АКАДЕМИЯ НАУК
СОЮЗА СОВЕТСКИХ СОЦИАЛИСТИЧЕСКИХ РЕСПУБЛИК

ТРУДЫ
МАТЕМАТИЧЕСКОГО ИНСТИТУТА
имени В. А. СТЕКЛОВА

XLIX

А. Я. ХИНЧИН

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР
МОСКВА 1955

АКАДЕМИЯ НАУК
СОЮЗА СОВЕТСКИХ СОЦИАЛИСТИЧЕСКИХ РЕСПУБЛИК

ТРУДЫ
МАТЕМАТИЧЕСКОГО ИНСТИТУТА
имени В. А. СТЕКЛОВА

XLIX

А. Я. ХИНЧИН

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР
МОСКВА 1955

Ответственный редактор
академик И. Г. Петровский

Зам. отв. редактора
профессор С. М. Никольский

ПРЕДИСЛОВИЕ

В науке, практической деятельности человечества и в быту каждодневно создаются такие положения, когда возникает массовый спрос на обслуживание какого-либо специального вида, причем обслуживающая организация, располагая лишь ограниченным числом обслуживающих единиц, не всегда способна немедленно удовлетворять все поступающие заявки. Примеры такой ситуации хорошо известны каждому. Очереди у магазинных и билетных касс, в буфетах, парикмахерских и т. д.; невозможность получить билет на нужный поезд из-за переполнения; задержка в посадке самолетов, вызываемая отсутствием свободных посадочных площадок; задержка в ремонте потерпевших аварию станков из-за нехватки ремонтных бригад — все эти и многие другие аналогичные, хорошо известные примеры, несмотря на существенные различия их реального содержания, с формальной стороны очень близки друг другу. Во всех подобных случаях перед теорией встает в сущности одна основная задача: установить с возможной точностью взаимную зависимость между числом обслуживающих единиц и качеством обслуживания. При этом качество обслуживания в различных случаях, естественно, измеряется различными показателями. По большей части таким показателем служит либо процент заявок, получающих отказ (процент пассажиров, не получивших билетов на данный поезд), либо среднее время ожидания начала обслуживания (очереди различного рода). Разумеется, качество обслуживания во всех случаях тем выше, чем больше число обслуживающих единиц; однако столь же очевидно, что чрезмерный рост этого числа сопряжен с излишним расходом сил и материальных средств; практически поэтому вопрос обычно ставится так, что сначала устанавливается необходимый уровень качества обслуживания, а затем находится минимальное число обслуживающих единиц, при котором этот уровень может быть достигнут.

В задачах подобного рода почти всегда приходится учитывать влияние случайного элемента на течение изучаемого явления. Количество поступающих заявок не является, как правило, постоянным, а испытывает случайные колебания. Время обслуживания заявок в большинстве задач не является стандартным, а подвержено случайным колебаниям от одной заявки к другой. Все эти элементы случайности отнюдь не имеют характера небольших «возмущений», нарушающих собой плавный и закономерный ход явления; напротив, они составляют собой основную черту в картине изучаемых процессов. Естественно поэтому, что математическим инструментом теории массового обслуживания должны стать понятия и методы теории вероятностей — математической дисциплины, посвященной изучению закономерностей случая.

Цель настоящей книги — ознакомить читателя с основными идеями, методами и отдельными ходами мысли, господствующими в приложениях теории вероятностей к вопросам массового обслуживания. Необходимость создания монографии подобного рода уже давно ощущается как математиками, так и практическими работниками (в первую очередь связистами, в свое время вызвавшими к жизни изучаемую теорию и до сих пор остающимися ее главными потребителями). Эта потребность усиливается еще тем обстоятельством, что сколько-нибудь доступных изложений общей теории не имеется и за границей.

Само собой разумеется, что предлагаемая монография ни в какой мере не может претендовать на полноту сообщаемых сведений. Теория массового обслуживания в настоящее время разрослась очень широко, и для сколько-нибудь полного изложения хотя бы только важнейших из ее достижений понадобился бы толстый том. Однако такой цели я себе и не ставил. Моей задачей было — осветить на небольшом числе важнейших примеров общий характер, основной стиль использования вероятностных рассуждений в вопросах массового обслуживания. Исходя из этой «методологической» целевой установки, я и производил отбор материала для моей монографии. Особое внимание было при этом уделено трудам основоположника теории А. К. Эрланга, далеко не все исследования которого известны у нас в той степени, какой они заслуживают. В значительном количестве мною введены в книгу также идеи К. Пальма —крупнейшего современного продолжателя дела Эрланга. В известной мере я учел и пробивающееся все чаще в наши дни стремление искать более элементарных методов исследования в противовес почти безраздельно господствующим со времени классических трудов Эрланга аналитическим методам, связанным с составлением систем дифференциальных, разностно-дифференциальных и интегральных уравнений. В этом направлении я обращаю внимание читателя на § 2 и 25, а также на всю главу 11. Попытка создания общей элементарной теории предпринята недавно в интересной статье Лундквиста (Ericsson Technics, 1953).

Составление этой монографии было очень сильно затруднено тем обстоятельством, что вся основная литература принадлежит перу практических специалистов и в математическом отношении является неудовлетворительной. Чтобы придать всему изложению форму, сколько-нибудь приемлемую в математическом смысле, я не мог оставить почти ни одного рассуждения в его первоначальном виде; приходилось либо существенно дополнять приводимую автором аргументацию, либо выбрасывать ее и заменять другой. В равной степени и вводимые новые понятия во многих случаях пришлось определять по-иному, так как определения, даваемые авторами, оказывались недостаточно четкими.

Чтобы помочь читателю связывать общие понятия с конкретными представлениями, я на протяжении всей книги пользуюсь терминологией, заимствованной из телефонной практики, запросы которой до сих пор остаются важнейшим стимулом развития теории массового обслуживания. Так, я говорю о «вызовах» (вместо «требований» или «заявок»), о «потерях» (вместо «отказов»), о длительности «разговора» (вместо «обслуживания») и т. п. Однако во всех случаях все сказанное относится, разумеется, к любому виду массового обслуживания при соответствующем изменении терминологии.

Книга разделена на три части, из которых первая — самая большая — посвящена изучению «потока» поступающих вызовов и вовсе не содержит вопросов обслуживания (которым посвящены две остальные части). Такое отчетливое выделение в особую часть изучения входящего потока казалось мне целесообразным не только потому, что для правильного обслуживания надо прежде всего хорошо знать то, что обслуживаешь, хорошо знать все черты чередования поступающих заявок; другой, не менее важный повод к такому выделению состоял в том, что теория входящего потока вызовов есть не что иное, как общая теория случайного чередования однородных событий, которая находит себе широкий круг применений и в областях, ни с каким обслуживанием не связанных (например, в вопросах радиоактивного распада атомов).

