ГЛАВА 10

МАРКОВСКИЕ ПРОЦЕССЫ. ПОТОКИ СОБЫТИЙ. ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ

Говорят, что в физической системе X происходит случайный процесс, если она с течением времени может под влиянием случайных факторов переходить из состояния в состояние.

Система X называется *системой с дискретными состояниями*, если она имеет счетное (в частном случае — конечное) множество возможных состояний $x_1, x_2, \ldots, x_n, \ldots$, и переход из одного состояния в другое осуществляется скачком. Ниже будут рассматриваться только системы с дискретными состояниями.

Возможные состояния системы X наглядно изображаются с помощью так называемого *графа состояний* (рис. 10а), на котором состояния системы изображены прямоугольниками, а возможные переходы системы из состояния в состояние — стрелками, соединяющими соответствующие прямоугольники.

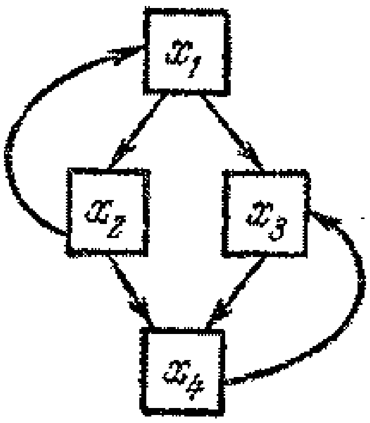

Рис. 10а.

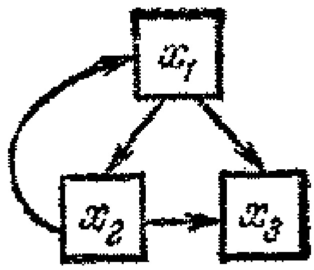

Рис. 10б.

На рис. 10а показан граф состояний системы, имеющей четыре возможных состояния: x_1, x_2, x_3, x_4. Из состояния x_1 возможны переходы в x_2 или x_3; из состояния x_2 — в x_4 или обратно в x_1; из состояния x_3 — в x_4, из состояния x_4 — обратно в x_3.

Состояние системы называется «состоянием без выхода», если из него невозможен переход ни в какое другое состояние (см. состояние x_3 на рис. 10б).

Для описания случайного процесса, протекающего в системе с дискретными состояниями $x_1, x_2, \ldots, x_n$, часто пользуются *вероятностями состояний*

$$p_1(t), \quad p_2(t), \ \ldots, \ p_n(t),$$

где $p_k(t)$ $(k=1, 2, \ldots, n)$ — вероятность того, что в момент t система находится в состоянии x_k. Вероятности $p_k(t)$ удовлетворяют условию

$$\sum_{k=1}^{n} p_k(t) = 1.$$

Случайный процесс, протекающий в системе X, называется *процессом с дискретным временем*, если переходы системы из состояния в состояние возможны только в определенные моменты времени $t_1, t_2, \ldots$ Если переходы возможны в любой момент времени, процесс называется *процессом с непрерывным временем.*

Если в системе X с дискретными состояниями происходит случайный процесс с непрерывным временем, то переходы системы из состояния в состояние можно рассматривать как происходящие под влиянием некоторых *потоков событий* (см. гл. 5 стр. 92). Случайный процесс с дискретными состояниями называется *марковским*, если все вероятностные характеристики процесса в будущем зависят лишь от того, в каком состоянии этот процесс находится в настоящий момент времени, и не зависят от того, каким образом этот процесс протекал в прошлом («будущее зависит от прошлого только через настоящее»). Если процесс марковский, то все потоки событий, переводящие систему из состояния в состояние, являются пуассоновскими.

Если процесс, протекающий в системе с дискретными состояниями и непрерывным временем, является марковским, то для вероятностей состояний $p_1(t)$, $p_2(t)$, $\ldots$, $p_n(t)$ можно составить систему линейных дифференциальных уравнений.

