедуру нахождения оптимального значения порога j_1^*.

2.8 Система типа $M|G|1$ с двумя возможными режимами функционирования и ненадежным прибором

Во всех моделях СМО, рассматриваемых в данной главе, предполагалось, что обслуживающий прибор является абсолютно надежным, т. е. он постоянно находится в готовности к обслуживанию требований. В реальных системах это не всегда имеет место. Поэтому изучим одну из моделей СМО, в которой возможен выход обслуживающего прибора из строя.

Эта модель отличается от модели, изученной в параграфе 2.6, тем, что при попадании системы в свободное состояние через время, имеющее функцию распределения $D(t)$ с преобразованием Лапласа—Стилтьеса $D^*(s)$, прибор может выйти из строя. В этом случае он восстанавливается в течение времени, имеющего функцию распределения $Q(t)$ с преобразованием Лапласа—Стилтьеса $Q^*(s)$. После этого прибор вновь оказывается способным к обслуживанию требований. В занятом состоянии прибор не может выходить из строя. Кроме того, предположим, что число N возможных режимов функционирования СМО равно двум.

Пусть, как и раньше,

$$\pi_i = \lim_{n\to\infty} \mathbf{P}\{N(\tau_n) = i\}, \quad i \geq 0. \tag{2.89}$$

Можно убедиться, что необходимым и достаточным условием существования пределов (2.89) является выполнение условий

$$\rho_2 < 1, \quad q_1 = \int_0^\infty (1 - Q(t))\, dt < +\infty.$$

Будем считать эти условия выполненными.

Введем обозначение: φ_m — вероятность того, что если после окончания обслуживания некоторого требования система оказалась пустой, то в следующий после этого момент начала обслуживания требования в системе будет m требований, $m \geq 1$.

Используя изученный выше метод введения дополнительного события, можно убедиться, что производящая функция $\Phi(z) = \sum\limits_{m=1}^{\infty} \varphi_m z^m$ распределения φ_m, $m \geq 1$, имеет следующий вид:

$$\Phi(z) = \frac{z(1 - D^*(\lambda_1)) + D^*(\lambda_1)(Q^*(\lambda_1(1-z)) - Q^*(\lambda_1))}{(1 - D^*(\lambda_1)Q^*(\lambda_1))}. \tag{2.90}$$

При получении (2.90) мы учитываем то обстоятельство, что до момента начала обслуживания требования возможно многократное повторение периодов ожидания прихода требования или поломки прибора и периодов восстановления прибора.

Введем в рассмотрение частичные производящие функции

$$\Phi_1(z) = \sum_{m=1}^{j} \varphi_m z^m, \quad \Phi(z) = \sum_{m=j+1}^{\infty} \varphi_m z^m,$$

$$\prod\nolimits_1(z) = \sum_{i=0}^{j} \pi_i z^i, \quad \prod\nolimits_2(z) = \sum_{i=j+1}^{\infty} \pi_i z^i, \quad |z| < 1.$$

Здесь j — порог, задающий стратегию переключения режимов работы рассматриваемой СМО.

Теорема 2.3 *Стационарное распределение π_i, $i \geq 0$, вероятностей состояний рассматриваемой системы за-дается следующим образом:*

$$\prod\nolimits_1(z) = -\pi_0 \sum_{i=0}^{j} z^i \sum_{m=0}^{i} \varphi_m \Delta_{i-m}, \tag{2.91}$$

$$\prod\nolimits_2(z) = \prod\nolimits_1(z) \frac{B_1^*(\lambda_1(1-z)) - z}{z - B_2^*(\lambda_2(1-z))} +$$
$$+ \pi_0 \frac{(\varPhi_1(z) - 1)B_1^*(\lambda_1(1-z)) + \varPhi_2(z) B_2^*(\lambda_2(1-z))}{z - B_2^*(\lambda_2(1-z))}, \tag{2.92}$$

где

$$\pi_0 = (1-\rho_2) \left\{ \varPhi'(1) + (\rho_1 - \rho_2) \left[\varPhi_1(1) - \sum_{i=1}^{j} \sum_{m=0}^{i} \varphi_m \Delta_{i-m} \right] \right\}^{-1}, \tag{2.93}$$

$$\Delta_i = \sum_{l=0}^{i} \alpha_l^{(1)},$$

вероятности φ_m, $m \geq 1$, заданы своей производящей функцией (2.90), а $\varphi_0 = -1$.

Доказательство. Используя формулу полной вероятности, можно получить следующую систему линейных алгебраических уравнений для вероятностей π_l, $l \geq 0$:

$$\pi_l = \pi_0 \sum_{m=1}^{l+1} \varphi_m f_{l-m+1}^{(1)} + \sum_{i=1}^{l+1} \pi_i f_{l-i+1}^{(1)}, \quad l = \overline{0, j-1}, \tag{2.94}$$

$$\pi_l = \pi_0 \left[\sum_{m=1}^{j} \varphi_m f_{l-m+1}^{(1)} + \sum_{m=j+1}^{l+1} \varphi_m f_{l-m+1}^{(2)} \right] +$$
$$+ \sum_{i=1}^{j} \pi_i f_{l-i+1}^{(1)} + \sum_{i=j+1}^{l+1} \pi_i f_{l-i+1}^{(2)}, \quad l \geq j. \tag{2.95}$$

Здесь, как и в п. 2.6, $f_l^{(\nu)}$ — вероятность поступления в систему l требований за время обслуживания одного требования при работе системы в ν-м режиме, $\nu = 1, 2$, $l \geq 0$. Уточним, что в данной модели мы предполагаем, что решение о режиме обслуживания требований выносится в моменты начала обслуживания требований (в надежных системах, рассматриваемых ранее, в отличие от рассматриваемой сейчас, число требований в системе в момент начала обслуживания строго определялось числом требований в системе в момент окончания обслуживания требования).

Решим систему (2.94), (2.95). Используем подмеченную нами в п. 2.6 особенность матрицы переходных вероятностей рассматриваемой системы, в силу которой значения вероятности π_i полностью определялось видом уравнений с номерами от 0 до $i-1$ и не зависило от вида последующих уравнений.

Будем временно считать, что уравнения (2.94) справедливы для всех i, $i \geq 0$. Обозначим $\prod_1(z) = \sum_{l=0}^{\infty} \pi_l z^l$. Умножая уравнения системы (2.94) на соответствующие степени z и суммируя, после некоторых преобразований имеем:

$$\prod(z) = \pi_0 \frac{(\Phi(z) - 1) B_1^*(\lambda_1(1-z))}{z - B_1^*(\lambda_1(1-z))}. \tag{2.96}$$

Разлагая правую часть (2.96) в ряд Маклорена, получаем:

$$\pi_i = -\pi_0 \sum_{m=0}^{i} \varphi_m \Delta_{i-m}. \tag{2.97}$$

Поскольку на самом деле система (2.94) конечна, вероятности π_i задаются формулой (2.97) только для i от 0 до j. Формула (2.91) доказана. Теперь, умножая уравнения (2.95) на соответствующие степени z и суммируя, после некоторых преобразований получаем соотношение

$$\prod\nolimits_1(z) + \prod\nolimits_2(z) = \pi_0 \left\{ \Phi_1(z) \frac{B_1^*(\lambda_1(1-z))}{z} + \Phi_2(z) \frac{B_2^*(\lambda_2(1-z))}{z} \right\} +$$

$$+ \frac{\prod_1(z) - \pi_0}{z} B_1^*(\lambda_1(1-z)) + \frac{\prod_2(z)}{z} B_2^*(\lambda_2(1-z)),$$

откуда и получаем (2.92). Выражение (2.93) для вероятности π_0, как обычно, находим из условия нормировки: $\prod_1(1) + \prod_2(1) = 1$. Теорема доказана.

Предположим снова, что $\lambda_1 = \lambda_2 = \lambda$. Аналогично тому, как это делалось в п. 2.6, можно убедиться, что величины V, $P_l, l = \overline{1,2}$, входящие в функционал (2.56), определяются следующим образом:

$$V = \lambda^{-1} \Bigg\{ \prod\nolimits_1(1) \frac{Y}{(1-\rho_2)^2} + \frac{\rho_1 - \rho_2}{1 - \rho_2} \prod\nolimits_1'(1) + \pi_0 \Big[Y(\Phi_1(1) - 1) - \Phi_1'(1)(1-\rho_2) \times$$

$$\times (\rho_2 - \rho_1) + \Phi''(1) \frac{1-\rho_2}{2} + \Phi'(1) \left(\frac{\lambda^2 b_2}{2} + \rho_2(1-\rho) \right) \Big] \Bigg\} / (1-\rho_2)^2,$$

$$P_1 = \rho_1 \left(\prod\nolimits_1(1) - \pi_0 \Phi_2(1) \right),$$

$$P_2 = \rho_2 \left(\prod\nolimits_2(1) + \pi_0 \Phi_2(1) \right) \tag{2.98}$$

где

$$Y = \frac{\lambda^2}{2} \left(b_2^{(1)}(1-\rho_2) - b_2^{(2)}(1-\rho_1) \right).$$

Подставляя выражения (2.98) в (2.56) и используя утверждение теоремы, и получаем явную зависимость значения функционала качества (2.56) от порога переключения j, что позволяет реализовать на ЭВМ процедуру нахождения оптимального значения порога.

При формулировке постановки задачи мы сделали весьма ограничительное предположение о том, что прибор может выходить из строя только в свободном состоянии. Однако, полученные результаты довольно легко распространяются и на случай, когда прибор может выходить из строя и в занятом состоянии, если время до выхода прибора из строя в занятом состоянии имеет экспоненциальное распределение с параметром α_l при работе системы в данный момент в l–м режиме, $l = 1, 2$. Время восстановления имеет произвольное распределение с преобразованием Лапласа–Стилтьеса $P_l^*(s)$, $l = 1, 2$.

Для такой системы воспользуемся следующим приемом. Переопределим понятие "время обслуживания требования", включив в это время наряду с собственно временем обслуживания времена всех ремонтов прибора за время обслуживания данного требования. Т. е. будем считать, что время обслуживания требования есть время с момента поступления требования на прибор до момента ухода требования из системы. При таком переопределении понятия "время обслуживания" модель с выходом прибора из строя и в занятом состоянии, очевидно, сводится к модели системы с выходом прибора из строя только в свободном состоянии. Для расчета характеристик системы мы должны заменить выражение $B_l^*(s)$ для преобразования Лапласа–Стилтьеса времени обслуживания требования на выражение $\tilde{B}_l^*(s) = B_l^*(s + \alpha_l - \alpha_l P_l^*(s)),\ l = 1, 2$. Вид преобразования $\tilde{B}_l^*(s)$ легко устанавливается с помощью метода введения дополнительного события (для того, чтобы не наступило катастрофы за переопределенное время обслуживания, очевидно, необходимо и достаточно, чтобы за собственно время обслуживания не наступило катастрофы, а также не произошло поломок прибора, во время ликвидации которых произошла бы катастрофа).

Отметим, что если при попадании прибора в свободное состояние прибор немедленно выходит из строя (распределение $D(t)$ вырожденное со значением в нуле), а восстановление его завершается при достижении числом требований в системе уровня $k,\ k \geq 1$, из рассматриваемой нами модели получается модель системы с удаляющимся на отдых прибором (одну из моделей СМО с удаляющимся на отдых прибором мы уже рассмотрели в п. 2.3) и k–стратегией включения прибора.

В этом случае можно убедиться, что явная зависимость значения критерия качества (2.56) от порога j имеет следующий вид:

$$I = I(j) = \frac{\xi_k^{(0)} + \rho_1^{j+1}(\xi_k^{(1)} + \xi_k^{(2)})}{\varepsilon_k^{(0)} + \rho_1^{j+1}\varepsilon_k^{(1)}}, \quad \text{если } j \geq k,\ \rho_1 \neq 1,$$

$$I = I(j) = \frac{\varphi_k^{(0)} + \varphi^{(1)} j + \varphi^{(2)} j^2}{\psi_k^{(0)} + j}, \quad \text{если } j \geq k,\ \rho_1 = 1,$$

$$I = I(j) = \frac{\theta_k^{(0)} + \theta^{(1)} j + \theta^{(2)} j^2 + \rho_1^{j+1}(\theta^{(3)} j + \theta^{(4)})}{æ_k^{(0)} + æ^{(1)} j + æ^{(2)} \rho_1^{j+1}}, \quad \text{если } j < k,\ \rho_1 \neq 1,$$

$$I = I(j) = \frac{\delta_k^{(0)} + \delta^{(1)} j + \delta^{(2)} j^2 + \delta^{(3)} j^3}{\eta_k^{(0)} + \eta^{(1)} j + \eta^{(2)} j^2}, \quad \text{если } j \geq k,\ \rho_1 = 1, \tag{2.99}$$

где величины $\xi^{(l)}, \varepsilon^{(l)}, \varphi^{(l)}, \psi^{(l)}, \theta^{(l)}, æ^{(l)}, \delta^{(l)}, \eta^{(l)}$ —- константы, зависящие от параметров СМО и стоимостных коэффициентов, а также, где это указано, от величины k.

Имея явные зависимости (2.99), можно реализовать эффективные процедуры нахождения оптимального значения порога j^*.

Лабораторная работа: "Исследование и оптимизация функционирования ненадежной двухскоростной системы массового обслуживания типа M|G|1"

Цель работы. Приобретение навыка расчета характеристик и оптимизация функционирования ненадежных СМО.

Задание 1. Реализовать на ЭВМ расчет вероятностей $\varphi_m,\ m \geq 1$, для различных видов распределения времен до выхода прибора и восстановления прибора.

Задание 2. Получить явную зависимость значения функционала (2.56) от порога j на основе использования (2.98) для различных видов распределения времен обслуживания, восстановления и выхода прибора из строя.

Задание 3. Реализовать на ЭВМ процедуру нахождения оптимального значения порога j^* и провести исследование поведения величины выигрыша, даваемого пороговой стратегией по сравнению со стратегией — использовать только один режим, от вида распределений: времени обслуживания, времени до выхода прибора из строя, времени восстановления.

Указание. При фиксированных значениях первых моментов соответствующих распределений исследовать зависимость от коэффициентов вариации этих распределений. При подсчете значения критерия качества при использовании только l–го режима учесть, что $P_l = \rho_l$, а преобразование Лапласа–Стилтьеса распределения времени ожидания в системе имеет вид

$$W^*(s) = \frac{(1-\rho_l)(1-D^*(\lambda_l))}{1-D^*(\lambda_l)(1-\lambda_l q_1)} +$$

$$+ \frac{\lambda_l(1-\rho_l)(1-B_l^*(s)+D^*(\lambda_l)(B_l^*(s)-Q^*(s)))}{(s-\lambda_l(1-B_l^*(s)))(1-D^*(\lambda_l)(1-\lambda_l q_1))}.$$

Задание 4. Получить явный вид коэффициентов в (2.99) и реализовать на ЭВМ процедуру нахождения оптимального значения порога j^*.

Задание 5. Условно считая параметр j непрерывным, получить трансцендентные уравнения для оптимального значения порога j^* и решить их с помощью ЭВМ.

Глава 3

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СМО И СеМО

3.1 Методы моделирования и методы обработки результатов.

Из предыдущих глав видно, что решение математических задач ТМО в предположениях, отличающихся от классических, например, отсутствие марковости, является довольно сложным и приводит либо к громоздким результатам, либо к уравнениям, допускающим только численные решения. Поэтому для получения аналитических результатов приходится прибегать к приближениям, точность которых аналитически оценить не всегда удается. К тому же задачи анализа многих СМО и СеМО, особенно немарковских, являются слишком сложными, чтобы исследовать их лишь аналитическими методами. В этой ситуации используется имитационное моделирование. Процесс имитационного моделирования состоит из трех последовательно выполняемых этапов:

1. Строится математическая модель СМО или СеМО;

2. Функционирование этой модели воспроизводится на ЭВМ при помощи соответствующих алгоритмов, т. е. проводятся имитационные эксперименты;

3. Результаты этих экспериментов обрабатываются статистическими методами, что позволяет сделать необходимые выводы об интересующих характеристиках исследуемого объекта.

Привлечение имитационного моделирования объясняется и тем, что появляется возможность численными экспериментами на ЭВМ подтвердить имеющиеся аналитические выкладки относительно тех или иных характеристик исследуемых СМО или СеМО. Кроме этого, преимущество и достоинство моделирования заключается в том, что оно применимо к системам и сетям любой сложности, где проведение аналитического исследования сопряжено с большими трудностями или вообще невозможно.

Имитационные модели предназначаются для:

— статистического оценивания параметров, характеризующих работу СМО;

— проверки статистических гипотез относительно параметров этих СМО;

— оптимизации режима функционирования СМО;

— проверки точности численного анализа СМО и т. д.

Эти иммитационные модели должны обеспечивать указанные свыше возможности при следующих условиях:

— при различных способах формирования входящего потока требований и при различных видах зависимости во входящем потоке требований;

— при различных дисциплинах обслуживания требований;

— при различных видах зависимости между фрагментами механизма обслуживания требований и т. д.

Для осуществления моделирования необходимы исходные данные, которые подразделяются на:

— исходные данные, позволяющие выбрать конкретную модель СМО;

— исходные данные, задающие характеристики входного потока и механизма обслуживания;

— исходные данные, задающие конкретные значения параметров СМО;

— уровень значимости, который должен быть достигнут при моделировании и т. д.

Так как вопросы построения математических моделей СМО и СеМО рассматривались в гл. 1 и 2, опишем этапы 2 и 3 процесса имитационного моделирования более подробно.