Я стремился сделать книгу доступной всякому, кто владеет основными понятиями теории вероятностей и хотя бы кратким (втузовским) курсом математического анализа. Во многих случаях изложение могло бы быть сокращено, если бы я, вместо проведения полного доказательства того или иного утверждения, позволил себе сослаться на общие результаты соответствующей вероятностной теории (цепи Маркова, случайные процессы, эргодическая теория). Однако, желая сделать книгу более доступной возможно более широкому кругу читателей, я почти нигде не поддавался такого рода искушению.

В теории вероятностей принято называть законом распределения данной случайной величины ξ (неубывающую) функцию $F(x)$, выражающую собой вероятность неравенства $\xi < x$. В теории массового обслуживания часто оказывается более удобным называть законом распределения величины ξ (невозрастающую) функцию $\Phi(x)$, выражающую собой вероятность неравенства $\xi > x$. Мы будем, в зависимости от обстоятельств, пользоваться тем или другим пониманием закона распределения.

А. Хинчин

2 октября 1954 г.

ЧАСТЬ I

ВХОДЯЩИЙ ПОТОК ВЫЗОВОВ

Глава 1

Теория простейшего потока

Общая теория потоков однородных событий должна естественно начинаться с определения основных общих понятий, связанных с такими потоками. Однако мы отложим такой общий подход до главы 2, где он будет проведен в нужной широте. Мы предпочитаем сразу ввести читателя в круг конкретных исследований, связанных с потоками некоторого простейшего типа, чтобы тем самым с первых страниц дать ему наглядное представление об основных ходах мысли и математических орудиях теории массового обслуживания, о стиле этого учения как математической дисциплины. После того как эти конкретные представления будут в достаточной мере усвоены, изучение более абстрактной общей теории уже не должно будет показаться трудным.

Добавим к этому, что тот простейший тип потока, который мы будем изучать в настоящей главе, в течение долгого времени оставался почти единственным, употреблявшимся в приложениях; лишь в сравнительно недавнее время отчетливо выяснилась необходимость изучения потоков более общего типа; впрочем и в наши дни значительное большинство приложений теории массового обслуживания (в частности, и приложений к телефонному делу) исходит еще из предпосылки, что поступающий поток требований (вызовов) принадлежит простейшему типу. Приложения (особенно технические) теории простейшего потока за последние десятилетия настолько расширились, что в настоящее время даже элементарные курсы теории вероятностей, как правило, включают в свою программу специальные главы, посвященные этой теории.

§ 1. Определение и постановка задачи

Простейшим мы будем называть поток однородных событий, если он обладает следующими тремя свойствами.

1°. Стационарность. Каковы бы ни были $t>0$ и целое $k\geqslant 0$, вероятность того, что за промежуток времени $(a, a+t)$ произойдет k событий, одна и та же для всех $a\geqslant 0$ (и, значит, зависит только от k и t); мы будем во всем дальнейшем обозначать эту вероятность через $v_k(t)$. На протяжении всей книги мы будем иметь дело только с такими потоками, в которых за конечный промежуток времени с вероятностью 1 происходит лишь конечное число событий. Мы будем, таким образом, всегда иметь $\sum_{k=0}^{\infty} v_k(t) = 1$ при любом t. Стационарность потока выражает собой неизменность его вероятностного режима во времени.

2°. Отсутствие последействия. Вероятность $v_k(t)$ наступления k событий за промежуток времени $(a, a+t)$ не зависит от чередования событий до момента a; другими словами, условная вероятность наступления k событий за промежуток времени $(a, a+t)$, вычисленная при любом предположении о чередовании событий до момента a, равна безусловной вероятности $v_k(t)$ того же события. Отсутствие последействия выражает собой взаимную независимость протеканий потока в непересекающихся между собой промежутках времени.

3°. Ординарность. Пусть для данного стационарного потока $\psi(t)$ означает вероятность того, что за (где угодно расположенный) промежуток времени длины t наступит по меньшей мере два события $\left[\text{очевидно, } \psi(t) = 1 - v_0(t) - v_1(t) = \sum_{k=2}^{\infty} v_k(t)\right]$. Тогда мы имеем

$$\psi(t) = o(t) \quad (t\to 0),$$

или, что то же,

$$\frac{\psi(t)}{t} \to 0 \quad (t\to 0).$$

Как мы увидим далее, ординарность потока выражает собой практическую невозможность совмещения двух или более событий в один и тот же момент времени.

Итак, ***простейшим потоком однородных событий мы называем всякий стационарный ординарный поток без последействия.***

Основная задача теории простейшего потока состоит в определении вида функций $v_k(t)$; иначе говоря, целью нашей будет отыскание закона распределения числа событий за промежуток времени длины t, рассматриваемого как случайная величина. При этом мы ради определенности реальной интерпретации изучаемого потока будем во всем дальнейшем предполагать, что речь идет о потоке вызовов, поступающих на некоторую телефонную установку, и в соответствии с этим называть наши однородные события «вызовами».

§ 2. Элементарное решение

Разобьем промежуток времени (0, 1) на произвольное число n равных промежутков длины $1/n$. Вероятность того, что в какой-либо из этих частей не произойдет ни одного вызова, равна $v_0(1/n)$; а так как наш поток — без последействия, то

$$v_0(1) = \left[v_0\left(\frac{1}{n}\right)\right]^n,$$

или, полагая $v_0(1) = \theta$,

$$v_0\left(\frac{1}{n}\right) = \theta^{\frac{1}{n}}.$$

Если мы имеем промежуток длины k/n $(k = 1, 2, \ldots)$, то он разбивается на k промежутков длины $\frac{1}{n}$, вследствие чего

$$v_0\left(\frac{k}{n}\right) = \left[v_0\left(\frac{1}{n}\right)\right]^k = \theta^{k/n}. \tag{2.1}$$

Пусть, наконец, t — любое положительное число и пусть натуральное число k определяется из неравенств

$$\frac{k-1}{n} < t \leqslant \frac{k}{n};$$

так как $v_0(t)$, очевидно, есть невозрастающая функция от t, то

$$v_0\left(\frac{k-1}{n}\right) \geqslant v_0(t) \geqslant v_0\left(\frac{k}{n}\right),$$