При составлении этих дифференциальных уравнений удобно пользоваться графом состояний системы, на котором против каждой стрелки, ведущей из состояния в состояние, проставлена плотность (интенсивность) потока событий, переводящего систему из состояния в состояние по данной стрелке. Образец такого графа (*размеченного графа состояний*) показан на рис. 10в. Здесь $\lambda_{i,j}$ обозначает плотность потока событий, переводящего систему из состояния x_i в состояние x_j.

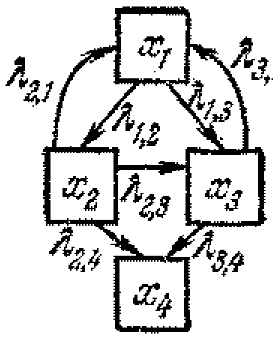

Рис. 10в.

Если имеется размеченный граф состояний системы X, то систему дифференциальных уравнений для вероятностей состояний $p_k(t)$ $(k=1, 2, \ldots, n)$ можно сразу написать, пользуясь следующим простым правилом. В левой части каждого уравнения стоит производная $\frac{dp_k(t)}{dt}$, а в правой части — столько членов, сколько стрелок связано непосредственно с данным состоянием; если стрелка ведет в данное состояние, член имеет знак плюс, если ведет из данного состояния, член имеет знак минус. Каждый член равен плотности потока событий, переводящего систему по данной стрелке, умноженной на вероятность того состояния, из которого исходит стрелка. Например, для системы X, размеченный граф состояний которой показан на рис. 10в, система дифференциальных уравнений будет:

$$\frac{dp_1(t)}{dt}=\lambda_{2,1}p_2(t)+\lambda_{3,1}p_3(t)-(\lambda_{1,2}+\lambda_{1,3})p_1(t),$$

$$\frac{dp_2(t)}{dt}=\lambda_{1,2}p_1(t)-(\lambda_{2,1}+\lambda_{2,3}+\lambda_{2,4})p_2(t),$$

$$\frac{dp_3(t)}{dt}=\lambda_{1,3}p_1(t)+\lambda_{2,3}p_2(t)-(\lambda_{3,1}+\lambda_{3,4})p_3(t),$$

$$\frac{dp_4(t)}{dt}=\lambda_{2,4}p_2(t)+\lambda_{3,4}p_3(t).$$

Число уравнений может быть уменьшено на единицу, если учесть условие: для любого t

$$p_1(t)+p_2(t)+p_3(t)+p_4(t)=1.$$

Начальные условия для интегрирования такой системы отражают состояние системы в начальный момент. Если, например, система при $t=0$ была в состоянии x_k, то полагают

$$p_k(0)=1;\ p_i(0)=0 \text{ при } i\neq k.$$

Предельным режимом для системы X называется случайный процесс, устанавливающийся в системе при $t\to\infty$.

Если в числе состояний системы имеются состояния без выхода, то при $t\to\infty$ система с практической достоверностью оказывается в одном из них.

Если все потоки событий, переводящие систему из состояния в состояние, стационарны ($\lambda_{i,j}=\text{const}$), общее число состояний конечно и состояний без выхода нет, то предельный режим существует и характеризуется предельными вероятностями состояний $p_1, p_2, \ldots, p_n$ $\left(\sum_{k=1}^{n}p_k=1\right)$. Чтобы найти эти вероятности, приравнивают нулю левые части уравнений для вероятностей состояний $\left(\text{полагают все производные } \frac{dp_k(t)}{dt} \text{ равными } 0\right)$ и решают полученную систему линейных алгебраических уравнений. К ним добавляется нормировочное условие

$$\sum_{k=1}^{n}p_k=1.$$

Например, для системы X, размеченный граф состояний которой дан на рис. 10в, система алгебраических уравнений, определяющая предельный режим, будет:

$$\lambda_{2,1} p_2 + \lambda_{3,1} p_3 - (\lambda_{1,2} + \lambda_{1,3}) p_1 = 0,$$
$$\lambda_{1,2} p_1 - (\lambda_{2,1} + \lambda_{2,3} + \lambda_{2,4}) p_2 = 0,$$
$$\lambda_{1,3} p_1 + \lambda_{2,3} p_2 - (\lambda_{3,1} + \lambda_{3,4}) p_3 = 0,$$
$$\lambda_{2,4} p_2 + \lambda_{3,4} p_3 = 0, \quad p_1 + p_2 + p_3 + p_4 = 1.$$

Потоком Пальма (потоком с ограниченным последействием) называется поток событий, у которого промежутки между соседними событиями представляют собой независимые случайные величины. Если эти случайные величины распределены одинаково, то поток Пальма называется *стационарным*.