Воспроизведение процесса функционирования системы (сети) МО на ЭВМ фактически состоит в получении вектора состояний системы (сети) в какие-то дискретные моменты времени $\tau_i,\ i \geq 0$. При моделировании как правило задается начальное состояние системы (сети) в момент τ_0, а затем согласно принятой математической модели разыгрываются последующие состояния в моменты $\tau_1, \tau_2, \ldots$. Иными словами, строится траектория вектора состояний системы в интервалах $[\tau_0, \tau_1),\ [\tau_1, \tau_2),\ [\tau_2, \tau_3)$ и т. д. Моменты $\tau_1, \tau_2, \ldots$ могут выбираться двумя способами. Первый из них, называемый *способом "$\Delta\tau$"*, состоит в том, что моменты $\tau_1, \tau_2, \ldots$ являются неслучайными и $\tau_i - \tau_{i-1} = \Delta\tau,\ i \geq 1$, где $\Delta\tau$ — некоторый заранее выбранный интервал, такой, что возможностью более чем одного изменения вектора состояний системы за время $\Delta\tau$ можно пренебречь. Тогда по имеющемуся вектору состояний в момент τ_i случайным образом имитируется возможный переход в другое состояние за время $\Delta\tau$ (отметим, что этого перехода может и не быть), и в результате получается вектор состояний в момент τ_{i+1}.

Описанный способ выбора моментов τ_i обладает как достоинствами, так и недостатками. Достоинства заключаются в простоте реализации процесса моделирования. Недостатки проявляются в следующем. При слишком больших $\Delta\tau$ результаты имитационного моделирования не адекватно аппроксимируют траекторию вектора состояний системы, а при слишком малых $\Delta\tau$ требуются значительные затраты машинного времени, так как на каждое изменение вектора состояний требуется большое число имитаций.

Второй способ выбора моментов τ_i, называемый *способом "особых состояний"*, состоит в том, что значения τ_i являются случайными, моделируются на ЭВМ и при этом обладают тем свойством, что вектор состояний системы изменяется лишь в моменты τ_i, причем изменяется лишь одна компонента этого вектора, а в промежутке (τ_i, τ_{i-1}) изменений не происходит.

Поэтому при таком моделировании системы (сети) достаточно лишь следить за ее особыми состояниями, что позволяет значительно уменьшить затраты машинного времени по сравнению с первым способом выбора моментов τ_i. Алгоритм моделирования способом "особых состояний" отличается от процедуры моделирования способом "$\Delta\tau$" лишь тем, что требуется специальный алгоритм моделирования интервалов $\tau_i - \tau_{i-1}$. К тому же здесь не возникает проблемы выбора "оптимального" значения $\Delta\tau$.

После того как получена реализация вектора состояний системы, необходимо обработать имеющиеся данные методами математической статистики. Проблемы, возникающие на этом этапе, заключаются в следующем.

Во-первых, в большинстве случаев оцениванию подлежат стационарные характеристики систем и, если учитывать данные с момента начала моделирования, то сказывается влияние начального заполнения системы. Поэтому возникает вопрос о длительности "периода стабилизации", т. е. до какого момента данные моделирования не следует учитывать, а с какого — принимать во внимание.

Во-вторых, практически всегда данные являются сильно коррелированными. По таким данным хотя и можно построить точечную оценку интересующего параметра, но не удается построить соответствующий доверительный интервал с заданным уровнем значимости. Это связано с тем, что теория доверительного оценивания использует, как правило, совокупности независимых и одинаково распределенных случайных величин из некоторого распределения вероятностей.

Проиллюстрируем сказанное на следующем примере. Предположим, что исследуется немарковская однолинейная СМО и необходимо оценить W — среднее время ожидания в стационарном режиме. Использование имитационного моделирования в этой задаче является естественным, так как отсутствуют простые в вычислительном отношении аналитические результаты для точного определения необходимой характеристики.

Обозначим ω_k — время ожидания начала обслуживания k–м требованием. Состоятельной оценкой математического ожидания случайной величины является выборочное среднее и в нашем случае в качестве оценки для W будем использовать $(1 / N) \sum_{k=1}^{N} \omega_k$, где N — число требований, сгенерированных входящим потоком за время моделирования. Но выписанная оценка в силу влияния начального заполнения системы будет в общем случае смещенной, так как если, например, $\omega_1 = 0$, то и последующие несколько значений ω_k будут малыми. К тому же, если для какого–то k–го требования время ожидания ω_k велико, то и несколько последующих требований будут иметь большие времена ожидания. Все это говорит о том, что последовательность случайных величин $\{\omega_k,\ k \geq 1\}$ является коррелированной, что создает трудности при обработке результатов наблюдений.

Но необходимо отметить, что ряд процессов, описывающих поведение отдельных характеристик систем (сетей) МО, обладает определенным свойством, а именно свойством регенерации, что дает возможность использовать специальную методику обработки результатов наблюдений над такими процессами, преодолевающую отмеченные трудности. Перед описанием этой методики дадим ряд необходимых определений.

Определение 3.1 *Последовательность случайных величин $\{\xi_n,\ n = 0, 1, 2, \ldots\}$ является процессом восстановления, если $\xi_0 = 0$ и $\{\xi_n - \xi_{n-1},\ n \geq 1\}$ есть независимые, одинаково распределенные неотрицательные величины.*

Определение 3.2 *Последовательность $\{X(n),\ n \geq 1\}$ случайных векторов называется регенерирующим процессом, если существует последовательность $1 \leq \beta_1 < \beta_2 < \beta_3 < \ldots$ случайных дискретных моментов времени, являющихся процессом восстановления, такая, что $\forall\ k > 0$ и для любой другой последовательности моментов времени $0 < n_1 < n_2 < \ldots < n_m$ $(m \geq 1)$ случайные векторы $\{\mathbf{X}(n_1), \mathbf{X}(n_2), \ldots, \mathbf{X}(n-m)\}$ и $\{\mathbf{X}(n_1 + \beta_k), \mathbf{X}(n_2 + \beta_k), \ldots, \mathbf{X}(n_m + \beta_k)\}$ имеют одинаковое распределение, а процессы $\{\mathbf{X}(n),\ n < \beta_k\}$ и $\{\mathbf{X}(p),\ p > \beta_k\}$ являются независимыми. Моменты времени $\{\beta_k,\ k \geq 0\}$ назовем точками (моментами) регенерации процесса $\mathbf{X_n}$, случайную величину $\alpha_k = \beta_{k+1} - \beta_k$ назовем k–м периодом регенерации, а часть процесса $\{\mathbf{X}(n),\ \beta_k \leq n < \beta_{k+1}\}$ будем называть k–м циклом.*

Отметим, что многие процессы, описывающие функционирование СМО и СеМО, являются регенерирующими. Например, в однолинейных стационарных СМО с обслуживанием типа GI регенерирующими являются такие процессы, как число требований в системе в момент времени t, время ожидания n–го требования и виртуальное время ожидания в момент времени t. Для первого процесса точками регенерации являются номера требований с нулевым временем ожидания, а для двух последующих процессов — моменты начала периодов занятости. В таких системах циклы регенерации совпадают с периодами занятости.

В экспоненциальных СеМО регенерирующим является векторный случайный процесс $X(t)$, компоненты которого равны числу требований в узлах обслуживания в момент времени t. Точками регенерации в рассматриваемом случае являются моменты возвращения процесса $X(t)$ в некоторое заранее фиксированное состояние.

Из введенных определений следует, что последовательность периодов регенерации $\{\alpha_k,\ k \geq 1\}$ состоит из независимых и одинаково распределенных случайных величин. Аналогичное утверждение справедливо и для последовательности циклов регенерационного процесса. Если теперь результаты моделирования сгруппировать по циклам, то получившиеся выборки будут независимы и одинаково распределены, что позволяет наряду с оценками средних значений интересующих параметров указать и соответствующие доверительные интервалы. Методика такого построения в общем виде следующая.

Пусть имеется реализация $X(t),\ 1 \le t \le N$, некоторого регенерирующего процесса и по ней необходимо построить точечную оценку и соответствующий доверительный интервал для параметра $\theta = E[X(t)]$, представляющего собой какую—то характеристику системы, функционирующей в стационарном режиме. Считаем, что за время моделирования N сгенерировалось n периодов регенерации, и для простоты предположим, что первый цикл регенерации начался в самом начале моделирования, а n—й цикл закончился в момент N, т. е. $\beta_1 = 1$ и $\beta_n = N$.

Известно, что состоятельной оценкой математического ожидания является выборочное среднее. Поэтому оценку параметра запишем в виде

$$\hat{\theta} = \frac{1}{N}\sum_{t=1}^{N} X(t). \tag{3.1}$$

Обозначим

$$Y_j = \sum_{t=\beta_j}^{\beta_{j+1}-1} X(t).$$

Так как $N = \sum_{i=1}^{n} \alpha_i$ и $\sum_{t=1}^{N} X(t) = \sum_{i=1}^{n} Y_i$, то имеет место представление

$$\hat{\theta} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} \alpha_i} = \frac{\sum_{i=1}^{n} Y_i/n}{\sum_{i=1}^{n} \alpha_i/n} = \frac{\overline{Y}}{\overline{\alpha}} \xrightarrow[n\to\infty]{p} \frac{E[Y_i]}{E[\alpha_i]} \tag{3.2}$$

в силу закона больших чисел при достаточно общих условиях на распределения циклов и периодов регенерации. Обозначим $\sigma^2 = E[Y_j - \theta\alpha_j)^2]$. Так как из определения регенерирующего процесса следует, что последовательность $\{(Y_j, \alpha_j),\ 1 \le j \le n\}$ состоит из независимых и одинаково распределенных случайных векторов, то в силу центральной предельной теоремы при $\sigma^2 < +\infty$ для статистики

$$\frac{\sum_{i=1}^{n} (Y_i - \theta\alpha_i)}{\sigma\sqrt{n}}$$

имеет место асимптотическая при $n \to \infty$ нормальность:

$$\frac{\sum_{i=1}^{n} (Y_i - \theta\alpha_i)}{\sigma\sqrt{n}} \xrightarrow{F} \mathrm{N}(0,1). \tag{3.3}$$

Последнее соотношение предоставляет возможность построить доверительный интервал для параметра θ; необходимо лишь заменить σ ее статистической оценкой. Согласно определению σ^2 получаем

$$\sigma^2 = E\left[(Y_j - \theta\alpha_j)^2\right] = \sigma_Y^2 - 2\theta\sigma_{Y\alpha}^2 + \theta^2\sigma_\alpha^2. \tag{3.4}$$

Оценки теоретических характеристик $\sigma_Y^2,\ \sigma_{Y\alpha}^2,\ \sigma_\alpha^2$ в правой части (3.4) имеют соответствующий вид:

$$\hat{\sigma}_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2, \quad \hat{\sigma}_{Y\alpha}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})(\alpha_i - \overline{\alpha})$$

и

$$\hat{\sigma}_{\alpha}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\alpha_i - \overline{\alpha})^2,$$

и в результате состоятельной оценкой параметра σ^2 является статистика

$$\hat{\sigma}^2 = \hat{\sigma}_Y^2 - 2\hat{\theta}\hat{\sigma}_{Y\alpha} + \hat{\theta}^2\hat{\sigma}_{\alpha}^2.$$

Следовательно, доверительный интервал для параметра θ с уровнем значимости 2γ записывается в виде

$$\left[\hat{\theta} - \frac{z_{1-\gamma}\hat{\sigma}}{\sqrt{n}\overline{\alpha}},\ \hat{\theta} + \frac{z_{1-\gamma}\hat{\sigma}}{\sqrt{n}\overline{\alpha}}\right], \tag{3.5}$$

где $\hat{\theta} = \overline{Y} \,/\, \overline{\alpha}$; $z_{1-\gamma}$ — квантиль уровня γ стандартного нормального распределения; $\hat{\sigma} = \sqrt{\hat{\sigma}_Y^2 - 2\hat{\theta}\hat{\sigma}_{Y\alpha} + \hat{\theta}^2\hat{\sigma}_{\alpha}^2}$.

Так как при построении доверительного интервала используется параметр n — число циклов регенерации, то время моделирования N заранее не фиксируется. Ясно, что $n \approx N \,/\, \overline{\alpha}$, и так как величина $\overline{\alpha}$ также не известна, то для оценивания ее и, следовательно, для оценивания времени моделирования N, можно провести несколько пробных этапов моделирования. Если же при проведении имитационных экспериментов существуют ограничения на затраты машинного времени и в процессе моделирования мы последовательно отслеживаем периоды и циклы регенерации, т.е. n заранее не зафиксировано, то возникает необходимость с ростом n получать последовательность оценок для интересующих нас параметров. Это осуществляется по следующим формулам (аргумент n указывает на тот факт, что оценка получена на основе наблюдений за n циклами регенерации):

$$\overline{Y}(n) = \frac{n-1}{n}\overline{Y}(n-1) + \frac{Y_n}{n}; \quad \overline{Y}(0) = 0;$$

$$\overline{\alpha}(n) = \frac{n-1}{n}\overline{\alpha}(n-1) + \frac{\alpha_n}{n}; \quad \overline{\alpha}(0) = 0;$$

$$\hat{\sigma}_Y^2(n) = \frac{\sigma_{11}(n)}{n-1},$$

где

$$\sigma_{11}(k) = \sigma_{11}(k-1) + \frac{\left[\sum\limits_{j=1}^{k-1} Y_j - (k-1)Y_k\right]^2}{(k-1)k},$$

$$k = 2, 3, \ldots, n, \quad \sigma_{11}(1) = 0;$$

$$\hat{\sigma}_{Y\alpha}(n) = \frac{\sigma_{12}(n)}{n-1}$$

где

$$\sigma_{12}(k) = \sigma_{12}(k-1) + \frac{\left[\sum\limits_{j=1}^{k-1} Y_j - (k-1)Y_k\right]\left[\sum\limits_{j=1}^{k-1} \alpha_j - (k-1)\alpha_k\right]}{(k-1)k},$$

$$k = 2, 3, \ldots, n, \quad \sigma_{12}(1) = 0;$$

$$\hat{\sigma}_{\alpha}^2(n) = \frac{\sigma_{22}(n)}{n-1}$$

где

$$\sigma_{22}(k)=\sigma_{22}(k-1)+\frac{\left[\sum\limits_{j=1}^{k-1}\alpha_j-(k-1)\alpha_k\right]^2}{(k-1)k},$$

$$k=2,3,\ldots,n,\quad \sigma_{22}(1)=0.$$

Кроме того можно отметить, что если имеется возможность выбора из более чем одной последовательности точек регенерации, то длины соответствующих доверительных интервалов будут приблизительно равными при большом времени моделирования N. Отсюда следует, что можно выбирать любую наиболее удобную последовательность точек регенерации.

Но при выборе этих точек необходимо проявлять определенную осторожность. Дело в том, что можно попасть в ситуацию, когда средняя длина периода регенерации бесконечна, или эта средняя длина конечна, но достаточно велика, чтобы при имеющихся возможностях на затраты машинного времени промоделировать необходимое число циклов регенерации.

Предположим, например, что моделируется однолинейная стационарная СМО с целью определения среднего времени ожидания. Пусть ω_n — время ожидания n—го требования. Точками регенерации в рассматриваемом случае являются моменты возвращения процесса $\{\omega_n,\ n\geq 1\}$ в какое—то фиксированное состояние, представляющее собой точку на неотрицательной половине числовой оси. Если это состояние нулевое, то средняя длина периода регенерации конечна. Если же это состояние зафиксировано ненулевым, то средняя длина периода регенерации бесконечна и изложенная ранее методика получения доверительного интервала для W не применима.

В общем случае выход из возникшей ситуации состоит в том, чтобы в качестве моментов регенерации выбрать некоторое множество S, состоящее более чем из одной точки. Это множество должно быть таким, чтобы среднее время возвращения в S было конечно и не велико до такой степени, чтобы в процессе моделирования получилось достаточное число циклов регенерации. Затем в предположении, что процесс регенерируется каждый раз при попадании в S следует применить ранее изложенную методику по расчету точечных оценок и доверительных интервалов для исследуемых параметров.

Описанная модификация носит название *метода приближенной регенерации*. Недостаток этого метода заключается в отсутствии количественного расчета предположения о приближенной регенерации на точность доверительного интервала, достоинство — в простоте численной реализации.

Для регенеративного случайного процесса точечная оценка параметра записывалась в виде (3.2), а соответствующий доверительный интервал имел форму (3.5). Укажем другие точечные оценки для и соответствующие доверительные интервалы.

Оценка Биля:

$$\hat{\theta}_b=(\overline{Y}\,/\,\overline{\alpha})\left((1+\hat{\sigma}^2_Y)\,/\,n\overline{Y}\overline{\alpha}\right)\,/\,\left((1+\hat{\sigma}^2_Y\,/\,n\overline{\alpha}^2\right)\tag{3.6}$$

с доверительным интервалом

$$\left[\hat{\theta}_b-\frac{z_{1-\gamma}\hat{\sigma}_b}{\sqrt{n}\overline{\alpha}},\ \hat{\theta}_b+\frac{z_{1-\gamma}\hat{\sigma}_b}{\sqrt{n}\overline{\alpha}}\right],$$

где $\hat{\sigma}^2_b=\hat{\sigma}^2_Y-2\hat{\theta}_b\hat{\sigma}_{Y\alpha}+\hat{\theta}_b\hat{\sigma}^2_\alpha$.