или, в силу (2.1),

$$\theta^{\frac{k-1}{n}} \geqslant v_0(t) \geqslant \theta^{\frac{k}{n}};$$

но при $n \to \infty$ мы имеем $\frac{k}{n} \to t$, вследствие чего крайние члены написанных неравенств стремятся к θ^t, и мы находим

$$v_0(t) = \theta^t$$

для любого $t > 0$. При этом постоянное число θ нами определено как $v_0(1)$, и следовательно, $0 \leqslant \theta \leqslant 1$. Однако случаи $\theta = 0$ и $\theta = 1$ не представляют интереса, и мы можем их не рассматривать. В самом деле, при $\theta = 1$ мы имеем $v_0(t) = 1$ при любом $t > 0$, что означает достоверное отсутствие вызовов в любом промежутке времени, т. е. отсутствие какого бы то ни было потока. Если же $\theta = 0$, то $v_0(t) = 0$ при любом $t > 0$; это означает, что вызовы с достоверностью будут получены в любом, сколь угодно малом промежутке времени; но тогда, сколь бы велико ни было k, число вызовов в любом промежутке с достоверностью будет больше, чем k; другими словами, число вызовов в любом промежутке бесконечно с вероятностью 1; но такие потоки мы в § 1 раз навсегда исключили из рассмотрения. Итак, мы можем считать, что $0<\theta<1$; поэтому можно положить $\theta=e^{-\lambda}$, где λ — постоянное положительное число, и писать

$$v_0(t)=e^{-\lambda t}. \tag{2.2}$$

Отметим, что при этом выводе мы нигде не пользовались ординарностью нашего потока, так что соотношение (2.2) имеет силу для любого стационарного потока без последействия; это замечание будет для нас важно в дальнейшем.

Теперь обращаемся к отысканию функций $v_k(t)$ при $k>0$. Существует много различных способов решения этой задачи, и почти все они поучительны, так как заложенные в них методы позволяют решать и ряд более сложных задач. Мы начнем с наиболее элементарного способа. Будем считать число t постоянным и разобьем промежуток $(0, t)$ на произвольное число $n>k$ равных частей (ячеек) длины $t/n=\delta$. Относительно расположения вызовов в этих ячейках возможны две гипотезы:

H_1 — ни в одной из n ячеек не будет более одного вызова;

H_2 — по крайней мере в одной из ячеек произойдет более одного вызова.

Мы, очевидно, имеем

$$v_k(t)=\mathsf{P}(H_1, k)+\mathsf{P}(H_2, k), \tag{2.3}$$

где $\mathsf{P}(H_i, k)$ $(i=1,2)$ означает вероятность двойного события: 1) реализуется гипотеза H_i и 2) в промежутке $(0, t)$ поступает k вызовов. Очевидно, что $\mathsf{P}\{H_1, k\}$ есть вероятность такого положения вещей, когда из наших n ячеек какие-либо k содержат по одному вызову, а остальные $n-k$ вообще вызовов не содержат, поэтому

$$\mathsf{P}(H_1, k)=\binom{n}{k}[v_1(\delta)]^k[v_0(\delta)]^{n-k}.$$

В силу формулы (2.2) и ординарности данного потока мы имеем при $n\to\infty$ $(\delta\to 0)$ и постоянном k

$$[v_0(\delta)]^{n-k}=e^{-\lambda\delta(n-k)}=e^{-\lambda t}e^{k\lambda\delta}=e^{-\lambda t}[1+o(1)];$$
$$[v_1(\delta)]^k=[1-e^{-\lambda\delta}-\psi(\delta)]^k=[1-e^{-\lambda\delta}+o(\delta)]^k=$$
$$=(\lambda\delta)^k[1+o(1)]=\frac{(\lambda t)^k}{n^k}[1+o(1)],$$

и, следовательно,

$$\mathsf{P}(H_1, k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}[1+o(1)]\to e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$
$$(n\to\infty).$$

С другой стороны, $\mathsf{P}(H_2, k)$, очевидно, не превосходит вероятности гипотезы H_2, т. е. того, что по меньшей мере одна из n ячеек содержит более одного вызова; так как для отдельной ячейки вероятность содержать более одного вызова есть $\psi(\delta)$, то поэтому

$$\mathsf{P}(H_2, k) \leqslant n\psi(\delta) = t\frac{\psi(\delta)}{\delta} \to 0 \quad (n \to \infty);$$

таким образом, правая часть равенства (2.3) при $n \to \infty$ имеет пределом

$$e^{-\lambda t}\frac{(\lambda t)^k}{k!};$$

а так как левая часть (2.3) от n не зависит, то

$$v_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \quad (k = 0, 1, \ldots).$$

Таким образом, для простейшего потока число вызовов в промежутке длины t распределено по закону Пуассона с параметром λt.

§ 3. Метод дифференциальных уравнений

В специальной литературе поставленная нами задача решается обычно другим методом, менее элементарным, но зато легко распространяемым на более сложные задачи. Рассмотрим теперь этот метод. Пусть t и τ — любые положительные числа и пусть $k > 0$. Пусть k_1 — число вызовов в промежутке $(0, t)$, а k_2 — в промежутке $(t, t+\tau)$. Для того, чтобы в промежутке $(0, t+\tau)$ произошло k вызовов, необходимо и достаточно наступление одного из следующих двойных событий: $k_1 = k$, $k_2 = 0$; $k_1 = k-1$, $k_2 = 1$; $k_1 = k-2$, $k_2 = 2; \ldots$; $k_1 = 0$, $k_2 = k$. Но вероятности событий $k_1 = l$ и $k_2 = m$ соответственно равны $v_l(t)$ и $v_m(\tau)$; а так как эти события взаимно независимы (поток без последействия!), то

$$\begin{aligned} v_k(t+\tau) = v_k(t)\,v_0(\tau) + v_{k-1}(t)\,v_1(\tau) &+ \\ + v_{k-2}(t)\,v_2(\tau) + \cdots &+ v_0(t)\,v_k(\tau). \end{aligned} \tag{3.1}$$

Но при $\tau \to 0$ мы имеем:

$$\begin{gathered} v_0(\tau) = e^{-\lambda\tau} = 1 - \lambda\tau + o(\tau); \\ v_1(\tau) = 1 - v_0(\tau) - \psi(\tau) = \lambda\tau + o(\tau); \\ \sum_{l=2}^{k} v_{k-l}(t)\,v_l(\tau) \leqslant \sum_{l=2}^{\infty} v_l(\tau) = \psi(\tau) = o(\tau), \end{gathered}$$

и, следовательно, (3.1) дает

$$\begin{gathered} v_k(t+\tau) = v_k(t)(1-\lambda\tau) + v_{k-1}(t)\lambda\tau + o(\tau); \\ \frac{v_k(t+\tau) - v_k(t)}{\tau} = \lambda[v_{k-1}(t) - v_k(t)] + o(1). \end{gathered}$$

Это показывает, что функция $v_k(t)$ дифференцируема при любом $t > 0$ и что

$$v_k'(t) = \lambda[v_{k-1}(t) - v_k(t)] \quad (k = 1, 2, \ldots); \tag{3.2}$$

если положить для общности $v_{-1}(t) \equiv 0$, то уравнение (3.2), как мы непосредственно убеждаемся из (2.2), имеет место и при $k = 0$.