Простейший (стационарный пуассоновский) поток является потоком Пальма.

Нестационарный пуассоновский поток потоком Пальма не является.

Потоком Эрланга k-го порядка называется поток событий, получаемый из простейшего путем операции «разрежения», когда выбрасывают из потока k точек подряд, а сохраняют только $(k+1)$-ю (рис. 10г). Простейший поток есть поток Эрланга нулевого порядка.

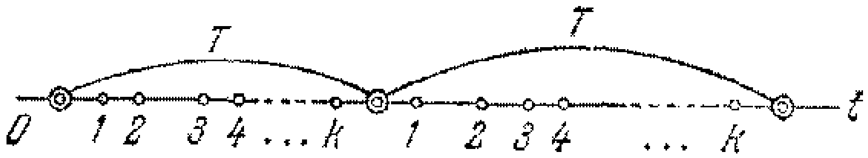

Рис. 10г.

Промежуток времени T между двумя соседними событиями в потоке Эрланга k-го порядка есть неотрицательная случайная величина с плотностью распределения

$$f_k(t) = \frac{\lambda (\lambda t)^k}{k!} e^{-\lambda t} \qquad (t > 0)$$

(закон Эрланга, см. стр. 231) и функцией распределения

$$F_k(t) = \mathrm{P}(T < t) = 1 - \sum_{s=0}^{k} \frac{(\lambda t)^s}{s!} e^{-\lambda t} \qquad (t > 0).$$

При $k = 0$ (простейший поток) получаем

$$f_0(t) = \lambda e^{-\lambda t} \qquad (t > 0)$$

(показательный закон).

Как плотность распределения $f_k(t)$, так и функцию распределения $F_k(t)$ для закона Эрланга любого порядка можно вычислять, пользуясь таблицами пуассоновского распределения:

$$P(k, a) = \frac{a^k}{k!} e^{-a},$$

В этих обозначениях

$$f_k(t)=\lambda P(k, \lambda t) \quad (t>0),$$
$$F_k(t)=1-R(k, \lambda t),$$

где

$$R(k, \lambda t)=\sum_{s=0}^{k} \frac{(\lambda t)^s}{s\,!} e^{-\lambda t}$$

— табулированная функция [см. приложение, табл. 1, где приведены значения функции

$$Q(m, a)=1-R(m, a)].$$

Функцию $P(k, a)$ можно вычислять по тем же таблицам $R(k, a)$:

$$P(k, a)=R(k, a)-R(k-1, a)=Q(k-1, a)-Q(k, a).$$

Между функциями $P(k, a)$ и $R(k, a)$ существует следующее соотношение:

$$\frac{\partial}{\partial a} R(k, a)=-P(k, a).$$

Полезно знать предельные соотношения:

$$\lim_{k\to\infty} R(k, a)=1,$$

$$\lim_{k\to\infty} P(k, a)=0.$$

Регулярным потоком событий называется поток, в котором события следуют одно за другим через строго определенные промежутки времени.

При увеличении порядка k потока Эрланга (и одновременном уменьшении масштаба по оси Ot делением на $k+1$) поток Эрланга приближается к регулярному.

Системой массового обслуживания (СМО) называется любая система, предназначенная для обслуживания какого-то потока заявок (например, ремонтная мастерская, телефонная станция, билетная касса и т. д.).

Системы массового обслуживания делятся *на системы с отказами и системы с ожиданием*.

В системе с отказами заявка, пришедшая в момент, когда все каналы обслуживания заняты, получает отказ и покидает систему.