Оценка Филлера:

$$\hat{\theta}_f = \frac{\overline{Y}\overline{\alpha}n - z^2_{1-\gamma}\hat{\sigma}_{Y\alpha}}{\overline{\alpha}^2 n - z^2_{1-\gamma}\hat{\sigma}^2_\alpha} \tag{3.7}$$

с доверительным интервалом

$$\left[\hat{\theta}_f - \frac{\sqrt{D}}{\overline{\alpha}^2 - \frac{z^2_{1-\gamma}\hat{\sigma}^2_\alpha}{n}},\ \hat{\theta}_f + \frac{\sqrt{D}}{\overline{\alpha}^2 - \frac{z^2_{1-\gamma}\hat{\sigma}^2_\alpha}{n}}\right],$$

где

$$D = \left(\overline{Y}\overline{\alpha} - \frac{z^2_{1-\gamma}\hat{\sigma}_{Y\alpha}}{n}\right) - \left(\overline{\alpha}^2 - \frac{z^2_{1-\gamma}\hat{\sigma}_{Y\alpha}}{n}\right)\left(\overline{Y}^2 - \frac{z^2_{1-\gamma}\hat{\sigma}_{Y\alpha}}{n}\right).$$

Оценка Тина:

$$\hat{\theta}_t = \frac{\overline{Y}}{\overline{\alpha}} + \left(1 + \frac{1}{n}\left(\frac{\hat{\sigma}_{Y\alpha}}{\overline{Y}\overline{\alpha}} - \frac{\hat{\sigma}^2_\alpha}{\hat{\alpha}^2}\right)\right) \tag{3.8}$$

с доверительным интервалом

$$\left[\hat{\theta}_t - \frac{z_{1-\gamma}\hat{\sigma}_t}{\sqrt{n}\hat{\alpha}},\ \hat{\theta}_t + \frac{z_{1-\gamma}\hat{\sigma}_t}{\sqrt{n}\hat{\alpha}}\right],$$

где

$\hat{\sigma}^2_b = \hat{\sigma}^2_Y - 2\hat{\theta}_t\hat{\sigma}_{Y\alpha} + \hat{\theta}_t\hat{\sigma}^2_\alpha.$

Оценка "складного ножа":

$$\hat{\theta}_k = \frac{\overline{Y}}{\overline{\alpha}} - \frac{n-1}{n}\sum_{i=1}^{n}\frac{\sum\limits_{j\neq i} Y_j}{\sum\limits_{j\neq i}\alpha_j} \tag{3.9}$$

с доверительным интервалом

$$\left[\hat{\theta}_k - \frac{z_{1-\gamma}\hat{\sigma}_k}{\sqrt{n}},\ \hat{\theta}_k + \frac{z_{1-\gamma}\hat{\sigma}_t}{\sqrt{n}}\right],$$

где

$$\hat{\sigma}^2_k = \frac{1}{n-1}\left(\frac{\overline{Y}}{\overline{\alpha}} - (n-1)\sum_{i=1}^{n}\frac{\sum\limits_{j\neq i} Y_j}{\sum\limits_{j\neq i}\alpha_j}\right).$$

Оценки(3.6)–(3.9) являются *сильно состоятельными* и более предпочтительными по сравнению с оценкой (3.2), так как имеют меньшее смещение. Для малых выборок наименьшее смещение и более точный доверительный интервал для θ дает оценка "складного ножа" (3.9). Эта оценка наиболее предпочтительна для малых значений n. При больших выборках все приведенные оценки дают примерно одинаковые результаты.

3.2. Моделирование СМО

Для иллюстрации методов, изложенных в п. 3.1, приведем алгоритмы моделирования конкретных СМО и укажем точечные оценки некоторых характеристик этих систем .

Необходимо отметить, что иногда точечные оценки параметров приводятся в виде отношения средних типа (3.2). Для определения доверительных интервалов надо определить соответствующие регенерирующие процессы и обрабатывать получаемые экспериментальные данные регенеративным методом. Для выяснения вопроса о том, является ли конкретный случайный процесс регенерирующим, можно учитывать, например, тот факт, что регенерирующими процессами являются неприводимые и возвратные (с непрерывным или дискретным временем) цепи Маркова и полумарковские процессы с конечным или счетным пространством состояний.

Цель каждой лабораторной работы заключается в реализации на ЭВМ процесса функционирования этой СМО методами иммитационного моделирования и проведении статистического анализа по оцениванию параметров этой СМО регенеративным методом.

Несмотря на то, что СМО, описанные в лабораторных работах (см. далее), хорошо изучены аналитически и не требуют исследования иммитационными методами, все же приведем алгоритмы, описывающие функционирование этих систем. Это делается для обработки и методики моделирования, и способа обработки получающихся экспериментальных данных регенеративным методом. Параметры, характеризующие работу СМО (интенсивности входящего потока и обслуживания, число приборов и т. д.) следует выбирать такими, чтобы у рассматриваемых систем существовал стационарный режим. Число циклов регенерации n необходимо взять таким, чтобы выборочные оценки параметров отличались от теоретических значений этих параметров не более чем на заранее заданную величину.

Лабораторная работа: "Моделирование СМО M|M|1|0“

Цель работы. Разработка алгоритма моделирования СМО $M|M|1|0$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. Пусть λ — интенсивность простейшего потока требований, μ — параметр распределения времени обслуживания. Если в момент прихода требования в систему прибор свободен, то требование немедленно начинает обслуживаться. Если же прибор занят, то требование покидает систему необслуженным. В момент $t = 0$ прибор свободен.

Рассматриваемая СМО описывается случайным процессом $N(t)$ — числом требований в системе или числом занятых приборов в момент времени t. $N(t)$ есть цепь Маркова с множеством состояний $0, 1$ и с вероятностями переходов

$$\begin{aligned}
P[N(t+\Delta t) = 0|N(t) = 0] &= \lambda\Delta_t + o(\Delta_t),\\
P[N(t+\Delta t) = 1|N(t) = 0] &= 1 - \lambda\Delta_t + o(\Delta_t),\\
P[N(t+\Delta t) = 0|N(t) = 1] &= 1 - \mu\Delta_t + o(\Delta_t),\\
P[N(t+\Delta t) = 1|N(t) = 1] &= \mu\Delta_t + o(\Delta_t). \qquad (3.10)
\end{aligned}$$

$N(t)$ — регенерирующий случайный процесс, точками регенерации которого являются моменты начала (или окончания) обслуживания требований.

Стационарные вероятности $p_k = \lim\limits_{t\to\infty} P[N(t) = k],\ k \in \{0,1\}$, в соответствии с (3.10) существуют при всех положительных λ и μ:

$$p_0 = \mu \,/\, (\lambda + \mu), \quad p_1 = \lambda \,/\, (\lambda + \mu) \qquad (3.11)$$

Введем обозначения:

τ_i — момент поступления в СМО i—го требования, $0 = \tau_0 < \tau_1 < \tau_2 < \dots$; $t_i = \tau_i - \tau_{i-1}$ — интервал между моментами поступления в СМО $(i-1)$—го и i—го требования. Так как входящий поток требований является простейшим, то случайные величины t_i независимы и распределены по показательному закону с параметром λ;

x_i — возможная длительность обслуживания i—го требования в предположении, что это требование не было потеряно; $\{x_i\}$ есть последовательность независимых случайных величин, распределенных по показательному закону с параметром μ;

$$\Delta_i = \begin{cases} 1, & \text{если } i\text{— е требование обслужилось;} \\ 0, & \text{если } i\text{— е требование было потеряно.} \end{cases}$$

Реализация $X_i = (t_i, x_i, \Delta_i)$ описывает процесс функционирования рассматриваемой СМО. Значения $\{\Delta_i\}$ получаются при помощи следующего алгоритма:

$\beta_0 \equiv 0; \quad i = 0;$
1. $i = i + 1;$
2. Если $\tau_i \geq \beta_{i-1}$, то $\Delta_i = 1,\ \beta_i = \tau_i + x_i;$
 если $\tau_i < \beta_{i-1}$, то $\Delta_i = 0,\ \beta_i = \beta_{i-1};$
3. Если $i < n$, то переход на 1.

Далее в последовательности Δ_i отбрасываем повторяющиеся значения (первые из каждого повтора оставляем), в результате чего получаем строго возрастающую последовательность $\Delta_0^* = 0 < \Delta_1^* < \Delta_2^* < \ldots < \Delta_n^*$ моментов окончания обслуживания требований и $\{N(\Delta_k^*),\ k \geq 0\}$ есть регенеративный процесс.

Задание 1. Реализовать описанный алгоритм, получить точечные оценки стационарных вероятностей (3.11) и построить соответствующие доверительные интервалы. Исследовать, как эти доверительные интервалы накрывают значения параметров (3.11) в зависимости от уровня значимости γ и числа циклов регенерации n.

Задание 2. Построить оценки Биля, Филлера, Тина и "складного ножа" вместе с соответствующими доверительными интервалами; сравнить полученные результаты с классической оценкой в виде отношения средних (3.2) из задания 1.

Лабораторная работа. "Моделирование СМО M|M|1|0 с ненадежным прибором"

Цель работы. Разработка алгоритма моделирования СМО $M|M|1|0$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. Рассматриваемая СМО отличается от предыдущей системы тем, что работающий прибор может выходить из строя (отказывать); поток отказов прибора — простейший с интенсивностью ν. Восстановление (ремонт) вышедшего из строя прибора начинается сразу же после его отказа; время ремонта — показательное с параметром ε. Требование, которое обслуживалось в момент выхода прибора из строя, покидает СМО необслуженным.

Рассматриваемая система может находиться в одном из трех состояний: E_0 — прибор свободен; E_1 — прибор исправен и обслуживает заявку; E_2 — прибор ремонтируется. Стационарные вероятности этих состояний существуют при всех λ, μ, ν и ε:

$$p_0 = \left[1 + \frac{\lambda}{\mu+\nu} + \frac{\lambda\nu}{\varepsilon(\mu+\nu)}\right]^{-1}, \quad p_1 = \frac{\lambda}{\mu+\nu}p_0, \quad p_2 = \frac{\lambda\nu}{\varepsilon(\mu+\nu)}p_0. \tag{3.12}$$

Стационарная вероятность того, что поступившее в СМО требование будет обслужено,

$$p_{\text{обс}} = \frac{\mu}{\mu+\nu+\lambda(1+\nu\,/\,\varepsilon)}. \tag{3.13}$$

Введем обозначения:

$\{\tau_i\}$, $\{t_i\}$ имеют тот же смысл, что в лабораторной работе по моделированию СМО $M|M|1|0$;

x_i — возможная длительность обслуживания i—го требования в предположении, что за время обслуживания этого требования прибор не выйдет из строя;

α_i — длительность безотказной работы прибора во время обслуживания i—го требования;

ξ_i — возможная длительность ремонта прибора при условии, что он вышел бы из строя во время обслуживания i—го требования; x_i, α_i, ξ_i есть независимые случайные величины, распределенные по показательному закону с параметрами μ, ν и α соответственно;

$$\Delta_i = \begin{cases} 1, & \text{если } i\text{—е требование было обслужино;} \\ 0, & \text{если } i\text{—е требование было потеряно.} \end{cases}$$

Процесс функционирования рассматриваемой СМО может быть описан реализацией $\{\Delta_i, \Delta_i^{(0)}, \Delta_i^{(1)}, \Delta_i^{(2)}\},\ 1 \le i \le N$, где:

$\Delta_i^{(0)}$ — возможный последний промежуток времени, в течение которого прибор находился в свободном исправном состоянии перед приходом в СМО i—го требования; $\Delta_i^{(0)} \equiv 0$, если в момент прихода i-го требования в СМО прибор не находился в состоянии E_0;

$\Delta_i^{(1)}$ — фактическое время обслуживания i-го требования. $\Delta_i^{(1)} \equiv 0$, если i—е требование не попало на обслуживание;

$\Delta_i^{(2)}$ — фактическое время ремонта, связанное с отказом прибора во время обслуживания i-го требования; $\Delta_i^{(2)} \equiv 0$, если i—е требование не попало на обслуживание, либо во время его обслуживания прибор не выходил из строя.

Величины $\{\Delta_i, \Delta_i^{(0)}, \Delta_i^{(1)}, \Delta_i^{(2)}\}$ получаются N—кратной реализацией шагов 2—3 следующего алгоритма.

$\beta_0 \equiv 0$;

1. $i = 1$;

$$\Delta_i^{(0)} = \tau_1; \quad \Delta_i^{(1)} = \min\{x_i, \alpha_i\};$$

$$\beta_1 = \tau_1 + \Delta_i^{(1)} + \Delta_i^{(2)};$$

$$\Delta_i^{(2)} = \begin{cases} 0, & \text{если } \alpha_1 > x_1, \\ \xi_1, & \text{если } \alpha_1 \le x_1; \end{cases}$$

$$\Delta_1 = \begin{cases} 1, & \text{если } \alpha_1 > x_1, \\ 0, & \text{если } \alpha_1 \le x_1; \end{cases}$$

2. $i = i + 1$;

3. Если $\tau_i \ge \beta_{i-1}$, то

$$\Delta_i^{(0)} = \tau_i - \beta_{i-1}, \quad \Delta_i^{(1)} = \min\{x_i, \alpha_i\}$$

$$\Delta_i^{(2)} = \begin{cases} 0, & \text{если } \alpha_i > x_i, \\ \xi_i, & \text{если } \alpha_i \le x_i; \end{cases}$$

$$\Delta_1 = \begin{cases} 1, & \text{если } \alpha_i > x_i, \\ 0, & \text{если } \alpha_i \le x_i; \end{cases}$$

$\beta_i = \tau_i + \Delta_1^{(1)} + \Delta_i^{(2)}$ и переход на 2.

Если $\tau_i < \beta_{i-1}$, то

$\Delta_i \equiv 0,\ \beta_i = \beta_{i-1},\ \Delta_i^{(0)} = \Delta_i^{(1)} = \Delta_i^{(2)} = 0$ и переход на 2.

Точечные оценки стационарных вероятностей p_0, p_1, p_2 могут быть оценены по следующим формулам:

$$\hat{p}_j = \frac{1}{N^*} \sum_i \Delta_i^{(j)}, \quad j = 0, 1, 2, \tag{3.14}$$

где суммирование ведется по тем номерам требований, для которых $\tau_i \geq \beta_{i-1}$, N^* — число таких требований;

$$\hat{p}_{\text{обс}} = \frac{1}{N}\sum_{i=1}^{N}\Delta_i, \tag{3.15}$$

где N — число требований, поступивших в систему в течение периода моделирования.

Задание. Реализовать описанный алгоритм моделирования. Получить точечные оценки стационарных вероятностей как по формулам (3.14)–(3.15), так и по формулам (3.2), (3.6)–(3.9). Сравнить полученные оценки. Учесть, что точками регенерации в рассматриваемой системе являются моменты начала обслуживания. Построить доверительные интервалы. Исследовать, как эти доверительные интервалы накрывают значения вероятностей (3.12)–(3.13) в зависимости от уровня значимости γ и числа циклов регенерации n.

Лабораторная работа. "Моделирование СМО $\mathbf{M|E_r|1|0}$"

Цель работы. Разработка алгоритма моделирования СМО $M|E_r|1|0$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. Рассматриваемая СМО отличается от системы с потерями, описанной в лабораторной работе по моделированию СМО $M|M|1|0$, лишь тем, что время обслуживания требования распределено по закону Эрланга r–го порядка с параметром μ (гамма-распределение с параметрами (r,μ)).

Система может находиться в следующих состояниях: E_0 – прибор свободен, E_1 — прибор занят. Стационарные вероятности состояний существуют при всех положительных λ и μ:

$$p_0 = \frac{1}{1+r\rho}, \quad p_1 = \frac{r\rho}{1+r\rho}, \tag{3.16}$$

где $\rho = \lambda\,/\,\mu$.

Статистический анализ СМО можно проводить двумя способами. 1-й способ использует алгоритм из лабораторной работы по моделированию СМО $M|M|1|0$ с тем изменением, что случайные величины $\{x_i\}$ подчиняются гамма–распределению с параметрами (r,μ).

2-й способ заключается в следующем. Известно, что случайную величину, имеющую распределение Эрланга r–го порядка с параметром μ можно представить в виде суммы r независимых случайных величин, распределенных по показательному закону с параметром μ. Поэтому длительность обслуживания, имеющую распределение Эрланга r–го порядка, можно рассматривать как сумму r последовательных стохастически независимых случайных фаз, каждая из которых имеет показательное распределение.

В связи с этим СМО $M|E_r|1|0$ представляется как многофазная система. В процессе обслуживания требование последовательно проходит r фаз и время обслуживания на каждой фазе имеет показательное распределение с параметром μ. Эта система с потерями, под этим понимается тот факт, что если в момент прихода требования в СМО занята какая-либо (любая!) фаза, то пришедшее требование покидает систему необслуженной.

Состояние такой многофазной СМО описывается цепью Маркова $N(t)$ с множеством состояний $\{0,1,2,\ldots,r\}$. $N(T)=0$, если в момент t в системе нет требований; $N(t)=k$, если в момент t требование обслуживается на k–й фазе, $1\leq k\leq r$. Стационарные вероятности состояний $p_j = \lim\limits_{t\to\infty} P\{N(t)=j\}$ следующие:

$$p_0 = \frac{1}{1+r\rho}, \quad p_1 = p_2 = \ldots = p_r = \rho p_0. \tag{3.17}$$

Введем обозначения:

$\{\tau_i\}, \{t_i\}, \{\Delta_i\}$ имеют тот же смысл, что и в лабораторной работе по моделированию СМО $M|M|1|0$; $x_i^{(j)}$ — длительность обслуживания i—го требования на j—й фазе, $1 \le j \le r$. $\{x_i^{(j)}\}$ независимы и распределены по показательному закону с параметром μ при $i \ge 1$, $1 \le j \le r$; β_i — момент окончания обслуживания i—го требования (если только оно обслуживалось);
Траектория $\{t_i, \Delta_i, \beta_i,\ i \ge 1\}$ описывает процесс функционирования рассматриваемой СМО. Значения β_i получаются при помощи следующего алгоритма.