Таким образом, для определения искомых функций $v_k(t)$ мы получили систему линейных дифференциальных уравнений (3.2). Эта система легко решается различными методами, из которых мы рассмотрим два наиболее поучительных для дальнейшего.

А. Метод замены искомых функций

Положим

$$v_k(t) = e^{-\lambda t} u_k(t) \quad (k = 0, 1, 2, \ldots),$$

отсюда

$$v_k'(t) = e^{-\lambda t} [u_k'(t) - \lambda u_k(t)] \quad (k = 0, 1, 2, \ldots);$$

вставляя же эти выражения $v_k(t)$ и $v_k'(t)$ в уравнения (3.2), легко находим

$$u_k'(t) = \lambda u_{k-1}(t) \quad (k = 0, 1, 2, \ldots),$$

где по определению $u_{-1}(t) \equiv 0$. Отсюда, интегрируя, находим

$$u_k(t) - u_k(0) = \lambda \int_0^t u_{k-1}(z)\, dz \quad (k = 0, 1, 2, \ldots).$$

Очевидно, мы имеем при любом $k \geqslant 0$

$$v_k(0) = u_k(0);$$

но по определению функций $v_k(t)$

$$v_0(0) = 1, \quad v_k(0) = 0 \quad (k > 0);$$

поэтому и $u_0(0) = 1$, $u_k(0) = 0$ $(k = 1, 2, \ldots)$, и мы находим при $k \geqslant 1$

$$u_k(t) = \lambda \int_0^t u_{k-1}(z)\, dz. \tag{3.3}$$

Замечая, что в силу (2.2) мы имеем по определению функций $u_k(t)$

$$u_0(t) \equiv 1,$$

мы по формуле (3.3) рекуррентно находим

$$u_1(t) = \lambda t;$$
$$u_2(t) = \frac{(\lambda t)^2}{2};$$
$$\ldots\ldots\ldots;$$
$$u_k(t) = \frac{(\lambda t)^k}{k!};$$
$$\ldots\ldots\ldots,$$

и, следовательно,

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

т. е. получаем прежнее решение задачи.

В. Метод производящих функций

Положим

$$\sum_{k=0}^{\infty} v_k(t)\, x^k = \Phi(t, x);$$

ряд в левой части этого равенства во всяком случае абсолютно сходится при $|x| \leqslant 1$. Умножая на x^k все члены уравнения (3.2) и суммируя по k от 0 до ∞, мы легко находим

$$\frac{\partial \Phi}{\partial t} = \lambda \sum_{k=0}^{\infty} v_{k-1}(t)\, x^k - \lambda\Phi = \lambda(x-1)\,\Phi,$$

или

$$\frac{\partial \ln \Phi}{\partial t} = \lambda(x-1);$$

отсюда

$$\ln \Phi(t, x) - \ln \Phi(0, x) = \lambda(x-1)\,t; \tag{3.4}$$

но легко видеть, что при любом x

$$\Phi(0, x) = v_0(0) = 1,$$

поэтому (3.4) дает

$$\Phi(t, x) = e^{\lambda(x-1)t} = e^{-\lambda t} e^{\lambda t x} = e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}x^k.$$

Сопоставляя это с определением функции $\Phi(t, x)$, мы непосредственно видим, что

$$v_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \quad (k = 0, 1, 2, \ldots),$$

т. е. снова приходим к прежнему решению задачи.

§ 4. Интенсивность простейшего потока

Полученные нами результаты показывают, что те три свойства (стационарность, отсутствие последействия, ординарность), которыми мы определили простейший поток, полностью характеризуют его структуру с точностью до значения параметра λ, которое может быть любым положительным числом. Два простейших потока могут отличаться друг от друга только значениями этого параметра.

Условимся обозначать в дальнейшем для любого стационарного потока через $w(t)$ вероятность того, что за промежуток времени t произойдет по меньшей мере один вызов. Очевидно, мы имеем

$$w(t) = 1 - v_0(t) = \sum_{k=1}^{\infty} v_k(t) = v_1(t) + \psi(t),$$

где $\psi(t) = \sum_{k=2}^{\infty} v_k(t)$ попрежнему означает вероятность поступления по меньшей мере двух вызовов за промежуток времени длины t. Для простейшего потока с параметром λ $v_0(t) = e^{-\lambda t}$ и, значит, при $t \to 0$

$$w(t) = 1 - e^{-\lambda t} = \lambda t + o(t),$$

или, что то же,

$$\lim_{t \to 0} \frac{w(t)}{t} = \lambda. \tag{4.1}$$

Мы можем считать это соотношение *определением* параметра λ для данного потока. Мы узнаем в дальнейшем, что предел (4.1) существует у любого стационарного потока и определенный соотношением (4.1) параметр λ служит одной из важнейших характеристик этого потока.

Но вернемся к простейшему потоку и найдем теперь математическое ожидание числа вызовов, поступающих за промежуток времени длины t. Оно равно

$$\sum_{k=1}^{\infty} k v_k(t) = e^{-\lambda t} \sum_{k=1}^{\infty} \frac{(\lambda t)^k}{(k-1)!} = e^{-\lambda t} \lambda t \sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} = \lambda t,$$

так как последняя сумма, очевидно, равна $e^{\lambda t}$. Мы могли бы предвидеть этот результат и заранее, так как известно, что математическое ожидание величины, распределенной по закону Пуассона, равно параметру этого закона, т. е. в данном случае равно λt.

Математическое ожидание числа вызовов в единицу времени называют **интенсивностью** данного потока; мы будем обозначать эту интенсивность через μ. Как мы только что установили, для простейшего потока $\mu = \lambda$. Однако для стационарных потоков более сложной структуры это равенство не только не очевидно, но и не всегда верно; в этом вопросе мы подробно разберемся в дальнейшем. Сейчас же убедимся только, что для любого стационарного потока $\mu \geqslant \lambda$. В самом деле, математическое ожидание числа вызовов за время t для данного потока равно

$$\mu t = \sum_{k=1}^{\infty} k v_k(t) \geqslant \sum_{k=1}^{\infty} v_k(t) = w(t),$$

$$\mu \geqslant \frac{w(t)}{t},$$

а так как левая часть этого неравенства от t не зависит, то в силу (4.1) $\mu \geqslant \lambda$; разумеется, самое существование предела (4.1) для любого стационарного потока еще должно быть доказано.