В системе с ожиданием такая заявка не покидает систему, а становится в очередь и ждет, пока не освободится какой-нибудь канал. Время ожидания и число мест в очереди могут быть как неограниченными, так и ограниченными.

Система массового обслуживания называется *пуассоновской*, если все потоки событий, переводящие ее из состояния в состояние, являются пуассоновскими.

Ниже мы будем рассматривать только пуассоновские СМО, причем с простейшими потоками переходов.

Работа системы массового обслуживания с отказами определяется следующими параметрами:

1) число каналов n;

2) плотность потока заявок λ;

3) плотность «потока обслуживаний» одного канала μ (плотность потока заявок, обслуживаемых одним непрерывно занятым каналом).

Величина μ обратна среднему времени обслуживания одной заявки:

$$\mu = \frac{1}{\overline{t}_{об}},$$

где

$$\overline{t}_{об} = \mathrm{M}\,[T_{об}];$$

$T_{об}$ — случайное время обслуживания.

На рис. 10д показан размеченный граф состояний n-канальной СМО с отказами. Состояние x_k $(0 \leqslant k \leqslant n)$ состоит в том, что занято

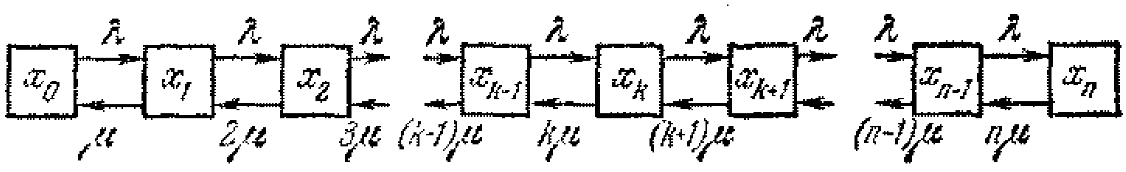

Рис. 10д.

ровно k каналов из n *). Из этого графа следуют дифференциальные уравнения для вероятностей состояний (*уравнения Эрланга*)

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t),$$

$$\frac{dp_1(t)}{dt} = -(\lambda+\mu)\,p_1(t) + \lambda p_0(t) + 2\mu p_2(t),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{dp_k(t)}{dt} = -(\lambda + k\mu)\,p_k(t) + \lambda p_{k-1}(t) + (k+1)\,\mu p_{k+1}(t),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{dp_n(t)}{dt} = -n\mu p_n(t) + \lambda p_{n-1}(t).$$

Эту систему обычно интегрируют при начальных условиях

$$p_0(0) = 1; \quad p_k(0) = 0 \qquad (k > 0)$$

(в начальный момент все каналы свободны). При $t \to \infty$ существует предельный (установившийся) режим работы СМО, при котором вероят-

*) Предполагается, что каждый канал может обслуживать только одну заявку, а каждая заявка обслуживается только одним каналом.

ности состояний определяются *формулами Эрланга*

$$p_k=\frac{\dfrac{\alpha^k}{k!}}{\sum\limits_{k=0}^{n}\dfrac{\alpha^k}{k!}}\qquad (k=0,\ 1,\dots,\ n),\tag{10.1}$$

где $\alpha=\dfrac{\lambda}{\mu}$.

Вероятности p_k могут быть вычислены с помощью таблиц пуассоновского распределения (см. приложение, *табл.* 1):

$$p_k=\frac{P(k,\alpha)}{R(n,\ \alpha)}=\frac{R(k,\alpha)-R(k-1,\alpha)}{R(n,\ \alpha)}\qquad (k=0,\ 1,\dots,\ n).\tag{10.2}$$

Вероятность того, что заявка будет обслужена (не получит отказа) выражается формулой

$$P_{\text{обс}}=1-p_n=1-\frac{P(n,\alpha)}{R(n,\alpha)}=\frac{R(n-1,\alpha)}{R(n,\alpha)}.\tag{10.3}$$