1. $i = 1, \quad \Delta_1 = 1, \quad \beta_1 = \tau_1 + \sum_{j=1}^{r} x_1^{(j)}$;
2. $i = i + 1$

$$\Delta_i = \begin{cases} 1, & \text{если } \tau_i \ge \beta_{i-1}, \\ 0, & \text{если } \tau_i < \beta_{i-1}; \end{cases}$$

3. Если $\Delta_i = 0$, то переход на 2. Если $\Delta_i = 1$, то $\beta_i = \tau_i + \sum\limits_{j=1}^{r} x_i^{(j)}$ и переход на 2.

Задание 1. Реализовать первый способ моделирования системы. Получить точечные оценки стационарных вероятностей (3.16) по формулам (3.2), (3.6)–(3.9). Сравнить полученные оценки. Построить соответствующие доверительные интервалы. Исследовать, как эти доверительные интервалы накрывают значения вероятностей (3.16) в зависимости от уровня значимости γ и числа циклов регенерации n. Учесть, что точками регенерации являются моменты начала обслуживания требований.

Задание 2. Реализовать второй способ моделирования системы. Получить точечные оценки стационарных вероятностей (3.17) по формулам (3.2), (3.6)–(3.9). Сравнить полученные оценки. Построить соответствующие доверительные интервалы. Исследовать, как эти доверительные интервалы накрывают значения вероятностей (3.17) в зависимости от уровня значимости γ и числа циклов регенерации n. Моменты начала обслуживания требований первой фазой являются точками регенерации.

Лабораторная работа. ”Моделирование СМО M|M|m|0“

Цель работы. Разработка алгоритма моделирования СМО $M|M|m|0$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. На m одинаковых параллельных приборов поступает простейший поток требований интенсивности λ. Если в момент поступления требования имеется хотя бы один свободный прибор,то это требование немедленно начинает обслуживаться. Номер прибора для обслуживания в этой ситуации выбирается случайным образом. Если же все приборы заняты, то требование покидает систему необслуженным. Время обслуживания требования распределено по показательному закону с параметром μ.

Процесс функционирования рассматриваемой СМО может быть описан случайным процессом $N(t)$ — числом занятых приборов в момент времени t. $N(t)$ — это цепь Маркова с множеством состояний $\{0, 1, 2, \ldots, m\}$. При всех положительных значениях λ и μ $N(t)$ имеет стационарное распределение следующего вида:

$$p_j = \lim_{t \to \infty} P[N(t) = j] = \frac{\rho^j}{j!} p_0,$$

$$p_0 = \left(\sum_{i=0}^{n} \frac{\rho^i}{i!} \right)^{-1}, \quad \rho = \frac{\lambda}{\mu}, \quad j = 0, 1, \ldots, m. \tag{3.18}$$

Важнейшими характеристиками системы $M|M|m|0$ являются число занятых приборов и вероятность того, что требование будет потеряно. В соответствии с (3.18) среднее число занятых приборов в стационарном режиме

$$E[N(t)] = \rho - \frac{\rho^{m+1}}{m!} p_0, \tag{3.19}$$

а стационарная вероятность того, что требование будет потеряно

$$p_{\text{отк}} = p_m = \frac{\rho^m}{m!} p_0. \tag{3.20}$$

$N(t)$ есть регенеративный случайный процесс, точками регенерации которого являются моменты начала обслуживания требований каким-либо прибором при условии, что все остальные приборы в этот момент свободны.

Введем обозначения: $\{\tau_i\}, \{t_i\}, \{x_i\}$ имеют тот же смысл, что и в лабораторной работе по моделированию СМО M|M|1|0;

z_i — момент окончания обслуживания i–го требования (если только это требование обслуживалось); $z_i = \qquad = x_i + \tau_i$;

Δ_j — j–й момент изменения состояния рассматриваемой системы, т. е. момент изменения числа работающих приборов;

$S = \{\tau_i\},\ \Delta = \{\Delta_j\}$ — совокупности моментов времени прибытия и ухода из системы обслуженных требований, $\Delta_j \in \{S, \Delta\}$;

m_j — число работающих приборов в момент $(\Delta_j - 0)$.

Введенные случайные величины описывают процесс функционирования рассматриваемой СМО. Наблюдаемая реализация вида $\{\Delta_j, m_j\},\ 1 \le j \le N$, позволяет оценить моменты числа занятых приборов и вероятность потери требований. Алгоритм получения случайных величин Δ_j и m_j следующий.

$\Delta_0 = 0, \quad \Delta_1 = \tau_1, \quad m_1 = 0;$

$\Delta_2 = \min\{\tau_2,\ z_1\}, \quad m_2 = 1;$

$$i > 2, \quad m_i = \begin{cases} \min\{m_{i-1} - 1, m\}, & \text{если } \Delta_{i-1} \in S, \\ \max\{m_{i-1} - 1, 0\}, & \text{если } \Delta_{i-1} \in \Delta; \end{cases}$$

$\Delta_i = \min\{A, \tau_i\}$,

где A есть множество моментов окончания обслуживания требований, которые обслуживались в момент $(\Delta_{i-1} + 0)$; $(j-1)$ есть номер последнего по счету требования, поступившего в систему до момента $(\Delta_{i-1} - 0)$ и принятого на обслуживание.

Оценка среднего числа занятых приборов подсчитывается по формуле

$$E[N(t)] = \frac{1}{\Delta_N} \sum_{i=1}^{N} m_i (\Delta_i - \Delta_{i-1}), \tag{3.21}$$

а оценка вероятности потери требования

$$\hat{p}_{\text{отк}} = \frac{1}{\Delta_N} \sum_{i=1}^{N} \delta_{m_i, m} (\Delta_i - \Delta_{i-1}), \tag{3.22}$$

где $\delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \ne j \end{cases}$ — символ Кронекера.

В соотношениях (3.21) и (3.22) N есть число требований, принятых на обслуживание системой в течение периода моделирования.

Задание. Реализовать алгоритм моделирования системы. Получить точечные оценки для среднего числа занятых приборов и вероятности потери требования как при помощи формул (3.21)–(3.22), так и при помощи формул (3.2), (3.6)–(3.9). Сравнить полученные оценки. Построить доверительные интервалы. Исследовать факт покрытия этими доверительными интервалами характеристик (3.19)–(3.20) в зависимости от уровня значимости γ и числа циклов регенерации n.

Лабораторная работа. "Моделирование СМО $\mathbf{M|M|1|\infty}$"

Цель работы. Разработка алгоритма моделирования СМО $M|M|1|\infty$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. На вход одноприборной СМО с ожиданием поступает простейший поток требований с интенсивностью λ. Если в момент прихода требования в систему прибор свободен, то требование немедленно начинает обслуживаться. Если же прибор занят, то требование становится в очередь и начинает обслуживаться после того, когда закончится обслуживание всех требований, стоявших в очереди перед ним. Выбор требования из очереди на обслуживание осуществляется в порядке поступления требований в СМО по принципу "первым пришел — первым обслужился". Время обслуживания каждого требования распределено по показательному закону с параметром μ. Число мест в очереди для ожидания неограниченное.

Процесс функционирования рассматриваемой системы может быть описан случайным процессом $N(t)$ — числом требований в СМО в момент времени t. $N(t)$ — цепь Маркова с множеством состояний $\{0, 1, 2, \ldots\}$. Если $\rho = \lambda \,/\, \mu < 1$, то указанный процесс имеет стационарное распределение вида

$$p_j = \lim_{t\to\infty} P[N(t) = j] = \rho^j(1-\rho), \quad j = 0, 1, 2, \ldots \tag{3.23}$$

Кроме распределения (3.23) числа требований в системе важнейшими характеристиками являются средняя длина очереди и среднее время ожидания требований в системе. Отметим, что случайные процессы $N(t)$ — число требований в системе в момент времени t и ω_n — время ожидания n–м требованием являются регенеративными. Для первого процесса точками регенерации являются моменты начала периодов занятости прибора, а для второго — номера требований с нулевым временем ожидания.

Введем обозначения:

τ_i — момент поступления в СМО i–го требования, $1 \le i \le N$; $0 = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_N$; $S = \{\tau_1, \tau_2, \ldots, \tau_N\}$;

$t_i = \tau_i - \tau_{i-1}$ — промежуток времени между моментами поступления в СМО $(i-1)$–го и i–го требования. Из определения простейшего потока следует, что случайные величины $\{t_i,\ 1 \le i \le N\}$ независимы и распределены по показательному закону с параметром λ;

x_i — длительность обслуживания i-го требования; $\{x_i\}$ независимы и распределены по показательному закону с параметром μ;

z_i — момент окончания обслуживания i–го требования, $Z = \qquad = \{z_1, \ldots, z_N\}$;

$\omega_i,\ s_i$ — время ожидания и время пребывания соответственно i–го требования в СМО;

ξ_j — j–й момент изменения состояния рассматриваемой системы. Это либо момент прихода требования в СМО, либо момент окончания обслуживания требования (момент поступления следующего требования на прибор); $\xi_i \in \{Z, S\}$;

m_j — число требований в очереди в момент $(\xi_j + 0)$.

Реализация функционирования системы $M|M|1|\infty$ описывается введенными случайными величинами. Они определяются по следующему алгоритму.

Если $i = 1$, то $\omega_1 = 0, \quad z_1 = \tau_1 + x_1, \quad s_1 = z_1$;

При $2 \le i \le N, \quad \omega_i = \max\{0, v_{i-1} - s_i\}, \quad T_i = \omega_i + \tau_i, \quad v_i = s_i + T_i$;

$m_1 = 0$;

$$m_j = \begin{cases} \max\{0, m_{j-1} - 1\}, & \text{если } \xi_j \in Z; \\ m_{j-1} + 1, & \text{если } \xi_j \in S \text{ и в момент } (\xi_j - 0) \\ & \text{прибор был занят;} \\ 0, & \text{если } \xi_j \in S \text{ и в момент } (\xi_j - 0) \\ & \text{прибор был свободен.} \end{cases}$$

Задание. Реализовать алгоритм моделирования системы. Используя регенеративный метод анализа данных, найти оценки стационарного распределения числа требований в системе, средней длины очереди и среднего времени ожидания по формулам (3.2), (3.6)–(3.9). Сравнить полученные оценки с соответствующими теоретическими значениями. Построить доверительные интервалы. Исследовать, как эти доверительные интервалы накрывают теоретические значения в зависимости от уровня значимости γ и числа циклов регенерации n.

Лабораторная работа. "Моделирование СМО $\mathbf{M|M|m|\infty}$"

Цель работы. Разработка алгоритма моделирования СМО $M|M|m|\infty$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. На m одинаковых приборов поступает простейший поток требований интенсивности λ. Если в момент поступления требования имеется хотя бы один свободный прибор, то требование немедленно начинает обслуживаться. Если же все приборы заняты, то прибывшее требование становится в очередь за всеми теми требованиями, которые поступили раньше и еще не начали обслуживаться. Освободившийся прибор немедленно приступает к обслуживанию очередного требования, если только имеется очередь. Каждое требование обслуживается только одним прибором, и каждый прибор обслуживает в каждый момент не более одного требования. Время обслуживания требований распределено по показательному закону с параметром μ. Число мест в очереди для ожидания неограниченное.

Введем случайный процесс $N(t)$ — число требований в системе в момент времени t. $N(t)$ есть цепь Маркова с множеством состояний $\{0, 1, 2, \ldots\}$. $N(t) = 0$ означает, что СМО свободна; $N(t) = j$ при $1 \leq j < m$ означает, что j приборов заняты, $(m - j)$ приборов свободны и очереди нет; $N(t) = j$ при $j \geq m$ означает, что все m приборов заняты, $(j - m)$ требований находятся в очереди.

При $\rho = \lambda \,/\, \mu < m$ существуют стационарные вероятности

$$p_j = \lim_{t\to\infty} P\{N(t) = j\};$$

$$p_0 = \left(\sum_{i=0}^{m} \frac{\rho^i}{i!} + \frac{\rho^{m+1}}{m!(m-\rho)} \right)^{-1},$$

$$p_j = \frac{\rho^j}{j!} p_0,\ 1 \leq j \leq m; \quad p_j = \frac{\rho^j}{m^{j-m} m!} p_o,\ j \geq m+1.$$

Важнейшими характеристиками рассматриваемой СМО являются длина очереди, время ожидания, время пребывания.

Средняя длина очереди в стационарном режиме

$$\overline{N} = \frac{\rho^{m+1} p - o}{(m-1)!(m-\rho)^2}.$$

Среднее время ожидания в стационарном режиме

$$W = \frac{\rho^m p - o}{\mu(m-1)!(m-\rho)^2}.$$

Среднее время пребывания требований в СМО в стационарном режиме

$$T = \left(\overline{N} + \rho\right) \,/\, \lambda.$$

Введем обозначения:
$\tau_i, t_i, x_i, z_i, w_i, s_i, \xi_i, m_i$ — имеют тот же смысл, что и в лабораторной работе по моделированию СМО $M|M|1|\infty$;

$\{z_i\}, \{w_i\}, \{s_i\}, \{\xi_j\}, \{m_j\}$ получаются по следующему алгоритму.

Если $1 \leq i \leq m$, то $w_i = 0, \quad s_i = x_i, \quad z_i = x_i + \tau_i$,

Если $i > m$, то $w_i = \max\{0, z_{(i-m)} - \tau_i\}, \quad s_i = w_i + x_i, \quad z_i = \tau_i + s_i$, где $z_{(i-m)}$ — есть $(i - m)$−я порядковая статистика, построенная по выборке $z_1, z_2, \ldots, z_{i-1}$;

$m_1 = 0$,

$$m_j = \begin{cases} \max\{0, m_{j-1} - 1\}, & \text{если } \xi_j \in Z; \\ m_{j-1} + 1, & \text{если } \xi_j \in S \text{ и в момент } (\xi_j - 0) \\ & \text{все } m \text{ приборов были заняты;} \\ 0, & \text{если } \xi_j \in S \text{ и в момент } (\xi_j - 0) \\ & \text{имелся хотя—бы один свободный прибор.} \end{cases}$$

Задание. То же, что и в лабораторной работе 5. Оцениванию подлежат стационарное распределение числа требований в системе, средняя длина очереди, среднее время ожидания и среднее время пребывания требований в системе.

Лабораторная работа. "Моделирование СМО M|M|∞"

Цель работы. Разработка алгоритма моделирования СМО $M|M|\infty$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. На вход системы с бесконечным числом идентичных приборов поступает простейший поток требований с интенсивностью λ. Требование обслуживается за некоторое случайное время, распределенное по показательному закону с параметром μ.

Хотя систем с бесконечным числом приборов не существует, тем не менее, если приборов очень много, то результаты для $M|M|\infty$ будут давать в этой ситуации хорошие приближения. Состояния рассматриваемой СМО описывается процессом $N(t)$ — числом занятых приборов в момент времени t. $N(t)$ есть цепь Маркова с множеством состояний $\{0, 1, 2, \ldots\}$ и со стационарными вероятностями

$$p_j = \lim_{t\to\infty} P\{N(t) = j\} = \frac{\rho^j}{j!} e^{-\rho}, \quad \rho = \frac{\lambda}{\mu},\ j = 0, 1, 2, \ldots \tag{3.24}$$

В соответствии с (3.24) среднее число занятых приборов в стационарном режиме равно ρ.

$N(t)$ — регенеративный случайный процесс, точками регенерации которого являются моменты начала обслуживания требований каким-либо прибором при условии, что остальные приборы в этот момент свободны.

Введем обозначения:

$\{z_i\}, \{t_i\}, \{x_i\}, \{\xi_i\}, \{\tau_i\}$ имеют тот же смысл, что и в лабораторной работе по моделированию СМО $M|M|1|\infty$;

m_j — число работающих приборов в момент $(\xi_j + 0)$.

Реализация $\{\xi_j, m_j\}$ определяет процесс функционирования рассматриваемой СМО. Для ее получения моделируем N случайных величин τ_i и N случайных величин $z_i = x_i + \tau_i$. По реализациям $\{\tau_i,\ 1 \le i \le N\}$ и $\{z_i,\ 1 \le i \le N\}$ образуем совместный вариационный ряд и $\xi_j,\ 1 \le j \le 2N$, — есть j—я порядковая статистика этого ряда. Тогда $m_1 = 1$,

$$m_{j+1} = \begin{cases} m_j + 1, & \text{если } \xi_{j+1} \in S, \\ \max\{0, m_{j-1}\}, & \text{если } \xi_{j=1} \in Z,\ j = 1, 2, \ldots \end{cases}.$$

Стационарные вероятности p_j оцениваются по формулам

$$p_j = \frac{1}{\xi_N} \sum_{i=1}^{N} \delta_{j,m_i}(\xi_i - \xi_{i-1}), \quad \xi_0 \equiv 0. \tag{3.25}$$

Оценка среднего числа занятых приборов в стационарном режиме

$$\hat{m} = \frac{1}{\xi_N} \sum_{i=1}^{N} m_i(\xi_i - \xi_{i-1}). \tag{3.26}$$

Задание. То же, что и в лабораторной работе по моделированию СМО $M|M|1|\infty$. Оцениванию подлежат распределение и среднее число занятых приборов.

Лабораторная работа. "Моделирование СМО M|G|1|∞ с зависимыми временами обслуживания"

Цель работы. Разработка алгоритма моделирования СМО $M|G|1|\infty$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. На вход одноканальной СМО с неограниченной очередью поступает простейший поток требований интенсивности λ. Время обслуживания требования является случайным и определяется по правилу

$$x_1 = \theta_1,$$

$$x_{n+1} = \begin{cases} x_n \text{ с вероятностью } p, \\ \theta_{n+1} \text{ с вероятностью } (1\text{-}p). \end{cases} \tag{3.27}$$

Здесь n — номер обслуживаемого требования, x_n — время обслуживания n—го требования, $\{\theta_n\}$ — последовательность независимых случайных величин с плотностью распределения $f(x)$.