Итак, *интенсивность* μ *простейшего потока совпадает с его параметром* λ; *для произвольного же стационарного потока мы пока можем утверждать только, что* $\mu \geqslant \lambda$. При этом существование предела (4.1) для любого стационарного потока еще должно быть доказано*.

§ 5. Поток с переменным параметром

В этой книге мы будем изучать почти исключительно стационарные потоки вызовов. Однако для некоторых простейших задач решение в нестационарном случае настолько легко проводится и вместе с тем имеет столь ясное практическое значение, что было бы жаль оставить его совсем без рассмотрения. В частности, в настоящем параграфе мы подвергнем изучению потоки, не обладающие стационарностью, не являющиеся, подобно простейшему потоку, ординарными потоками без последействия. Мы сейчас более точно поясним смысл этих предпосылок.

Если поток не стационарен, то вероятность получить k вызовов за промежуток времени длины τ зависит не только от τ, но и от начального момента t этого промежутка: поэтому мы будем обозначать ее через $v_k(\tau, t)$. Таким образом, $v_k(\tau, t)$ есть вероятность того, что за промежуток времени $(t, t+\tau)$ произойдет k вызовов. По аналогии со стационарным случаем мы полагаем

$$1 - v_0(\tau, t) = w(\tau, t); \; 1 - v_0(\tau, t) - v_1(\tau, t) = \psi(\tau, t).$$

Мы будем называть исследуемый поток ординарным, если при $\tau \to 0$ и любом постоянном $t \geqslant 0$ имеет место соотношение

$$\frac{\psi(\tau, t)}{\tau} \to 0.$$

Далее мы должны допустить, что *для любого* $t \geqslant 0$ *существует*

$$\lim_{\tau \to 0} \frac{w(\tau, t)}{\tau} = \lambda(t) \tag{5.1}$$

(мгновенное значение параметра).

---

* В сущности и параметр λ может быть реально интерпретирован как «интенсивность» данного потока, так как соотношение (4.1) показывает, что вероятность поступления вызовов в промежутке бесконечно малой длины t асимптотически пропорциональна t, и коэффициентом пропорциональности служит как раз параметр λ. Можно было бы говорить о «верхней интенсивности» μ и «нижней интенсивности» λ (так как всегда $\mu \geqslant \lambda$). В дальнейшем (§ 11) мы узнаем, что для ординарного потока всегда $\mu = \lambda$.

Исходя из этих предпосылок, мы поставим себе задачей найти выражение функций $v_k(\tau, t)$. Рассмотрим, как и прежде, сначала случай $k=0$.

Так как мы имеем дело с потоком без последействия, то при $\Delta\tau > 0$

$$v_0(\tau + \Delta\tau, t) = v_0(\tau, t)\, v_0(\Delta\tau, t+\tau);$$

но по предположению при $\Delta\tau \to 0$ и постоянных t, τ

$$v_0(\Delta\tau, t+\tau) = 1 - w(\Delta\tau, t+\tau) = 1 - \lambda(t+\tau)\,\Delta\tau + o(\Delta\tau),$$

следовательно,

$$v_0(\tau + \Delta\tau, t) - v_0(\tau, t) = -v_0(\tau, t)\,\lambda(t+\tau)\,\Delta\tau + o(\Delta\tau);$$

это после почленного деления на $\Delta\tau$ в пределе приводит к соотношению

$$\frac{\partial v_0(\tau, t)}{\partial \tau} = -\lambda(t+\tau)\, v_0(\tau, t) \tag{5.2}$$

(причем существование производной, очевидно, попутно доказывается); отсюда же

$$\frac{\partial \ln v_0(\tau, t)}{\partial \tau} = -\lambda(t+\tau),$$

и, следовательно,

$$\ln v_0(\tau, t) - \ln v_0(0, t) = -\int_0^\tau \lambda(t+u)\,du;$$

а так как $\ln v_0(0, t) = 0$, то

$$\ln v_0(\tau, t) = -\int_0^\tau \lambda(t+u)\,du = -\Lambda(\tau, t);$$

$$v_0(\tau, t) = e^{-\Lambda(\tau, t)}.$$

В стационарном случае мы имели показателем $-\lambda\tau$; в общем случае, как мы теперь видим, мы должны заменить λ величиной

$$\frac{1}{\tau}\Lambda(\tau, t) = \frac{1}{\tau}\int_0^\tau \lambda(t+u)\,du,$$

которую естественно рассматривать как среднее значение «мгновенного параметра» $\lambda(t)$ в промежутке $(t, t+\tau)$.

Переходя теперь к случаю $k > 0$, мы аналогично предыдущему [см. § 3] при $\Delta\tau \to 0$ и постоянных t, τ легко находим

$$v_k(\tau + \Delta\tau, t) = v_k(\tau, t)\, v_0(\Delta\tau, t+\tau) + v_{k-1}(\tau, t)\, v_1(\Delta\tau, t+\tau) + o(\Delta\tau),$$

где

$$v_0(\Delta\tau, t+\tau) = 1 - \lambda(t+\tau)\,\Delta\tau + o(\Delta\tau)$$

и

$$v_1(\Delta\tau,\ t+\tau) = 1 - v_0(\Delta\tau,\ t+\tau) - \psi(\Delta\tau,\ t+\tau) = \\ = \lambda(t+\tau)\,\Delta\tau + o(\Delta\tau),$$

так что

$$v_k(\tau+\Delta\tau,\ t) = v_k(\tau,\ t)\,[1-\lambda(t+\tau)\,\Delta\tau] + \\ + v_{k-1}(\tau,\ t)\,\lambda(t+\tau)\,\Delta\tau + o(\Delta\tau)$$

откуда

$$\frac{v\ \ (\tau+\Delta\tau,\ t) - v_k(\tau,\ t)}{\Delta\tau} = \lambda(t+\tau)\,[v_{k-1}(\tau,\ t) - v_k(\tau,\ t)] + o(1),$$

и следовательно, в пределе

$$\frac{\partial v_k(\tau,\ t)}{\partial\tau} = \lambda(t+\tau)\,[v_{k-1}(\tau,\ t) - v_k(\tau,\ t)]. \tag{5.3}$$

Это соотношение, доказанное нами для любого $k > 0$, остается, как показывает (5.2), верным и при $k=0$, если положить

$$v_{-1}(\tau,\ t) \equiv 0.$$

Мы найдем нужное нам решение системы (5.3), применяя метод производящих функций. Положим

$$F(t,\ \tau,\ x) = \sum_{k=0}^{\infty} v_k(\tau,t)\,x^k.$$

Умножая все члены уравнения (5.3) на x^k и суммируя по k от 0 до ∞, мы находим в точной аналогии с § 3