Система массового обслуживания называется *чистой системой с ожиданием*, если ни время пребывания заявки в очереди, ни число заявок в очереди ничем не ограничено. Если имеются ограничения по какому-нибудь из этих признаков, система называется *системой смешанного типа*. Для системы массового обслуживания смешанного типа с ограничениями по числу мест в очереди предельные вероятности состояний выражаются формулами

$$\left.\begin{aligned}
p_k&=\frac{\dfrac{\alpha^k}{k!}}{\sum\limits_{k=0}^{n}\dfrac{\alpha^k}{k!}+\dfrac{\alpha^{n+1}}{n\cdot n!}\,\dfrac{1-\left(\dfrac{\alpha}{n}\right)^m}{1-\dfrac{\alpha}{n}}}\qquad (k=0,\ 1,\ \dots,\ n),\\
p_{n+s}&=\frac{\dfrac{\alpha^n}{n!}\left(\dfrac{\alpha}{n}\right)^s}{\sum\limits_{k=0}^{n}\dfrac{\alpha^k}{k!}+\dfrac{\alpha^{n+1}}{n\cdot n!}\cdot\dfrac{1-\left(\dfrac{\alpha}{n}\right)^m}{1-\dfrac{\alpha}{n}}}\qquad (s=1,2,\ \dots,\ m);
\end{aligned}\right\}\tag{10.4}$$

где n — число каналов обслуживания;

m — число мест в очереди; $\alpha=\dfrac{\lambda}{\mu}$;

λ — плотность потока заявок;

μ — плотность «потока обслуживаний» одного канала.

Для чистой системы с ожиданием ($m = \infty$) установившийся предельный режим существует только в случае $\frac{\alpha}{n} < 1$. Предельные вероятности выражаются формулами.

$$p_k = \frac{\frac{\alpha^k}{k!}}{\sum_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^{n+1}}{n \cdot n! \left(1 - \frac{\alpha}{n}\right)}} \qquad (k = 0, 1, \ldots, n),$$

$$p_{n+s} = \frac{\frac{\alpha^n}{n!} \left(\frac{\alpha}{n}\right)^s}{\sum_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^{n+1}}{n \cdot n! \left(1 - \frac{\alpha}{n}\right)}} \qquad (s = 1, 2, \ldots).$$

Ограничения по времени пребывания заявки в очереди (или в системе) при составлении уравнений для вероятностей состояний учитываются тем, что на каждую заявку, находящуюся в очереди (системе), действует «поток уходов» с плотностью ν, обратной среднему времени пребывания заявки в очереди (системе).

10.1. Поток машин, следующих по шоссе в одном направлении, представляет собой простейший поток с плотностью λ. Человек выходит на шоссе, чтобы остановить первую попавшуюся машину, идущую в данном направлении. Найти закон распределения времени T, которое ему придется ждать; определить его математическое ожидание m_t и среднее квадратическое отклонение σ_t.

Решение. Плотность распределения времени ожидания будет такая же, как плотность распределения промежутка между машинами, а именно

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

так как «будущее» в простейшем потоке никак не зависит от «прошлого», в частности от того, сколько времени тому назад прошла последняя машина.

Для показательного закона

$$m_t = \frac{1}{\lambda};$$

$$D_t = \frac{1}{\lambda^2};$$

$$\sigma_t = \frac{1}{\lambda} = m_t.$$

10.2. Тот же вопрос, что и в задаче 10.1, но поток машин — регулярный, с той же плотностью λ.

Решение. Закон распределения времени ожидания T будет законом постоянной плотности в промежутке времени между двумя машинами, равном $\frac{1}{\lambda}$:

$$f(t)=\lambda \qquad \left(0<t<\frac{1}{\lambda}\right).$$

Для закона постоянной плотности

$$m_t=\frac{1}{2\lambda};$$

$$D_t=\frac{\left(\frac{1}{\lambda}\right)^2}{12}=\frac{1}{12\lambda^2};$$

$$\sigma_t=\frac{1}{2\sqrt{3}\,\lambda}.$$

10.3. Показать, что для пуассоновского потока событий

$$\lim_{\Delta t\to 0}\frac{\mathsf{P}(X(\Delta t)\geqslant 1)}{\mathsf{P}(X(\Delta t)=1)}=1,$$

где $X(\Delta t)$ — число событий, попадающих на участок длиной Δt.