Трудность аналитического исследования рассматриваемой СМО заключается в нерекуррентности обслуживания требований. Обозначим через q_n — число требований, оставшихся в системе после ухода n—го требования. Тогда последовательность $(q_n, x_{n+1}),\ n = 1, 2, \ldots$ образует марковский процесс, и стационарные вероятности

$$P[i,y] = \lim_{n\to\infty} P_n[i,y] = \lim_{n\to\infty} \frac{d}{dy} P[q_n = i, x_{n+1} \le y] = \frac{d}{dy} P[q = i, x \le y],$$

если они существуют, удовлетворяют следующей системе уравнений:

$$P[i,y] = p\left\{P[0,y]\frac{(\lambda y)^i}{i!}\mathrm{e}^{-\lambda y} + \sum_{l=1}^{i+1} P[l,y]\frac{(\lambda y)^{i-l+1}}{(i-l+1)!}\mathrm{e}^{-\lambda y}\right\} +$$
$$+(i-p)f(y)P(i),$$

где

$$P(i) = P[q=i] = \int_0^\infty P[i,x]\,dx = \int_0^\infty P[0,y]\frac{(\lambda y)^i}{i!}\mathrm{e}^{-\lambda y}\,dy +$$
$$+\sum_{l=1}^{i+1}\int_0^\infty P[l,y]\frac{(\lambda y)^{i-l+1}}{(i-l+1)!}\mathrm{e}^{-\lambda y}\,dy.$$

При $\lambda E[x] < 1$ существует стационарный режим системы и средняя длина очереди подсчитывается по формуле

$$E[q] = (1-\lambda E[x])^{-1}\left[\frac{1+p}{2(1-p)}E[x^2] - \frac{p}{1-p}\lambda E[x] + \frac{\lambda}{1-p}\int_0^\infty yP[0,y]\,dy\right]. \tag{3.28}$$

Правая часть формулы (3.28) может быть вычислена, если известна функция $P[0,y]$, а нахождение этой функции представляет собой трудноразрешимую задачу.

Можно также показать, что в стационарном режиме для средней длины очереди имеют место неравенства:

$$\frac{\lambda^2(1-p)E[x^2] - 2\lambda E[x] + 2(1-p)(\lambda E[x])^2 + 2p\lambda^2 E[x^2]}{2(1-\lambda E[x])(1-p)} \le E[q] \le$$
$$\le \frac{\lambda^2 E[x^2](1-p) + 2(1-p)(\lambda E[x])^2}{2(1-\lambda E[x])(1-p)}. \tag{3.29}$$

Кроме того, можно получить приближенное выражение для $E[q]$:

$$E[q] \approx \frac{\lambda^2(1+p)E[x^2] - 2p(\lambda E[x])^2}{2(1-p)(1-\lambda E[x])}. \tag{3.30}$$

Дальнейшее аналитическое исследование системы не приводит к успеху, в связи с чем приходится прибегать к имитационному моделированию с целью как статистического оценивания параметров, характеризующих работу СМО, так и для проверки численного анализа этой СМО. Алгоритм моделирования системы такой же, как и в лабораторной работе по моделированию СМО $M|M|1|\infty$, с тем лишь изменением, что случайные величины $\{x_i\}$ моделируются по правилу (3.27).

Задание. Реализовать алгоритм моделирования системы. Найти оценки среднего времени ожидания и средней длины очереди требований по формулам (3.2), (3.6)–(3.9). Проверить, удовлетворяют ли оценки для средней длины очереди соотношениям (3.29)–(3.30). Исследовать зависимость оценок от вероятности p. Особо рассмотреть предельные случаи $f(x) = \quad = \mu e^{-\mu x}$, $p \downarrow 0$ и $p \uparrow 1$, т. е. варианты системы $M|M|1|\infty$ и $M|D|1|\infty$. Построить доверительные интервалы. Для применения регенеративного метода анализа данных моделирования необходимо учесть, что случайный процесс $N(t)$ — число требований в системе в момент t будет регенеративным, если в качестве точек регенерации брать моменты начала тех периодов занятости прибора, при которых время обслуживания требования, пришедшего в пустую систему, не зависит от времени обслуживания последнего требования предыдущего периода занятости.

Лабораторная работа. "Моделирование СМО G|M|1|∞ с зависимостью во входящем потоке“

Цель работы. Разработка алгоритма моделирования СМО $G|M|1|\infty$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. Имеется одноканальная СМО с ожиданием. Интервалы входящего потока (номер интервала совпадает с номером требования, поступившего в момент окончания интервала) определяются по правилу

$$t_1 = \theta_1,$$

$$t_{n+1} = \begin{cases} t_n \text{ с вероятностью } p; \\ \theta_{N+1} \text{ с вероятностью } (1-p). \end{cases} \tag{3.31}$$

Здесь n — номер интервала; $\{\theta_n\}$ — последовательность независимых случайных величин с плотностью распределения $f(x)$.

Время обслуживания требований распределено по показательному закону с параметром μ.

Нерекуррентность входящего потока требований создает существенные математические трудности для аналитического решения. Приходится исследовать вероятностно–временные характеристики СМО при имеющийся зависимости во входящем потоке требований методами имитационного моделирования. Алгоритм моделирования такой же, как и для системы $M|M|1|\infty$, с тем изменением, что случайные величины $\{t_i\}$ моделируются по правилу (3.31). По аналогии с системой с зависимыми временами обслуживания точками регенерации процесса $N(t)$ берутся моменты начала тех периодов занятости, в которых интервал, связанный с поступлением требования в пустую систему, не зависит от интервала, связанного с поступлением в систему последнего требования предыдущего периода занятости.

Задание. Реализовать алгоритм моделирования системы. Учесть, что стационарный режим существует при $E[t_1]\mu > 1$. Найти оценки среднего времени ожидания и средней длины очереди по формулам (3.2), (3.6)–(3.9). Исследовать зависимости оценок от вероятности p. Особо рассмотреть предельные случаи $f(x) = \lambda e^{-\lambda x}$, $p \downarrow 0$ и $p \uparrow 1$, т. е. варианты систем $M|M|1|\infty$ и $D|M|1|\infty$.

Построить доверительные интервалы.

Лабораторная работа. "Моделирование СМО $G|G|1|\infty$, управляемой цепью Маркова"

Цель работы. Разработка алгоритма моделирования СМО $G|G|1|\infty$, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. Рассматривается однолинейная СМО с ожиданием, входящий поток требований которой является рекуррентным с заданным распределением интервалов времени между поступлениями требований. Последовательные временные интервалы обслуживания управляются марковской цепью $\{\eta_k\}$ следующим образом. Эта марковская цепь с n состояниями может изменять свои состояния в моменты окончания обслуживания в соответствии с матрицей переходных вероятностей $P = \|p_{ij}\|$, $1 \leq i, j \leq n$. Если в течение времени обслуживания k−го требования цепь Маркова находилась в состоянии j, т. е. $\eta_k = j$, то время обслуживания x_k имеет условную функцию распределения

$$P[x_k \leq y] \,|\, \eta_k = j] = B_j(y), \quad 1 \leq j \leq n. \tag{3.32}$$

До настоящего времени не существует расчетных формул для моментов вероятностно-временных характеристик в системах $G|G|1|\infty$ при рекуррентном обслуживании, в частности, для средней длины очереди и среднего времени ожидания. Поэтому таких формул нет и для случая зависимостей между временами обслуживания. Приближенные методы расчета основаны на аппроксимации распределения интервалов между прибытиями требований и времен обслуживания распределениями фазового типа (H_R−распределениями),т. е. в качестве аппроксимации рассматриваемой СМО исследуется система $H_R|H_R|1|\infty$, такая, у которой распределения интервалов прибытия и времен обслуживания имеют такие же первые моменты, какими являются соответствующие моменты у СМО $G|G|1|\infty$, управляемой цепью Маркова. Необходимые процедуры расчета характеристик рассматриваемой СМО изложены в гл. 1.

К сожалению, аналитические результаты, позволяющие вычислять характеристики СМО с распределениями, не попадающими в класс H_R−распределений, не получены. Поэтому численного сравнения точности произвести не удается, тем более, что исследуется СМО с нерекуррентным обслуживанием вида (3.32). По−видимому исследование рассматриваемой системы следует искать на пути имитационного моделирования с целью определения на требуемом доверительном уровне как характеристик этой СМО, так и установления точности способа определения характеристик системы аппроксимацией H_R−распределениями. Алгоритм моделирования такой же, как и для СМО $M|M|1|\infty$, но случайные величины $\{t_i\}$ и $\{x_i\}$ генерируются в соответствии со своими законами распределения.

Задание. Реализовать алгоритм моделирования системы. Найти оценки среднего времени ожидания, средней длины очереди. Построить доверительные интервалы для этих оценок. Сравнить построенные оценки с результатами, полученными при помощи аппроксимации H_R−распределениями.

Лабораторная работа. "Оптимальное управление СМО с управляемым режимом функционирования"

Цель работы. Разработка алгоритма моделирования СМО, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. Рассматривается СМО, описанная в п. 2.6. Требуется указать стратегию назначения скорости обслуживания требований, минимизирующую функционал качества (2.56).

Алгоритм решения задачи следующий. Фиксируется набор пороговых значений j_ν, $1 \leq \nu \leq leN - 1$, и реализуется алгоритм моделирования СМО аналогично тому, как это делалось в лабораторной работе по моделированию СМО $M|M|1|\infty$ (для однолинейной СМО) или в лабораторной работе по моделированию СМО $M|M|m|\infty$ (для многолинейных СМО). Изменения состоят в том, что случайные величины $\{t_i\}$ и $\{x_i\}$ моделируются в соответствии со своими законами распределения и в соответствии с зафиксированной стратегией. В результате моделирования подсчитываются оценки среднего времени пребывания требований V и оценки вероятностей P_r, как доли времени, в течение которого используется r–я скорость, $1 \leq r \leq N$. Таким, образом мы получаем оценку значения функционала при заданном наборе пороговых значений j_ν, $1 \leq \nu \leq N-1$. Перебирая возможные наборы этих пороговых значений, мы получаем последовательность оценок функционалов I и в качестве оптимальной стратегии выбирается тот набор пороговых значений, который доставляет функционалу из этой последовательности наименьшее значение.

Задание 1. Смоделировать однолинейную СМО с пуассоновским входящим потоком и с изменяемыми интенсивностями экспоненциального времени обслуживания. Определить оценки стационарных вероятностей состояний этой системы, среднего времени пребывания, оценки вероятностей работы прибора с i–й скоростью и построить для них доверительные интервалы. Для случая $N=2$ построить графики зависимостей указанных параметров для единственного в этом случае порогового значения j. Определить оптимальную стратегию. Результаты моделирования сравнить с аналитическими результатами из 2.6.

Задание 2. Аналогично заданию 1, но интервалы входящего потока требований и длительности обслуживания являются случайными величинами фазового типа (распределение Эрланга, H_R–распределение). Сравнить полученные результаты с результатами задания 1 в случае, когда интенсивности λ_r и функционал распределения времени обслуживания $B_r(t)$ в каждом из заданий совпадают.

Лабораторная работа. "Анализ корреляционных свойств выходящих потоков СМО методом имитационного моделирования“

Цель работы. Разработка алгоритма моделирования СМО, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. Анализ корреляционных свойств выходящих потоков СМО представляет интерес по следующей причине. Как правило СМО являются элементами СеМО и совокупность выходящих потоков СМО формирует входящие потоки для систем этой сети. Поэтому если выходящие потоки коррелированы, т. е. нерекуррентны, то и входящие потоки будут обладать аналогичным свойством и влиять на вероятностно–временные характеристики СМО, что демонстрируется результатами лабораторной работы по моделированию СМО $G|M|1|\infty$ с зависимостью во входящем потоке.

В качестве показателя корреляционных свойств потока возьмем значение коэффициента корреляции $R(\xi,\eta)$ между соседними интервалами в потоке требований. Известно, что коэффициент корреляции есть мера зависимости случайных величин. Для любых ξ и η имеем $|R(\xi,\eta)| \leq 1$, и чем больше значение $R(\xi,\eta)$ по абсолютной величине, тем сильнее зависимость между случайными величинами (в нашем случае между соседними интервалами потока). И наоборот: чем меньше $R(\xi,\eta)$ по модулю, тем зависимость меньше и соответственно меньше влияние нерекуррентности потоков.

Имеющиеся аналитические результаты по корреляционным свойствам выходящих потоков относятся в основном к системам с простейшим входящим потоком требований. Получена производящая функция ковариации между интервалами $\xi_0, \xi_1, \ldots, \xi_m, \ldots$ стационарного потока, выходящего из СМО $M|G|1$:

$$f(s) = \sum_{m=1}^{\infty} \operatorname{cov}(\xi_0, \xi_m) s^m = \frac{1-\rho}{\lambda^2(1-s)} \left(\frac{v-s}{1-v} + \frac{sv'-v}{vv'} \right),$$

где $\operatorname{cov}(\xi_0, \xi_m)$ — есть ковариация между интервалами ξ_0, ξ_m; $|s| < 1$; λ —интенсивность входящего потока требований; ρ — коэффициент загрузки системы; $v = v(s)$ — наименьший по модулю корень уравнения $v = s\beta(\lambda(1-v))$, где $\beta(\ )$ — преобразование Лапласа–Стилтьеса функции распределения времени обслуживания в СМО; $v' = dv\,/\,ds$.

У СМО с простейшим входящим потоком требований и гиперэкспоненциальной функцией распределения времени обслуживания

$$B(t) = 1 - \sum_{i=1}^{k} p_i \mathrm{e}^{-\mu_i t},\ p_i \geq 0, \quad \sum_{i=1}^{k} p_i = 1,$$

выходящий поток имеет отрицательный коэффициент корреляции и принимает максимальное по модулю значение при $k = 2,\ \rho_1 = \lambda / \mu_1 = 1,\ \rho_2 = \lambda / \mu_2 << 1,\ p_1 = 2\,\mathrm{cov}\,((4/9)\pi) = 0,347$. Отметим, что в системах $M|M|1$ и $M|D|1|1$ выходящий поток является рекуррентным, а выходящие потоки всех остальных СМО вида $M|G|1$ свойством рекуррентности не обладают.

Цель настоящей лабораторной работы заключается в проверке корреляционных свойств выходящих потоков различных СМО. Необходимо смоделировать СМО, получить для нее последовательность $\xi_0, \xi_1, \ldots$ интервалов выходящего потока и проверить, являются ли случайные величины этой последовательности независимыми и одинаково распределенными. Проверку осуществлять статистическими критериями на требуемом доверительном уровне.

Задание 1. Смоделировать СМО $M|M|1$ и $M|D|1|1$ и убедиться в том, что выходящие потоки этих систем являются рекуррентными.

Задание 2. Смоделировать СМО $M|H_k|1$ при нескольких значениях k, подсчитать значение коэффициентов корреляции выходящих потоков и найти параметры СМО, при которых коэффициент корреляции максимален.

Задание 3. Имитационными методами исследовать корреляционные свойства выходящих потоков систем с зависимыми временами обслуживания, с зависимостями во входящем потоке, с распределением времени обслуживания, управляемом цепью Маркова из лабораторных работ по моделированию СМО $M|M|\infty$, СМО $M|G|1|\infty$, и СМО $G|M|1|\infty$ с зависимостью во входящем потоке соответственно. Выявить влияние параметров этих систем на значение коэффициента корреляции.

Задание 4. Исследовать корреляционные свойства выходящего потока требований системы $M|M|m$ при $m > 1$. Выяснить влияние числа приборов на значение коэффициента корреляции интервалов выходящего потока.

Лабораторная работа. "Моделирование СМО с повторными требованиями"

Цель работы. Разработка алгоритма моделирования СМО, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. Рассмотрим однолинейную СМО с повторными требованиями, в которую поступает рекуррентный поток требований с функцией распределения $A(t)$. При каждом поступлении с вероятностью p_k поступает $k \geq 1$ требований. Назовем их первичными. Групповое поступление рассматривается в связи с тем, что в реальности возникающее требование для передачи представляет собой совокупность пакетов данных. Если при поступлении в систему группы первичных требований прибор оказался свободным, то одно требование поступает на обслуживание, а остальные требования становятся повторными. Каждое повторное требование возобновляет попытку занять прибор через независимые и одинаково распределенные промежутки времени с функцией распределения $D(t)$, и его обслуживание ничем не отличается от обслуживания первичного требования.

Обслуживание на приборе состоит из двух типов. Длительность первого этапа (длительность конфликтного промежутка времени) и длительность второго этапа (времени передачи требования) имеют функции распределения $B_1(t)$ и $B_2(t)$ соответственно. Если на первом этапе обслуживания на прибор поступило какое—то требование (первичное или повторное), то находящееся на приборе и поступившие требования мгновенно становится повторными, а прибор освобождается. Если по истечении первого этапа обслуживания конфликт не наступил, то требование поступает на второй этап обслуживания и прибор становится недостижимым для вновь поступающих требований. Поступающие в систему на втором этапе обслуживания первичные требования становятся повторными, а при поступлении повторных требований состояние системы не меняется. После прохождения второго этапа требование покидает систему. Число повторных требований не ограничено. Каждое требование неоднократно поступает в систему до тех пор, пока не обслужится.