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

или

$$\frac{\partial \ln F}{\partial\tau} = (x-1)\,\lambda(t+\tau),$$

откуда

$$\ln F(t,\ \tau,\ x) - \ln F(t,\ 0,\ x) = (x-1)\int_0^{\tau}\lambda(t+u)\,du = \\ = (x-1)\,\Lambda(\tau,\ t). \tag{5.4}$$

При любых x и t мы имеем

$$F(t,\ 0,\ x) = v_0(0,\ t) = 1;$$

поэтому (5.4) дает

$$F(t,\ \tau,\ x) = e^{(x-1)\Lambda(\tau,\ t)} = e^{-\Lambda(\tau,\ t)}\,e^{x\Lambda(\tau,t)} = \\ = \sum_{k=0}^{\infty} e^{-\Lambda(\tau,\ t)}\frac{[\Lambda(\tau,\ t)]^k}{k!},$$

и сопоставление с определением функции $F(t, \tau, x)$ дает

$$v_k(\tau, t) = e^{-\Lambda(\tau, t)} \frac{[\Lambda(\tau, t)]^k}{k!} \quad (k = 0, 1, 2, \ldots). \tag{5.5}$$

Эти формулы полностью решают поставленную задачу. Мы видим, что и для потока с переменным параметром число вызовов в промежутке $(t, t+\tau)$ подчиняется закону Пуассона; однако параметр этого закона теперь зависит, кроме длины τ данного промежутка, еще и от его начального момента t. В случае стационарного потока мы имели закон Пуассона с параметром $\lambda\tau$; при переходе к нестационарному случаю мы должны, как видим, заменить постоянное число λ выражением

$$\frac{\Lambda(\tau, t)}{\tau} = \frac{1}{\tau}\int_0^\tau \lambda(t+u)\,du,$$

т. е. средним значением $\lambda(x)$ в промежутке $t \leqslant x \leqslant t+\tau$. Обратно, если функция $\lambda(x) = \lambda$ есть постоянная величина, то, очевидно, при любом t

$$\Lambda(\tau, t) = \int_0^\tau \lambda(t+u)\,du = \lambda\tau$$

и формулы (5.5) переходят в решения, полученные нами в § 3 для стационарного случая.

Заметим, наконец, что число вызовов в промежутке $(t, t+\tau)$, подчиняясь закону Пуассона (5.5), имеет своим математическим ожиданием параметр этого закона, т. е. величину

$$\Lambda(\tau, t) = \int_t^{t+\tau} \lambda(u)\,du;$$

поэтому величину $\frac{1}{\tau}\Lambda(\tau, t)$ можно понимать как *среднюю* интенсивность нашего потока в промежутке $(t, t+\tau)$; предел же этой величины при $\tau \to 0$ есть *мгновенная* интенсивность $\mu(t)$ данного потока в момент t; мы находим

$$\mu(t) = \lim_{\tau\to 0}\frac{1}{\tau}\Lambda(\tau, t) = \lim_{\tau\to 0}\frac{1}{\tau}\int_t^{t+\tau}\lambda(u)\,du = \lambda(t).$$

Таким образом, и в случае простейшего потока с переменным параметром мы имеем совпадение мгновенной интенсивности потока с мгновенным значением параметра.

Глава 2

Общие свойства стационарных потоков

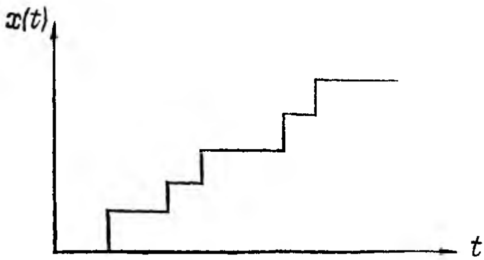
§ 6. Поток вызовов как случайный процесс

Как мы уже говорили в начале главы 1, потоки вызовов, с которыми мы встречаемся на практике, во многих случаях могут с достаточно хорошим приближением рассматриваться как простейшие; при изучении таких потоков поэтому обычно пользуются теми результатами, которые получены нами в главе 1. Однако в последние годы, когда усложняющаяся практика ставит перед наукой задачи все более сложные и требующие все более точного решения, встала настоятельная необходимость изучения и потоков более общего типа. Непосредственных поводов для такого расширения изучаемой области имеется в основном два. С одной стороны, статистика потоков вызовов даже в самых обычных условиях при возрастающей точности показывает, что выводы, основанные на предположении простейшего характера потока, недостаточно хорошо согласуются с опытными данными; можно без труда и теоретически предвидеть необходимость такого рода расхождений: нетрудно сообразить, например, что в действительности почти всегда следует ожидать в изучаемом потоке известного последействия и что не всегда этим последействием можно пренебрегать. С другой стороны, имеются и такие случаи, когда изучаемый поток заведомо и принципиально отличается от простейшего; таковы все виды потоков переменной интенсивности (они не стационарны, см. § 5); таковы же и потоки, поступающие на вторую, третью, и т. д. линию «полнодоступного пучка», даже в том случае, когда на первую линию поступает простейший поток; все эти потоки обладают значительным последействием, возрастающим с номером линии, и учет этого последействия обязателен для теории (см. гл. 8). Поэтому современные исследования по теории массового обслуживания не могут ограничиться рассмотрением простейших потоков и вынуждены расширить в той или другой мере исходные предпосылки.

Переходя к исследованию потоков более общего типа, мы должны в целях строгости и недвусмысленной ясности изложения начать с точного определения основных понятий; мы не сделали этого в главе 1, так как рассматривали эту главу как вводную, имеющую целью на простейшем примере показать характерные для всей теории потоков образцы применяемых в ней математических методов.

Если мы обозначим через $x(t)$ число вызовов, поступающих за промежуток времени $(0, t)$, то для каждого фиксированного значения $t > 0$ $x(t)$ представляет собой случайную величину. При переменном t,

$x(t)$ представляет собой однопараметрическое семейство случайных величин, которое называют **случайным процессом** или **случайной функцией**. Для функции $x(t)$ характерно то, что она: 1) может принимать только целые неотрицательные значения и 2) с возрастанием t никогда не убывает. График такой функции поэтому независимо от случая всегда имеет форму «лестницы», изображенную на черт. 1.