Решение. Имеем

$$\mathsf{P}(X(\Delta t)\geqslant 1)=1-\mathsf{P}(X(\Delta t)=0)=1-e^{-\lambda\Delta t};$$

$$\mathsf{P}(X(\Delta t)=1)=\lambda\Delta t e^{-\lambda\Delta t}.$$

Следовательно,

$$\lim_{\Delta t\to 0}\frac{\mathsf{P}(X(\Delta t)\geqslant 1)}{\mathsf{P}(X(\Delta t)=1)}=\lim_{\Delta t\to 0}\frac{1-e^{-\lambda\Delta t}}{\lambda\Delta t e^{-\lambda\Delta t}}=1.$$

10.4.* Пассажир выходит на автобусную остановку и ждет очередного автобуса. Автобусы подходят к остановке через случайные, взаимонезависимые и одинаково распределенные промежутки времени T_1, T_2, ... ($T_i>0$). Каждый из этих промежутков времени имеет одну и ту же плотность распределения $f(t)$. Требуется найти закон распределения времени ожидания очередного автобуса при условии, что выход пассажира на остановку некоррелирован с моментом прибытия автобуса (расписание движения автобусов пассажиру неизвестно).

Решение. Рассмотрим поток событий, состоящих в том, что на остановку прибывает автобус. Этот поток по условиям задачи будет стационарным потоком Пальма. Выход пассажира на автобусную остановку можно рассматривать как появление некоторой точки П на оси времени Ot.

Случайность выхода пассажира на остановку следует понимать в том смысле, что в интервале времени T^* между прибытием двух автобусов (рис. 10.4а) точка П распределена равномерно (подчеркнем, что речь идет об очередном автобусе и ему предшествующем).

Закон распределения интервала времени T^* между прибытием двух автобусов, на котором появился пассажир (на

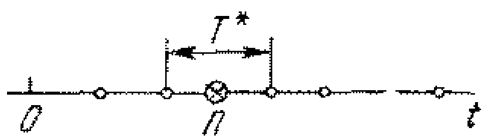

Рис. 10.4а.

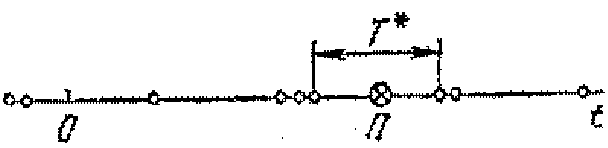

Рис. 10.4б.

который упала точка П), в общем случае *не совпадает* с законом $f(t)$. Этот (на первый взгляд парадоксальный) факт можно пояснить на следующем наглядном примере. Допустим, что интервал времени T (в часах) между появлениями двух соседних по времени автобусов может принимать только два значения: $t_1 = 0{,}9$ с вероятностью 0,5 и $t_2 = 0{,}1$ с вероятностью 0,5. Тогда на оси Ot мы будем иметь поток Пальма, в котором с одинаковой частотой будут встречаться длинные (0,9) и короткие (0,1) участки (см. рис. 10.4б). Пусть пассажир появился случайно в какой-нибудь момент на оси Ot. Спросим себя, что более вероятно: что он попадет на участок длины 0,9 или на участок длины 0,1? Очевидно, первое более вероятно: отрезков 0,9 и 0,1 на оси Ot в среднем одинаковое количество, но отрезки 0,9 длиннее в 9 раз; значит, они занимают в 9 раз большую протяженность оси Ot, чем малые отрезки, а следовательно, вероятность попадания точки П на отрезок 0,9 равна уже не 0,5, а 0,9, а вероятность попадания на отрезок 0,1 равна 0,1.

Таким образом, на этом простом примере можно убедиться, что закон распределения того промежутка, на который попала точка П, не совпадает с его априорным законом распределения.