Математическое исследование рассматриваемой системы в случае пуассоновского входящего потока требований, экспоненциальных интервалов возобновления попыток попасть на обслуживание и экспоненциальной длительности первого этапа показывает, что условия

$$\nu > 0, \quad æ = 2\lambda_1 / \mu + \lambda_1\beta < 1,$$

являются необходимыми и достаточными для существования стационарного режима в рассматриваемой СМО. Получены следующие характеристики системы:

1) вероятность того, что прибор свободен

$$p_0 = 1 - \lambda_1\mu^{-1} - \lambda_1\beta;$$

2) вероятность того, что прибор занят и требование находится на первом этапе обслуживания

$$p_1 = \lambda_1\mu^{-1};$$

3) вероятность того, что прибор занят и требование находится на втором этапе обслуживания

$$p_2 = \lambda_1\beta;$$

4) среднее число повторных требований в системе

$$\overline{n} = [(\lambda_1 - \lambda)\mu + æ(\lambda\mu + \nu\lambda_1) + \frac{1}{2}\mu\nu(\lambda_1^2\beta_1 + æc^{-1}c_1)](\nu\mu - \nu\mu æ^{-1});$$

5) среднее время ожидания для случая одиночного поступления требований

$$\overline{w} = \overline{n}\lambda_1^{-1};$$

6) приближенная формула для вероятности бесконфликтного обслуживания первичного требования

$$q \approx p_0 \sim \mu(\mu + \overline{n}\nu + \lambda)^{-1}$$

В приведенных формулах

ν — параметр экспоненциального распределения интервалов возобновления попыток попасть на обслуживание повторными требованиями;

μ — параметр экспоненциального распределения длительности первого этапа обслуживания;

$\lambda_1 = c\lambda$;

β — среднее время длительности второго этапа обслуживания;

λ — параметр входящего пуассоновского потока требований;

β_1 — второй начальный момент длительности второго этапа обслуживания;

$c = \sum\limits_{k=1}^{\infty} kp_k$ — среднее число требований в группе;

$$c_1 = \sum_{k=2}^{\infty} k(k-1)p_k.$$

В случае параметров, имеющих произвольные распределения, остается исследовать СМО имитационными методами. Обозначим для произвольного момента времени t $n(t)$ — число повторных требований в системе; $I(t)$ — состояние прибора системы ($I(t) = 0$ — прибор свободен; $I(t) = 1$ — прибор занят и требование находится на первом этапе обслуживания; $I(t) = 2$ — прибор занят и требование находится на втором этапе обслуживания). Тогда состояние системы описывается процессом $(n(t), I(t))$.

Моделирование системы будем производить способом "$\Delta\tau$". Для этого необходимо рассмотреть все возможные переходы в промежутке $[t, t+\Delta\tau[$, вероятности которых имеют порядок малости не более чем $0(\Delta\tau)$.

Рассмотрим эти переходы. Пусть $n(t) = n,\ I(t) = i,\ n \geq 0,\ i = 0, 1, 2$. Состояние $(n, 0)$ за время $\Delta\tau$ не изменится, если в систему не поступит ни одна группа требований и ни одно из n повторных требований не поступит на первый этап обслуживания. Из состояния $(n, 0)$ возможен переход в состояние $(n-1, 1)$, если в систему не поступит ни одна группа требований, а одно из имеющихся n повторных требований поступит на первый этап обслуживания. Из состояния $(n, 1)$ возможны следующие переходы: $(n, 1) \to (n, 1)$, если за $\Delta\tau$ не закончилось обслуживание требования на первом этапе, не поступило ни одной группы требований извне и ни одно из имеющихся n повторных требований не поступило на первый этап обслуживания; $(n, 1) \to (n, 2)$, если за $\Delta\tau$ закончилось обслуживание требования на первом этапе, за это время не поступило ни одной группы требований в систему и ни одно из n повторных требований не поступило не первый этап обслуживания; $(n, 1) \to (n+k, 0)$ $k \geq 0$, если за время обслуживания требования на первом этапе в систему пришла группа из l требований $(l = k > 0)$, либо хотя бы одно из n повторных требований затребовало обслуживание $(k = 0)$. Из состояния $(n, 2)$ возможны следующие переходы: $(n, 2) \to (n+k, 0)$, если за $\Delta\tau$ не закончится обслуживание на втором этапе, а в систему поступит группа из k требований $(k > 0)$; $(n, 2) \to (n, 0)$, если за $\Delta\tau$ обслуживание на втором этапе закончится и за время обслуживания в течение рассматриваемого времени в систему не поступят новые требования.

Выбрав промежуток $\Delta\tau$ малым, разыгрываем состояния системы в моменты $t = j\Delta\tau,\ j = 1, 2, \ldots$ Для статистической обработки экспериментальных данных необходимо учесть, что случайный процесс $(n(t), I(t), \xi(t))$, где $\xi(t)$ — время обслуживания находящегося на приборе требования, если $I(t) = 2$ и $\xi(t) = 0$ в противном случае, является регенерирующим. Точками регенерации этого процесса служат моменты обращения всех его компонент в нуль. Средняя длина цикла регенерации конечна.

Задание 1. Смоделировать систему в случае экспоненциальных функций распределения $A(t), D(t), B_1(t), B_2(t)$. Подсчитать оценки вероятностей $p_i,\ 0 \leq i \leq 2$, среднего числа повторных требований в системе, среднего времени ожидания для случая одиночного поступления требований. Построить доверительные интервалы для этих оценок.

Задание 2. Смоделировать систему в случае, когда интервалы входящего потока требований, длительности возобновления попыток вторичных требований попасть на обслуживание и длительности обслуживания первого этапа являются случайными величинами фазового типа (распределение Эрланга, H_R—распределение). Подсчитать оценки и построить доверительные интервалы для параметров, указанных в задании 1. Сравнить полученные результаты с результатами задания 1 в ситуации, когда функции распределения параметров систем $A(t), D(t), B_1(t), B_2(t)$ имеют в каждом из заданий одинаковые математические ожидания.

Лабораторная работа. "Моделирование нестационарных СМО"

Цель работы. Разработка алгоритма моделирования СМО, получение выборочных характеристик этой СМО и анализ полученных результатов.

Порядок выполнения работы. На практике часто возникает необходимость оценивания вероятностно-временных характеристик нестационарных систем. Под нестационарными будем понимать следующие СМО.

1. Системы, параметры которых являются функциями от времени. К таким относятся СМО с изменяемой интенсивностью входящего потока, изменяемой интенсивностью обслуживания требований, с изменяемым числом работающих приборов и т. д. У систем подобного типа стационарный режим функционирования может как существовать, так и не существовать.

2. Системы, параметры которых являются постоянными, но такими, что стационарного режима функционирования не существует. (Например однолинейные СМО с загрузкой $\rho \geq 1$.)

Аналитическое решение задач анализа подробных систем сопряжено с большими трудностями, так как для получения, например, распределения числа требований необходимо решать системы дифференциальных уравнений в частных производных с переменными коэффициентами либо системы интегро–дифференциальных уравнений с такими же коэффициентами, что в настоящее время невозможно. Поэтому приходится прибегать к исследованию методами имитационного моделирования. Особенность состоит в следующем. Ранее при моделировании систем в стационарном режиме мы получали одну реализацию достаточно большой длины, и при статистической обработке наблюдения из такой реализации группировались по циклам регенерации, в результате чего получались оценки стационарных характеристик. В нашем случае характеристики системы зависят от времени t. Поэтому необходимо моделировать систему достаточно большое число раз, при одним и тех же начальных состояниях СМО, получать соответствующие независимые реализации процесса функционирования рассматриваемой системы, а для получения оценки какой–либо характеристики в момент t следует усреднять данные этих реализаций, относящиеся к этому моменту. Доверительный интервал строится обычным образом. Например, пусть требуется оценить среднюю длину очереди $\overline{N}_q(t)$ некоторой нестационарной СМО в момент $t = t^*$. Если $N_{q,i}(t)$ — длина очереди в момент t в i–й реализации, $1 \le i \le N$, то искомая оценка запишется в виде

$$\hat{\overline{N}}_q(t^*) = \frac{1}{N}\sum_{i=1}^{N} N_{q,i}(t^*)$$

с доверительным интервалом

$$\left[\hat{\overline{N}}_q(t^*) - \frac{z_{1-\gamma}\hat{\sigma}(t^*)}{\sqrt{N}}, \hat{\overline{N}}_q(t^*) + \frac{z_{1-\gamma}\hat{\sigma}(t^*)}{\sqrt{N}}\right]$$

где

$$\hat{\sigma}^2(t^*) = \frac{1}{N-1}\sum_{i=1}^{N}\Big(N_{q,i}(t^*) - \hat{\overline{N}}_q(t^*)\Big)^2$$

— выборочная дисперсия длины очереди в момент t.

Отметим, что описанная методика годится для оценивания вероятностно–временных характеристик систем с простейшими параметрами, обеспечивающими существование стационарного режима, при тех значениях t, для которых система еще не вышла в этот режим.

Методы математической статистики позволяют ответить на вопрос о том, обеспечивают ли зафиксированные значения параметров СМО существование стационарного режима функционирования этих систем.

Задание 1. Смоделировать СМО $M|M|1|\infty$ с загрузкой, равной 1. Оценить функции средней длины очереди и среднего виртуального времени ожидания в зависимости от времени t, изобразить графики этих функций вместе с доверительными границами.

Задание 2. Смоделировать однолинейную марковскую СМО с ожиданием с заданными переменной интенсивностью входящего потока требований $\lambda(t)$ и переменной интенсивностью обслуживания $\mu(t)$. Методами математической статистики проверить гипотезу о том, что условие $\max_t\lambda(t) < \max_t\mu(t)$ обеспечивает существование стационарного режима рассматриваемой системы. Построить оценки средней длины очереди и среднего времени ожидания вместе с доверительными интервалами.

Задание 3. Методами математической статистики при помощи имитационного моделирования исследовать переходный режим системы $M|GI|1|\infty$ с загрузкой, меньшей 1, при условии, что распределение времени обслуживания относится к фазовому типу (распределение Эрланга, H_R–распределение). Оценить в зависимости от параметра загрузки среднюю длину периода стабилизации, оценить и построить графики средней длины очереди и среднего виртуального времени ожидания в течение этого периода. Построить соответствующие доверительные интервалы.

3.2 Моделирование марковских сетей массового обслуживания.

В предыдущем параграфе рассматривались задачи моделирования СМО, в которых каждое требование проходило только одну операцию обслуживания. Теперь рассмотрим вопросы моделирования СеМО, понимая под этим совокупность СМО, в которой циркулируют требования, переходя в соответствии с матрицей переходных вероятностей из одной системы в другую.

Различают два основных типа СеМО: *открытые* (разомкнутые) и *замкнутые*. В открытую сеть требования поступают из внешнего источника и после окончания обслуживания покидают ее. В замкнутой сети циркулирует постоянное число требований, т. е. требования не поступают извне и не уходят из сети.

Для задания любой СеМО необходимо определить:

1) число требований в сети (для замкнутых сетей) или вероятностные характеристики внешнего источника требований (для открытых сетей);

2) совокупность обслуживающих узлов (систем) $S_i,\ 1 \le i \le R$;

3) описание функционирования каждого узла сети, включая число обслуживающих приборов узла, дисциплину и механизм обслуживания;

4) матрицу переходных вероятностей, отражающую правило движения требований в сети.

В данном параграфе будем исследовать сети, для которых верны следующие предположения:

1) переходы требований между узлами сети описываются неприводимой цепью Маркова;

2) функции распределения длительностей обслуживания в узлах сети являются экспоненциальными, а входящий поток требований для открытой сети — пуассоновский;

3) дисциплина обслуживания: "первым пришел — первым обслужился";

4) существует стационарный режим распределения требований по узлам сети.

При сделанных предположениях сети описанного типа называются *однородными*, *экспоненциальными* или *марковскими сетями*, поскольку вектор состояния сети в момент времени t есть многомерный марковский процесс. Такой процесс является регенеративным, что дает возможность использовать регенеративные методы анализа экспериментальных данных, полученных в результате имитационного моделирования сетей.

Лабораторная работа. "Моделирование открытой марковской СеМО"

Цель работы. Разработка алгоритма моделирования СеМО, получение выборочных характеристик этой СеМО и анализ полученных результатов.

Порядок выполнения работы. Рассматривается разомкнутая СеМО, состоящая из R узлов, причем i—й узел содержит m_i параллельных приборов обслуживания и время обслуживания требований приборами i—го узла распределено по показательному закону с параметром $\mu_i, 1 \le i \le R$, а число мест ожидания перед каждым узлом неограничено. В i—й узел поступает извне (по отношению к сети) простейший поток требований с интенсивностью γ_i. После обслуживания в i—м узле требование с вероятностью r_{ij} переходит в j—й узел и с вероятностью $r_{i0} = 1 - \sum\limits_{j=1}^{R} r_{ij}$ требование покидает сеть.

Состояние рассматриваемой сети в момент времени t описывается случайным вектором $\mathbf{N(t)} = (N_1(t), N_2(t), \ldots, N_R(t))$, где $N_i(t)$ — число требований в i—м узле. При этом $\mathbf{N(t)}$ есть многомерная цепь Маркова. Найдем совместное распределение числа требований в узлах сети в стационарном режиме.

Обозначим через λ_i суммарную интенсивность потока требований в i—й узел в стационарном режиме. Очевидно, что для стационарного режима эта интенсивность состоит из интенсивности потока требований в этот узел извне и суммы интенсивностей требований, поступающих в i—й узел из других СМО, т. е. справедлива следующая система уравнений:

$$\lambda_i = \gamma_i + \sum_{j=1}^{R} \lambda_j r_{ji}, \quad 1 \le i \le R. \tag{3.33}$$

Соотношение (3.33) называют *законом сохранения потока требований*. Условие неприводимости цепи Маркова гарантирует существование единственного положительного решения системы (3.33).

Если существует стационарное распределение $P[\mathbf{K}] = \lim_{t\to\infty} P[\mathbf{N}(\mathbf{t}) = \mathbf{K}]$, где компоненты вектора $K = (k_1, k_2, \ldots, k_R)$ есть возможные неотрицательные целочисленные значения, то система уравнений для стационарных вероятностей $P[\mathbf{K}]$ имеет вид:

$$\lambda_i P[\mathbf{K} - \mathbf{I_i}] = \beta_J(i) P[\mathbf{K} + \mathbf{I_j} - \mathbf{I_i}], \quad 1 \le i, j \le R,$$

где I_i — вектор размерности R, j компоненты которого равны нулю, за исключением i–й, которая равна 1;

$$\beta_j(i) = \begin{cases} (k_j + 1 - \delta_{ji})\mu_i, & \text{если } k_j + 1 - \delta_{ji} < m_i, \\ m_i \mu_i, & \text{если } k_j + 1 - \delta_{ji} \ge m_i, \end{cases}$$

δ_{ij} — символ Кронекера.

При выполнении условия

$$\sum_K \prod_{i=1}^{R} \frac{\lambda_i^{k_i}}{\prod\limits_{l=1}^{k_i} \beta_i(l)} < +\infty$$

существует единственное стационарное распределение $P(\mathbf{K})$ и оно имеет форму

$$P[k_1, k_2, \ldots, k_R] = \prod_{i=1}^{R} P_i[k_i], \tag{3.34}$$

где

$$P_i[k_i] = P_i[0] \prod_{j=1}^{k_i} \frac{\lambda_i}{\beta_i(j)},$$

$$P_i^{-1}[0] = \sum_{j=1}^{\infty} \frac{\lambda_i}{\prod_{i=1}^{j} \beta_i(j)},$$

Из (3.34) следует, что совместные стационарные вероятности состояний рассматриваемой сети выражаются в виде произведения стационарных вероятностей составляющих сеть узлов. Это свидетельствует о том, что открытая марковская сеть является совокупностью R независимых СМО вида $M|M|m_i|\infty$, на которые поступают простейшие входящие потоки с интенсивностями λ_i, определяемыми системой (3.33). Полученный факт определяет первый способ моделирования открытой марковской сети.

В процессе моделирования можно получить оценки следующих характеристик сети:

1) среднего числа требований, находящихся в сети:

$$\overline{n} = \sum_{j=1}^{R} \overline{n}_j; \tag{3.35}$$

2) среднего числа требований, ожидающих обслуживание в сети:

$$\bar{l} = \sum_{j=1}^{R} \bar{l}_j; \tag{3.36}$$

3) среднего времени пребывания требования в сети:

$$\overline{T} = \sum_{j=1}^{R} \frac{\lambda_j}{\sum\limits_{i=1}^{R} \gamma_i} \overline{T}_j; \tag{3.37}$$

4) среднего времени ожидание требования в сети:

$$\overline{w} = \sum_{j=1}^{R} \frac{\lambda_j}{\sum\limits_{i=1}^{R} \gamma_i} \overline{w}_j. \tag{3.38}$$

Экспериментальные данные для получения оценок характеристик систем $M|M|m_i|\infty$, стоящих в правых частях формул (3.35)–(3.38), получаются при помощи алгоритма из лабораторной работы по моделированию СМО $M|M|m|\infty$.

При втором способе моделирования состояний $N_i(t)$ используется тот факт, что в стационарном режиме выходящий поток системы $M|M|m_i|\infty$ является простейшим с интенсивностью, равной интенсивности входящего в эту систему потока. Интенсивность выходящего потока из i–го узла в момент t равна $\alpha_i(t) = min\{N_i(t), m_i\}\mu_i$, а $\gamma(t) = \sum\limits_{i=1}^{R}(\alpha_i(t) + \gamma_i)$ есть интенсивность суммарного потока входящих и выходящих требований в рассматриваемой сети в момент времени t. Этот суммарный поток является простейшим.