Черт. 1

Для задания любого случайного процесса $x(t)$ как такового надо, чтобы для любой конечной группы положительных чисел $t_1, t_2, \ldots, t_n$ был задан n-мерный закон распределения вектора

$$x(t_1),\ x(t_2), \ldots, x(t_n).$$

Если процесс $x(t)$ представляет собой поток вызовов и, следовательно, $x(t)$ может принимать только целые неотрицательные значения, то для задания этого потока как случайного процесса надо поэтому задать для каждой группы положительных чисел $t_1, t_2, \ldots, t_n$ и каждой группы целых неотрицательных чисел $k_1, k_2, \ldots, k_n$ вероятность системы равенств $x(t_1) = k_1, x(t_2) = k_2, \ldots, x(t_n) = k_n$; очевидно, эта вероятность может быть отличной от нуля только в том случае, если при $t_1 < t_2 < \ldots < t_n$ мы имеем и $k_1 \leqslant k_2 \leqslant \ldots \leqslant k_n$. В частности ($n = 1$), для любого $t > 0$ и любого целого неотрицательного k должна быть известна вероятность равенства $x(t) = k$, которую мы в главе 1 обозначали через $v_k(t)$*. Таким образом, система функций $v_k(t)$ $(k = 0, 1, 2, \ldots)$ входит как обязательный элемент в состав описания каждого потока вызовов. В общем случае, однако, задания этой системы функций для полной характеристики потока еще недостаточно.

Поток вызовов называется **стационарным**, если при $0 < t_1 < < t_2 < \ldots < t_n$ и при любом положительном a закон распределения вектора $x(t_i)$ $(1 \leqslant i \leqslant n)$ совпадает с законом распределения вектора $x(a + t_i) - x(a)$ $(1 \leqslant i \leqslant n)$; иначе говоря, закон распределения вектора $x(a + t_i) - x(a)$ $(1 \leqslant i \leqslant n)$ зависит от чисел t_i, но не зависит от a. В частности ($n = 1$), $v_k(t)$ для стационарного процесса означает

---

* Необходимо, впрочем, отметить, что в главе 1 $v_k(t)$ (в силу предположенной стационарности потока) означало вероятность поступления k вызовов в *любом* промежутке времени длины t, в то время как здесь речь идет лишь о промежутке $(0, t)$.

вероятность поступления k вызовов в промежутке $(a, a+t)$, где $a \geqslant 0$ произвольно, т. е. в любом промежутке длины t*.

Данный поток вызовов называется **потоком без последействия**, если закон распределения вектора $x(a+t_i) - x(a)$ $(t_i > 0, 1 \leqslant i \leqslant n)$ при любом $a \geqslant 0$ не зависит от значений величины $x(t)$ при каких-либо значениях $t < a$. Это определение, очевидно, в точной форме выражает то требование, чтобы случайное течение потока вызовов после какого-либо момента времени a было независимым от его течения до момента a; но в этом и состоит отсутствие последействия в понимании теории вероятностей.

Легко видеть, что стационарный поток без последействия полностью характеризуется системой функций $v_k(t)$ $(k = 0, 1, 2, \ldots)$, т. е. законом распределения числа вызовов, поступающих в течение (где угодно расположенного) промежутка времени длины t. В самом деле, так как система равенств

$$x(t_i) = k_i \quad (1 \leqslant i \leqslant n,\ 0 < t_1 < t_2 < \ldots < t_n),$$

очевидно, равносильна системе равенств

$$x(t_i) - x(t_{i-1}) = k_i - k_{i-1} \quad (1 \leqslant i \leqslant n),$$

где для общности положено $t_0 = k_0 = x(t_0) = 0$, то (обозначая через $\mathsf{P}\{\ \}$ вероятность события, помещенного в фигурных скобках) мы будем иметь

$$\mathsf{P}\{x(t_i) = k_i,\ 1 \leqslant i \leqslant n\} = \mathsf{P}\{x(t_i) - x(t_{i-1}) = k_i - k_{i-1},\ 1 \leqslant i \leqslant n\};$$

так как промежутки (t_0, t_1), $(t_1, t_2), \ldots, (t_{n-1}, t_n)$ взаимно не перекрываются, то отсюда в силу отсутствия последействия

$$\mathsf{P}\{x(t_i) = k_i,\ 1 \leqslant i \leqslant n\} = \prod_{i=1}^{n} \mathsf{P}\{x(t_i) - x(t_{i-1}) = k_i - k_{i-1}\},$$

а так как в силу стационарности потока

$$\mathsf{P}\{x(t_i) - x(t_{i-1}) = k_i - k_i\} = \mathsf{P}\{x(t_i - t_{i-1}) = k_i - k_i\} = \\ = v_{k_i - k_{i-1}}(t_i - t_{i-1}) \quad (1 \leqslant i \leqslant n),$$

то**

$$\mathsf{P}\{x(t_i) = k_i,\ 1 \leqslant i \leqslant n\} = \prod_{i=1}^{n} v_{k_i - k_{i-1}}(t_i - t_{i-1}).$$

---

* Для читателя, знакомого с теорией случайных процессов, заметим, что определенный таким образом стационарный поток вызовов не является, конечно, стационарным случайным процессом в общепринятом смысле этого термина. В теории случайных процессов наш стационарный поток принадлежит к числу «процессов со стационарными приращениями».

** Здесь и в дальнейшем мы, разумеется, полагаем $v_r(t) = 0$ для любого отрицательного индекса r.

Это показывает, что заданием системы функций $v_k(t)$ действительно однозначно определяются вероятности вида $\mathsf{P}\{x(t_i)=k_i,\ 1\leqslant i\leqslant n\}$, т. е. полностью характеризуется данный поток как случайный процесс.

В частности, для простейшего потока с параметром λ мы имели (§ 2)

$$v_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}\quad (k=0,1,2,\ldots);$$

поэтому для простейшего потока с параметром λ мы имеем при

$$0=t_0<t_1<\ldots<t_n,\ 0\leqslant k_0\leqslant k_1\leqslant\ldots\leqslant k_n$$

$$\mathsf{P}\{x(t_i)=k_i,\ 1\leqslant i\leqslant n\}=\prod_{i=1}^{n}e^{-\lambda(t_i-t_{i-1})}\frac{\lambda^{k_i-k_{i-1}}(t_i-t_{i-1})^{k_i-k_{i-1}}}{(k_i-k_{i-1})!}=$$

$$=e^{-\lambda t_n}\lambda^{k_n}\prod_{i=1}^{n}\frac{(t_i-t_{i-1})^{k_i-k_{i-1}}}{(k_i-k_{i-1})!}.$$

§ 7. Основное свойство стационарных потоков

Для любого стационарного потока условимся, как в главе 1, обозначать через $w(t)$ вероятность того, что в течение промежутка времени длины t произойдет по крайней мере один вызов, так что

$$w(t)=1-v_0(t)=\sum_{k=1}^{\infty}v_k(t).$$

В § 4 мы убедились, что для стационарного потока без последействия (в частности, для простейшего потока) отношение $w(t)/t$ при $t\to 0$ стремится к определенному пределу λ, который мы называли параметром данного потока и который, как мы видели, имеет важнейшее значение для изучения основных свойств этого потока. Соотношение

$$\frac{w(t)}{t}\to\lambda\ (t\to 0),\tag{7.1}$$

равносильное соотношению

$$w(t)=\lambda t+o(t)\quad (t\to 0),\tag{7.2}$$

часто выражают, говоря, что $w(t)$ при малых t «асимптотически пропорционально» t. В подавляющем большинстве изложений теории простейшего потока соотношение (7.2) [или (7.1)] прямо включается в определение простейшего потока*, что, как мы видели, является излишним, так как это соотношение выводится как простое следствие требований стационарности и отсутствия последействия.