Обозначим Δ_j — j–й момент изменения состояний сети. Это либо момент прихода требования в сеть, либо момент окончания обслуживания. Случайная величина $r_j = \Delta_j - \Delta_{j-1}$ $(j \geq 1,\ \Delta_0 \equiv 0)$ распределена по показательному закону с параметром $\gamma(\Delta_j)$; $p_i(\Delta_j) = \gamma_i \,/\, \gamma(\Delta_j)$ — есть вероятность того, что в момент Δ_j требование поступит извне в i–й узел, а с вероятностью $q_i(\Delta_j) = \alpha_i(\Delta_j) \,/\, \gamma(\Delta_j)$ обслуженное требование выходит из i–го узла, $1 \leq i \leq R$.

Второй способ моделирования реализуется при помощи следующего алгоритма. Состояния сети рассматриваются в моменты Δ_j, $\Delta_j = \sum\limits_{i=1}^{j} \tau_i$. Пусть в момент Δ_j состоянию сети соответствует вектор $\mathbf{N}(\Delta_j) = (N_1(\Delta_j), N_2(\Delta_j), \ldots, N_R(\Delta_j))$. Если $j = 0$, то можно положить $\mathbf{N_i}(0) = 0$, $1 \leq i \leq R$. Моделируем новое значение τ_{j+1}. По распределению $\{p_1(\Delta_j), p_2(\Delta_j), \ldots, p_R(\Delta_j), q_1(\Delta_j), \ldots, q_R(\Delta_j)\}$ определяем i^* — номер узла, в котором произошло изменение состояния. Если в этот узел попадает требование извне по отношению к сети, то полагаем

$$N_{i^*}(\Delta_j + 0) = N_{i^*}(\Delta_j) + 1;$$

Если же в i^* узле в момент Δ_j требование окончило обслуживание, то в соответствии с распределением $(p_{i^*0}, p_{i^*1}, \ldots, p_{i^*R})$ определяем j^* — номер узла, в который это требование попадает, и

$$N_{i^*}(\Delta_j + 0) = N_{i^*}(\Delta_j) - 1,$$

$$N_{j^*}(\Delta_j + 0) = N_{j^*}(\Delta_j) + 1, \text{ если } j^* \neq 0.$$

Все остальные узлы в момент Δ_j изменений не претерпевают, т. е.

$$N_i(\Delta_j + 0) = N_i(\Delta_j), \text{ при } i \neq i^* \text{ и } i \neq j^*.$$

Для обработки экспериментальных данных отметим, что процесс $N(\Delta_j)$ является регенеративным, а моментами регенерации являются те значения Δ_j , при которых вектор $\mathbf{N(t)}$ попадает в некоторое фиксированное состояние.

Задание 1. Реализовать первый способ моделирования сети. Используя регенеративный метод, получить точечные оценки параметров (3.35)–(3.38) и построить соответствующие доверительные интервалы. Исследовать, как эти доверительные интервалы накрывают значения параметров (3.35)–(3.38) в зависимости от уровня значимости и числа циклов регенерации n.

Задание 2. Реализовать второй способ моделирования сети. Построить оценки и доверительные интервалы параметров (3.35)–(3.38). Сравнить полученные результаты с результатами задания 1.

Лабораторная работа. "Моделирование замкнутой марковской СеМО"

Цель работы. Разработка алгоритма моделирования СеМО, получение выборочных характеристик этой СеМО и анализ полученных результатов.

Порядок выполнения работы. Рассматривается замкнутая СеМО с R узлами, в которой циркулируют L требований; i–й узел представляет собой m_i параллельных приборов обслуживания, Время обслуживания требования приборами i–го узла распределено по показательному закону с параметром μ_i, и число мест в очереди перед каждым узлом неограничено. После окончания обслуживания в i–м узле требование с вероятностью r_{ij} поступает в j–й узел, $1 \le i, j \le R$.

Вектор $\mathbf{N}(\mathbf{t}) = (N_1(t), N_2(t), \ldots, N(t))$, где $N_i(t)$ — число требований в i–м узле в момент t, описывает состояние рассматриваемой сети. $N_i(t) \in \{0, 1, 2, \ldots, L\}$ и $\sum\limits_{i=1}^{R} N_i(t) = L$. Число различных состояний сети конечно и равно числу способов размещения L требований по R узлам, т. е. $\binom{L+R-1}{R-1}$.

Поскольку в замкнутых сетях требования не поступают извне и не покидают сеть, то $\gamma_i \equiv 0$, $r_{i0} \equiv 0$, $1 \le i \le R$. Соотношения для интенсивностей потока требований в узлах сети записываются в виде

$$\lambda_i = \sum_{j=1}^{R} \lambda_j r_{ji}, \quad 1 \le i \le R. \tag{3.39}$$

Интенсивности λ_i определяются из (3.39) с точностью до константы, и для однозначного определения λ_i можно ввести дополнительное ограничение

$$\sum_{j=1}^{R} \lambda_j = 1. \tag{3.40}$$

Тогда решение системы уравнений (3.39)–(3.40) определяется единственным образом и λ_i можно трактовать как вероятность того, что в стационарном режиме произвольно выбранный переход из одного узла в другой окажется переходом в i–й узел.

Система уравнений равновесия для стационарных вероятностей $P[K] = \lim\limits_{t\to\infty} P[\mathbf{N}(\mathbf{t}) = \mathbf{K}]$ имеет вид

$$\frac{\lambda_i}{\lambda_j}\beta_i(k_i)P[\mathbf{K}] = \beta_j(k_j + 1 - \delta_{ij})P[\mathbf{K} + \mathbf{I_j} - \mathbf{I_i}],\ 1 \le i, j, \le N \tag{3.41}$$

где $\beta_s(l) = \min\{m_s, l\}\mu_s$.

Из (3.41) получаем, что

$$P[\mathbf{K}] = \frac{1}{G(L,R)} \prod_{i=1}^{R} \frac{\lambda_i^{k_i}}{\prod\limits_{l=1}^{k_i} \beta_i(l)}, \tag{3.42}$$

где $G(L, R)$ — нормировочная константа, определяемая из условия

$$\sum_{\mathbf{K}\in Z(L,R)} P[\mathbf{K}] = 1, \quad Z(L,R) = \{\mathbf{K}\,|\,k_i \geq 0,\ 1 \leq i \leq R,\ \sum_{i=1}^{R} k_i = L\}.$$

Но при больших значениях L расчет нормировочной константы $G(L,R)$ требует значительных вычислительных усилий. Ведь требуется производить суммирование соответствующих произведений по $\begin{pmatrix} L+R-1 \\ R-1 \end{pmatrix}$ состояниям. Кроме того при подсчете $G(L,R)$ на ЭВМ возникают проблемы "исчезновения порядка" и обнуления результатов. Поэтому для вычисления $G(L,R)$ требуются специальные алгоритмы, один из которых, называемый *алгоритм свертки*, приводится ниже.

Обозначим

$$g(l,r) = \sum_{\mathbf{K}\in Z(l,r)} \prod_{i=1}^{r} y_i(k_i), \quad \text{где } y_i(k_i) = \frac{\lambda_i^{k_i}}{\prod\limits_{l=1}^{k_i} \beta_i(l)}.$$

Очевидно, что $G(L,R) = g(L,R)$, и при $n > 1$ получаем

$$g(l,r) = \sum_{s=0}^{l} \left[\sum_{\mathbf{K}\in Z(l,r)} \prod_{i=1}^{r} y_i(k_i) \right] =$$

$$= \sum_{s=0}^{l} y_r(s) \sum_{\substack{\mathbf{K}\in Z(l-s,r-1) \\ k_s=s}} \prod_{i=1}^{r-1} y_i(k_i) = \sum_{s=0}^{l} y_r(s) g(l-s,r-1),$$

то есть

$$G(L,R) = \sum_{s=0}^{L} y_R(s) G(L-s,R-1). \tag{3.43}$$

К тому же имеем

$$y_i(s) = \frac{\lambda_i}{\beta_i(s)} y_i(s-1). \tag{3.44}$$

Задав начальные условия: $y_i(0) = 1,\ G(l,1) = y_1(l),\ 1 \leq l \leq R$, и $G(0,r) = 1,\ 1 \leq r \leq L$ по рекуррентным формулам (3.43)–(3.44), подсчитывают значение константы $G(L,R)$. Для этого требуется $(R-1)(L+1)(L+2)\,/\,6$ операций умножения и сложения, что значительно меньше $\begin{pmatrix} L+R-1 \\ R-1 \end{pmatrix}$ — числа операций, необходимых для вычисления $G(L,R)$ по формулам, использующим условие нормировки при больших R и L.

Распределение (3.42) позволяет определить ряд вероятностных характеристик рассматриваемой СеМО. Можно получить стационарное распределение числа требований в i—м узле:

$$p_i(m) = \lim_{t\to\infty} P[N_i(t) = m] = \sum_{\mathbf{K}\in Z(L,R)} P(\mathbf{K}), \quad 0 \leq m \leq L. \tag{3.45}$$

Вероятность (3.45) дает возможность определить среднее число требований в i—м узле в стационарном режиме:

$$\overline{n}_i = \sum_{m=1}^{L} m P_i(m). \tag{3.46}$$

Но интерес представляют не только средние числа требований в каждом узле сети, но и другие характеристики. Многие из них могут быть получены с помощью алгоритмов, не использующих стационарное распределение (3.42). Эти алгоритмы относятся к так называемому анализу средних значений и базируются на законах сохранения потока требований в узлах сети. Приведем два таких алгоритма.

Алгоритм, рекуррентный по числу требований. Обозначим через $\overline{n}_i(L), \overline{\beta}_i(L)$ и $\overline{\omega}_i(L),\ 1 \leq i \leq R,$ — соответственно среднее число требований, среднее число занятых приборов и среднее время ожидания требований в i–м узле в стационарном режиме, когда в сети циркулируют L требований. Параметры $\overline{n}_i(L)$ и $\overline{\beta}_i(L)$ связаны соотношением

$$\overline{\beta}_i(L) = \min\{\overline{n}_i(L), m_i\}, \quad 1 \leq i \leq N. \tag{3.47}$$

Обозначим через $\xi_i(L)$ — среднее число требований в i–м узле в момент поступления нового требования. Тогда

$$\overline{\omega}_i(L) = \frac{1}{\mu_i}(1 + \xi_i(L)), \quad 1 \leq i \leq R,$$

так как среднее время ожидания $\overline{\omega}_i(L)$ состоит из среднего времени обслуживания поступившего требования, равного μ^{-1}, и среднего времени обслуживания всех требований, находившихся в i–м узле в момент поступления этого требования, равного $\xi_i(L)\mu^{-1}$.

Можно показать, что $\xi_I(L) = \overline{n}_i(L-1)$, и тогда

$$\overline{\omega}_i(L) = \frac{1}{\mu_i}(1 + \overline{n}_i(L-1)). \tag{3.48}$$

В стационарном режиме интенсивность потока требований, поступающих в i–й узел, равна интенсивности выходящего потока из этого узла и поэтому

$$\lambda_i = \mu_i \overline{\beta}_i(L), \quad 1 \leq i \leq R.$$

Из формулы Литтла определяется средняя длительность ожидания требования в узле:

$$\overline{\omega}_i(L) = \frac{\overline{n}_i(L) - \overline{\beta}_i(L)}{\mu_i \overline{\beta}_i(L)}, \quad 1 \leq i \leq R. \tag{3.49}$$

Задавшись начальным условием $\overline{n}(0) = 0,\ 1 \leq i \leq R$ по формулам (3.47)–(3.49) рассчитываются указанные средние характеристики.

Алгоритм, рекуррентный по моментам времени. Рассмотрим функционирование сети в интервале времени $[0, T^0]$, и пусть $\overline{n}_i(t), \overline{\beta}_i(t)$ и $\overline{w}_i(t),\ 1 \leq i \leq R,$ соответственно среднее число требований, среднее число занятых приборов и среднее время ожидания требований в i–м узле сети в интервале $[0, t[$. Рассуждениями, аналогичными тем, которые были проведены для предыдущего алгоритма, получаем следующий алгоритм для нахождения рассматриваемых средних характеристик:

$$\overline{\beta}_i(t) = \min\{\overline{n}_i(t), m_i\},$$

$$\overline{w}_i(t+1) = \frac{\overline{n}_i(t) - \overline{\beta}_i(t)}{\mu_i \overline{\beta}_i(t)},$$

$$\overline{n}_i(t+1) = L\overline{\omega}_i(t+1) \Big/ \sum_{j=1}^{R} \mathrm{e}_{ij} \overline{w}_i(t+1), \tag{3.50}$$

где $\mathrm{e}_{ij} = \mu_j \overline{\beta}_j(L) \,/\, \mu_i \overline{\beta}_i(L), \quad 1 \leq i, j \leq R.$

При $t \to \infty$ последовательность (3.50) сходится при любом начальном распределении требований по узлам сети, только бы

$$\sum_{i=1}^{R} \overline{n}_i(0) = L.$$

Моделирование процесса функционирования рассматриваемой сети проводится аналогично второму способу моделирования открытой марковской СеМО. При $t = 0$ необходимо задать (или смоделировать) начальное распределение требований по узлам.

Моментами $\{\Delta_j\}$ изменения состояний вектора $\mathbf{N(t)}$ являются моменты окончания обслуживания требований узлами сети. Последовательность этих моментов в стационарном режиме образует пуассоновский поток с интенсивностью, равной суммарной интенсивности обслуживания требований, т. е. в момент Δ_j эта интенсивность равна $\gamma(\Delta_j) = \sum_{l=1}^{R} \mu_l \min\{N_l(\Delta_j), m_l\}$. С вероятностью $p_i(\Delta_j) = \mu_i \min\{N_i(\Delta_j), m_i\} \,/\, \gamma(\Delta_j)$ в момент Δ_j обслуживание требований было закончено i–м узлом. Следовательно по распределению $\{p_1(\Delta_j), p_2(\Delta_j), \ldots, p_R(\Delta_j)\}$ определяем i^* — номер узла, который в момент Δ_j закончит обслуживание требования и полагаем

$$N_{i^*}(\Delta_j + 0) = N_{i^*}(\Delta_j) - 1, \quad N_i(\Delta_j + 0) = N_i(\Delta_j) \quad \forall i,\ i \neq i^*.$$

Далее по распределению $\{p_{i^*1}, p_{i^*2}, \ldots, p_{i^*R}\}$ определяем j^* — номер узла, в который было направлено требование после окончания обслуживания в момент Δ_j в узле i^* , и

$$N_{j^*}(\Delta_j + 0) = N_{j^*}(\Delta_j + 1).$$

Для остальных узлов в рассматриваемый момент Δ_j изменений не происходит, т. е.

$$N_i(\Delta_j + 0) = N_i(\Delta_j) \quad \text{при } 1 \leq i \leq R \text{ и } i \neq i^*,\ i \neq j^*.$$

Отметим, что процесс $N(t)$ является неприводимой и возвратной цепью Маркова, и, следовательно, этот процесс является регенерирующим. Последовательные моменты возвращения $N(t)$ в некоторое фиксированное состояние $\mathbf{K^0} = (k_1^0, k_2^0, \ldots, k_R^0)$ являются точками регенерации этого процесса. Таким образом, для изучения величин, связанных с числом требований в узлах сети можно применять регенеративный метод.

Задание 1. Реализовать оба рекуррентных алгоритма. При помощи алгоритма свертки подсчитать нормировочную константу $G(L, R)$, вычислить стационарные вероятности (3.45) и подсчитать параметры $\overline{n}_i$ по формуле (3.46). Сравнить полученные таким образом значения $\overline{n}_i$ со значениями этих характеристик, вычисленными при помощи рекуррентных алгоритмов.

Задание 2. Реализовать алгоритм моделирования сети, построить точечные оценки и доверительные интервалы параметров $\overline{n}_i, \overline{\beta}_i$ и $\overline{w}_i$. Сравнить результаты моделирования со значениями соответствующих параметров, подсчитанными в задании 1.

Лабораторная работа. "Моделирование замкнутой марковской СеМО в нестационарном случае“

Цель работы. Разработка алгоритма моделирования СеМО, получение выборочных характеристик этой СеМО и анализ полученных результатов.

Порядок выполнения работы. В лабораторных работах по моделированию открытой марковской СеМО и замкнутой марковской СеМО, анализ марковских СеМО проводился для случая, когда сеть находится в состоянии статистического равновесия, т. е. рассматривается стационарный режим работы сети, и вероятности состояний при этом не зависят от времени. Однако на практике часто встречаются случаи, когда желательно определить характеристики сети в конкретный момент времени t, например, при исследовании вычислительных сетей и сетей связи. В данной лабораторной работе рассматривается замкнутая марковская сеть, приводится метод нахождения вероятностей состояний, зависящих от времени, и описывается алгоритм статистической обработки данных моделирования.

Обозначим через $P(\mathbf{K},t) = P[N_1(t) = k_1, N_2(t) = k_2, \ldots, N_R(t) = k_R]$ — вероятность того, что в момент t сеть находится в состоянии $\mathbf{K} = (k_1, k_2, \ldots, k_R)$. Анализируя возможные переходы процесса $\mathbf{N(t)}$ за промежуток времени $\Delta\tau$ и переходя к пределу при $\Delta\tau \to 0$, получаем следующую систему прямых дифференциальных уравнений Колмогорова:

$$\frac{dP(\mathbf{K},t)}{dt} = -\Lambda(\mathbf{K})P(\mathbf{K},t) + \sum_{i=1}^{R}\sum_{j=1}^{R}\varepsilon(k_i)r_{ji}\beta_j(k_j+1-\delta_{ij})\times$$

$$\times P(\mathbf{K}+\mathbf{I_j}-\mathbf{I_i},t), \tag{3.51}$$

где

$$\varepsilon(k_i) = \begin{cases} 0, & k_i = 0, \\ 1, & k_i > 0, \end{cases} \quad \Lambda(\mathbf{K}) = \sum_{i=1}^{R}\beta_i(k_i).$$

Уравнение (3.51) можно решить методом последовательных приближений. Имеем

$$\frac{\partial}{\partial t}\left\{\mathrm{e}^{\Lambda(\mathbf{K})t}P(\mathbf{K},t)\right\}\mathrm{e}^{\Lambda(\mathbf{K})t}\sum_{i=1}^{R}\sum_{j=1}^{R}\varepsilon(k_i)r_{ji}\beta_j\times$$

$$\times (k_j+1-\delta_{ij})P(\mathbf{K}+\mathbf{I_j}-\mathbf{I_i},t). \tag{3.52}$$

Пусть $P_m(\mathbf{K},t),\ m = 0,1,2,\ldots$ — приближение функции $P(\mathbf{K},t)$ на m−й итерации и пусть $(m+1)$−я итерация этой функции определяется из (3.52) следующим образом:

$$P_{m+1}(\mathbf{K},t) = \mathrm{e}^{-\Lambda(\mathbf{K})t}P_m(\mathbf{K},0) + \mathrm{e}^{-\Lambda(\mathbf{K})t}\int_0^t \mathrm{e}^{\Lambda(\mathbf{K})\tau}\times$$

$$\times \sum_{=1}^{R}\sum_{j=1}^{R}\left[\varepsilon(k_i)r_{ji}\beta_j(k_j+1-\delta_{ij})P_m(\mathbf{K}+\mathbf{I_j}-\mathbf{I_i},\tau)\right]\,d\tau. \tag{3.53}$$

За начальное приближение можно взять стационарное распределение вероятностей состояний

$$P_0(\mathbf{K},t) = \lim_{t\to\infty}P(\mathbf{K},t) = P(\mathbf{K}),$$

вычисляемое по формуле (3.42). Из (3.51) видно, что функция $P(\mathbf{K})$ удовлетворяет уравнению

$$P(\mathbf{K})\Lambda(\mathbf{K}) = \sum_{i=1}^{R}\sum_{j=1}^{R}\varepsilon(k_i)r_{ji}\beta_j(k_j+1-\delta_{ij})P(\mathbf{K}+\mathbf{I_j}-\mathbf{I_i},t)$$

и поэтому

$$P_1(\mathbf{K},t) = \mathrm{e}^{-\Lambda(\mathbf{K})t}P(\mathbf{K},0) + \left(1-\mathrm{e}^{-\Lambda(\mathbf{K})t}\right)P(\mathbf{K}). \tag{3.54}$$

Имеют место два утверждения:

1) $\forall\, m > 0 \quad P_m(\mathbf{K},0) = P(\mathbf{K},0)$ и $P_m(\mathbf{K},t) \xrightarrow[t\,\to\,\infty]{} P(\mathbf{K})$;

2) последовательность $\{P_m(\mathbf{K},t),\ m = 0,1,2,\ldots\}$, полученная по схеме (3.53), сходится $\forall\, t \geq 0$ при $m \to \infty$.

Для реализации на ЭВМ метода последовательных приближений желательно функции $P_m(\mathbf{K},t)$ представлять в виде степенного ряда по t. Пусть

$$P_m(\mathbf{K},t) = \sum_{l=0}^{\infty}a_{ml}(\mathbf{K})t^l.$$

Тогда из формулы (3.53) следует, что

$$\sum_{l=0}^{\infty} a_{m+1,l}(\mathbf{K})t^l = \mathrm{e}^{-\Lambda(\mathbf{K})t}P(\mathbf{K},0) + \sum_{l=0}^{\infty} A_{ml}(\mathbf{K})B_l(t), \tag{3.55}$$

где

$$A_{ml}(\mathbf{K}) = \sum_{i=1}^{R}\sum_{j=1}^{R} \varepsilon(k_i)r_{ji}\beta_j(k_j+1-\delta_{ij})a_{ml}(\mathbf{K}+\mathbf{I_j}-\mathbf{I_i}),$$

$$B_l(t) = \mathrm{e}^{-\Lambda(\mathbf{K})t}\int\limits_0^t x^l \mathrm{e}^{\Lambda(\mathbf{K})\tau}\,d\tau = \frac{(-1)^{l+1}\mathrm{e}}{\{\Lambda(\mathbf{K})\}^{l+1}}\sum_{r=l+1}^{\infty}\frac{\{-\Lambda(\mathbf{K})t\}^r}{r!}.$$

Приравнивая в (3.55) коэффициенты при одинаковых степенях t, получаем

$$a_{m+1,l}(\mathbf{K}) = \frac{\{-\Lambda(\mathbf{K})\}^l}{l!}\left\{P(\mathbf{K},0) - \sum_{j-0}^{l-1} A_{mj}(\mathbf{K})j![-\Lambda(\mathbf{K})]^{-j-1}\right\}, \tag{3.56}$$

$$a_{m+1,0}(\mathbf{K}) = P(\mathbf{K},0), \tag{3.57}$$

$$a_{0,l}(\mathbf{K}) = P(\mathbf{K})\delta_{l0}. \tag{3.58}$$

Рекуррентные формулы (3.56)–(3.58) позволяют находить коэффициенты разложения $P_m(\mathbf{K},t)$ в степенной ряд.

Как и в стационарном случае распределение $P(\mathbf{K},t)$ позволяет определить вероятностно–временные характеристики рассматриваемой сети в момент t, а именно: распределение числа требований в i–м узле, среднее число требований в i–м узле и т. д.

Алгоритм моделирования сети изложен в предыдущей лабораторной работе. Методика стационарного исследования такая же, как и для нестационарных систем; т. е. поскольку характеристики сети зависят от времени t, необходимо моделировать сеть достаточно большое число раз при одних и тех же начальных состояниях сети, генерировать независимые реализации процесса функционирования сети, а для получения оценки какой-либо характеристики в момент t следует усреднять данные этих реализаций, относящиеся к этому моменту.

Задание 1. Реализовать рекуррентную процедуру подсчета вероятностей состояний $P(\mathbf{K},t)$ при помощи формул (3.56)–(3.58). Численно удостовериться в том, что при увеличении m и t функции $P_m(K,t)$ приближаются к стационарным вероятностям $P(\mathbf{K})$.

Задание 2. Реализовать алгоритм моделирования сети, оценить функции средней длины очереди и среднего виртуального времени ожидания по узлам сети, изобразить графики этих функций вместе с доверительными границами.

Лабораторная работа. "Моделирование открытой марковской СеМО в нестационарном случае“

Цель работы. Разработка алгоритма моделирования СеМО, получение выборочных характеристик этой СеМО и анализ полученных результатов.

Порядок выполнения работы. Рассматривается открытая марковская сеть, описанная в лабораторной работе по моделированию открытой марковской СеМО. Цепь исследования состоит в получении рекуррентных формул для подсчета вероятностей состояний этой сети, зависящих от времени, моделировании и обработки результатов наблюдений.

Система прямых дифференциальных уравнений Колмогорова для функций$P(\mathbf{K},t)$ имеет вид:

$$\frac{dP(\mathbf{K},t)}{dt} = -\Lambda(\mathbf{K})P(\mathbf{K},t) + \sum_{i=1}^{R}\beta_i(k_i+1)r_{i0}P(\mathbf{K}+\mathbf{I_i},t) +$$

$$\sum_{i=1}^{R}\varepsilon(k_i)\gamma_j P(\mathbf{K}-\mathbf{I_i},t)+\sum_{i=1}^{R}\sum_{j=1}^{R}\varepsilon(k_i)r_{ji}\beta_j(k_j+1-\delta_{ij})P(\mathbf{K}+\mathbf{I_j}-\mathbf{I_i},t),$$

где

$$\Lambda(\mathbf{K})=\sum_{i=1}^{R}\gamma_i+\sum_{i=1}^{R}\beta_i(k_i),$$

а далее вывод рекуррентных формул для подсчета вероятностей состояний $P(\mathbf{K},t)$ производится аналогично тому, как это осуществлялось в случае замкнутой марковской сети. Аналогично производится и статистическая обработка результатов имитационных экспериментов.

Задание 1. Получить рекуррентные формулы для подсчета вероятностей состояний $P(\mathbf{K},t)$, численно удостовериться в том, что при увеличении m и t функции $P_m(\mathbf{K},t)$ приближаются к стационарным вероятностям $P(\mathbf{K})$.

Задание 2. Реализовать алгоритм моделирования сети вторым способом, оценить функции среднего числа требований и среднего виртуального времени ожидания в i—м узле сети, среднего времени пребывания и среднего времени ожидания требования в сети, изобразить графики этих функций вместе с доверительными границами.

3.4. Моделирование марковских СеМО с разнотипными требованиями.

Обобщим рассмотренные в предыдущем параграфе однородные марковские сети на случай сетей, в которых циркулируют требования нескольких типов. Подобные сети являются хорошей математической моделью вычислительных систем.

Лабораторная работа. "Моделирование открытой марковской СеМО с разнотипными требованиями"

Цель работы. Разработка алгоритма моделирования СеМО, получение выборочных характеристик этой СеМО и анализ полученных результатов.

Порядок выполнения работы. Рассмотрим открытую СеМО, состоящую из R узлов, между которыми циркулируют требования Q различных типов; i—й узел $(1 \le i \le R)$ содержит m_i параллельных приборов обслуживания с неограниченным числом мест ожидания. Время обслуживания требований всех типов любым прибором i—го узла имеет показательное распределение с параметром μ_i, $1 \le i \le R$, дисциплина обслуживания — "первым пришел — первым обслужился". При переходе из одного узла в другой требования могут изменять свой тип. Структура движения требований в сети и возможность изменения типа требований задаются матрицей $P=\|p_{ir,ks}\|$, где $p_{ir,ks}$ — вероятность того, что требование r—го типа, закончившее обслуживание в i—м узле, перейдет в k—й узел и станет требованием s—го типа.

В сеть поступает пуассоновский поток требований с интенсивностью Λ. С вероятностью $q_{0,ir}$ требование, поступившее извне (по отношению к сети), становится требованием r—го типа и поступает для обслуживания в i—й узел. Требование типа r, закончившее обслуживание в i—м узле, покидает сеть с вероятностью

$$p_{ir,j}=1-\sum_{k=1}^{R}\sum_{s=1}^{R}p_{ir,ks}.$$

Исследование рассматриваемой цепи аналогично тому, которое проводилось для марковских цепей со статистически однородными требованиями, но при нахождении распределения числа требований по узлам сети вектор состояний должен включать в себя и типы требований, находящихся на обслуживании и стоящих в очереди.

Состояние i—го узла в момент времени t определяется вектором $\mathbf{N_i}(t) = (\mathbf{N_{i1}}(t),\mathbf{N_{i2}}(t),\dots,\mathbf{N_{in_i}}(t))$, где $N_{ij}(t)$ — номер типа требования, стоящего в узле на j—м месте. $1 \leq N_{ij}(t) \leq Q,\ 1 \leq i \leq R,\ 1 \leq j \leq n_i,\ n_i \geq 1$. Если в момент t i–й узел свободен, то полагаем $N_i(t) = 0$.

Состояние сети в момент t представляется вектором $\mathbf{N}(t) = (N_1(t), N_2(t), \ldots, N_R(t))$, распределение которого обозначим через $p(\mathbf{N}(t))$.

Уравнения для нахождения стационарных вероятностей $p(\mathbf{N}) = \lim\limits_{t\to\infty} p(\mathbf{N}(t))$ записываются в общем виде следующим образом:

$$\lambda(\mathbf{N})P(\mathbf{N}) = \sum_{N'} \lambda(\mathbf{N}', \mathbf{N})P(\mathbf{N}'), \tag{3.59}$$

где $\lambda(\mathbf{N})$ — интенсивность выхода сети из состояния $\mathbf{N}$, $\lambda(\mathbf{N}', \mathbf{N})$ — интенсивность перехода сети из состояния $\mathbf{N}'$ в состояние $\mathbf{N}$. Стационарное распределение вероятностей имеет вид

$$P(\mathbf{N}) = G^{-1}\Lambda^{\sum_{i=1}^{R} n_i} \prod_{i=1}^{R} \mu_i^{-n_i} \prod_{j=1}^{n_i} \mathrm{e}_j \mathbf{N_{ij}}, \tag{3.60}$$

где G, e_j — некоторые положительные константы. Стационарное распределение $P(\mathbf{N})$ существует, если выполняется условие

$$\sum_{N} \Lambda^{\sum_{i=1}^{R} n_i} \prod_{i=1}^{R} \mu_i^{-n_i} \prod_{j=1}^{n_i} \mathrm{e}_j N_{ij} < +\infty.$$

По распределению (3.60) отыскивается ряд средних характеристик сети таких, например, как среднее число требований определенного типа в каждом узле, суммарное по всем типам среднее число требований в каждом узле и т. д. Но нахождение этих характеристик представляет собой трудную задачу с точки зрения численной реализации, поэтому при исследовании сетей подобного типа разумно применять методы имитационного моделирования.

Алгоритм моделирования сети использует тот факт, что в стационарном режиме выходящие потоки из узлов сети являются пуассоновскими, и, следовательно, суммарный поток сети также является пуассоновским. В момент времени t интенсивность выходящего потока из $i0$–го узла равна $\alpha_i(t) = \min\{n_i(t), m_i\}\mu_i$, а $\gamma(t) = \sum\limits_{i=1}^{R} \alpha_i(t) + \Lambda$ есть интенсивность суммарного потока выходящих требований в момент времени t.

Моделирование сети происходит по схеме, аналогичной схеме моделирования марковской однородной открытой сети вторым способом с тем лишь добавлением, что необходимо моделировать изменение типа требования при переходе его от одного узла к другому. А именно, если требование r–го типа, закончившее обслуживание в i^* узле, переходит в j^*–й узел, то новый тип этого требования разыгрывается согласно вектору вероятностей

$$\{p_{i^*r,j^*1}, p_{i^*r,j^*2}, \ldots, p_{i^*r,j^*Q}\}.$$

При регенеративной обработке данных моделирования учесть, что вектор $\mathbf{N}(t)$ является неприводимой марковской цепью и, следовательно, регенеративным процессом.

Задание. Смоделировать рассматриваемую сеть. Подсчитать оценки среднего числа и среднего времени ожидания требованиями каждого типа в узлах сети и оценки среднего времени прохождения требованиями каждого типа всей сети, т. е. среднего промежутка времени между вхождением требования в сеть и оставлением сети этим требованиям. Построить доверительные интервалы этих параметров.

Лабораторная работа. "Моделирование замкнутой марковской СеМО с разнотипными требованиями“

Цель работы. Разработка алгоритма моделирования СеМО, получение выборочных характеристик этой СеМО и анализ полученных результатов.

Порядок выполнения работы. Данная сеть отличается от открытой марковской СеМО с разнотипными требованиями тем, что в сети циркулирует фиксированное число L требований, требования не приходят в сеть извне, т. е. $\Lambda \equiv 0$, и не могут покидать сеть, т. е. $p_{ir,j} \equiv 0$, $\forall\, 1 \leq i, r \leq R$, $\forall\, 1 \leq r \leq Q$.

Вектор состояний сети и уравнение для стационарных вероятностей состояний такие же, как и для открытой сети. Стационарные вероятности записываются в виде

$$p(N) = G^{-1} \prod_{i=1}^{R} \mu_i^{-n_i} \prod_{j=1}^{n_i} \mathrm{e}_j N_{ij}.$$

Сеть моделируется также, как и закрытая однородная марковская СеМО с тем добавлением, что необходимо моделировать изменение типа требования. Это делается так же, как и в работе по моделированию открытой марковской СеМО с разнотипными требованиями.

Задание. Смоделировать рассматриваемую сеть. Подсчитать оценки числа требований каждого типа в узлах сети и оценки среднего времени ожидания в узлах сети. Построить соответствующие доверительные интервалы.

Программа **COMBIN** вырабатывает всевозможные сочетания по M чисел из множества $\{1, 2, \ldots, N\}$. По каждому обращению в **COMBIN** получается один из возможных наборов.

Обращение в программе **COMBIN** имеет вид

COMBIN (M,N,J);
M, N : integer;
Type mas = array [1..X];
J : mas;

Примечание $N \geq X$.

Генерируемые наборы на выходе содержатся в массиве J.

Текст программы **COMBIN** следующий:

```
PROCEDURE COMBIN (ns, ks : integer; var : mas1);
 Label m5, m10;
 Var ia, ib, m1, mt, nb : integer;
 Begin
   for ib := 1 to ks do
   begin
     if J[ib] < ib then goto m10;
     ia := J[ib] - ib - 1;
     for mt := 1 to ib do J[mt] := mt + ia;
     goto m5;
m10 :
   end;
   nb := ns - ks - 1;
   for m1 := 1 to ks do J[m1] := nb + m1;
m5 :
 end;  {COMBIN}
```

Литература

[1] **Гнеденко Б.В.** Курс теории вероятностей. М.: Наука. 1988.

[2] **Гнеденко Б.В.**, **Коваленко И.Н.** Введение в теорию массового обслуживания. М.: Наука. 1987.

[3] **Ивченко Г.И.**, **Каштанов В.А.**, **Коваленко И.Н.** Теория массового обслуживания. М.: Высш. шк. 1987.

[4] **Кениг Д.**, **Штойян Д.** Методы теории массового обслуживания. М.: Радио и связь. 1981.

[5] **Клейнрок Л.** Теория массового обслуживания. М.: Машиностроение. 1979.

[6] **Матвеев В.Ф.**, **Ушаков В.Г.** Системы массового обслуживания. М.: Изд-во МГУ. 1984.

[7] **Чистяков В.П.** Курс теории вероятностей. М.: Наука. 1982.

Оглавление