* См. Erlang [7], Феллер [3], Фрай [4], Хинчин [5].

Однако соотношение (7.1) на самом деле обладает еще значительно более широкой областью применимости; оно имеет место, как мы теперь убедимся, *для любого стационарного потока*. Параметром, как мы его определяли в главе 1, обладает, таким образом, каждый стационарный поток, независимо от наличия или отсутствия последействия. Это обстоятельство дает нам, как мы увидим, весьма удобный опорный пункт для изучения общих свойств стационарных потоков.

Доказательство существования предела (7.1) для любого стационарного потока опирается на следующую элементарную лемму теории пределов, которая, как мы увидим, пригодится нам и в дальнейшем.

Лемма. *Пусть функция* $f(x)$ — *неотрицательная и неубывающая в отрезке* $0 < x \leqslant a$ *и* $f(x+y) \leqslant f(x) + f(y)$, *если* x, y *и* $x+y$ *принадлежат отрезку* $(0, a)$. *Тогда отношение* $f(x)/x$ *при* $x \to 0$ *либо безгранично возрастает, либо стремится к некоторому пределу; этот предел равен нулю только в тривиальном случае* $f(a) = 0$.

Доказательство. Из неравенства $f(x+y) \leqslant f(x) + f(y)$ легко следует, что

$$f(x) \leqslant m f\left(\frac{x}{m}\right) \tag{7.3}$$

для $0 < x \leqslant a$ и любого натурального числа m; в частности, при $x = a$

$$\frac{f\left(\frac{a}{m}\right)}{\frac{a}{m}} \geqslant \frac{f(a)}{a} \quad (m = 1, 2, \ldots);$$

это показывает, что [за исключением тривиального случая $f(a) = 0$]

$$\alpha = \overline{\lim_{x \to 0}} \frac{f(x)}{x} \geqslant \frac{f(a)}{a} > 0,$$

причем не исключен случай $\alpha = +\infty$.

Допустим сначала, что $\alpha < +\infty$. Пусть число $c > 0$ таково, что

$$\frac{f(c)}{c} > \alpha - \varepsilon, \tag{7.4}$$

где $\varepsilon > 0$ — сколь угодно малое наперед заданное число, и пусть $0 < x < c$. Определим натуральное число $m \geqslant 2$ из неравенств

$$\frac{c}{m} \leqslant x < \frac{c}{m-1};$$

тогда в силу (7.3) и предположенной монотонности $f(x)$

$$\frac{f(x)}{x} \geqslant \frac{f\left(\frac{c}{m}\right)}{\frac{c}{m-1}} = \frac{m-1}{m} \frac{m f\left(\frac{c}{m}\right)}{c} \geqslant \frac{m-1}{m} \frac{f(c)}{c}, \tag{7.5}$$

и, значит, в силу (7.4)

$$\frac{f(x)}{x} \geqslant \left(1 - \frac{1}{m}\right)(\alpha - \varepsilon),$$

а так как ε произвольно мало и $m \to \infty$ при $x \to 0$, то

$$\lim_{x \to 0} \frac{f(x)}{x} = \alpha,$$

и лемма доказана. Рассуждение остается в принципе тем же при $\alpha = +\infty$. Мы берем произвольно большое $A > 0$ и выбираем число c так, что $f(c)/c > A$; тогда из (7.5)

$$\frac{f(x)}{x} \geqslant \frac{m-1}{m} A,$$

и значит, $f(x)/x \to +\infty$ при $x \to 0$.

Теорема. *Для любого стационарного потока существует*

$$\lim_{t \to 0} \frac{w(t)}{t} = \lambda > 0,$$

причем не исключен случай $\lambda = +\infty$.

Доказательство. Достаточно показать, что функция $w(t)$ в некотором отрезке $(0, a)$ удовлетворяет всем предпосылкам только что доказанной леммы. Очевидно, что $w(t) \geqslant 0$ и с возрастанием t не может убывать; очевидно также, что $w(a) > 0$ при достаточно большом a, если мы отбросим тривиальный случай потока, в котором вызовы вообще невозможны. Наконец, если произошел по меньшей мере один вызов в промежутке $(0, t_1 + t_2)$, то, очевидно, то же самое должно иметь место для хотя бы одного из двух промежутков $(0, t_1)$ и $(t_1, t_1 + t_2)$, откуда

$$w(t_1 + t_2) \leqslant w(t_1) + w(t_2) \quad (t_1 > 0,\ t_2 > 0,\ t_1 + t_2 < a);$$

таким образом, и эта последняя предпосылка оказывается выполненной. Применяя лемму, мы видим, что теорема доказана.

Если данный стационарный поток есть поток без последействия, то для него, как мы это показали в § 2, мы имеем

$$v_0(t) = e^{-\lambda t} \quad (\lambda > 0 \text{ постоянная}),$$

за исключением случаев, когда в любом промежутке времени либо с достоверностью вовсе не поступает вызовов, либо с достоверностью поступает бесконечное множество вызовов; эти два случая, как не имеющие практического значения, мы условились выше исключить из рассмотрения. Отсюда для потока без последействия

$$w(t) = 1 - e^{-\lambda t}, \quad \lim_{t \to 0} \frac{w(t)}{t} = \lambda;$$

предел, существование которого мы доказали в последней теореме, в случае потока без последействия всегда есть, таким образом, некоторое конечное положительное число. Но если допустить возможность последействия, то λ может обращаться в $+\infty$ и в случаях, не исключенных нами из рассмотрения.

Чтобы в этом убедиться, рассмотрим следующий пример стационарного потока. Вообразим себе простейший поток, параметр λ которого представляет собой случайную величину, распределенную по некоторому закону $F(x)\,[F(+0)=0,\ F(+\infty)=1]$. В конце § 6 мы определили вероятность системы равенств $x(t_i)=k_i\,(1\leqslant t\leqslant n)$ для простейшего потока с данным значением параметра λ; будем теперь для краткости обозначать эту вероятность через $\mathsf{P}_\lambda(t_i, k_i)$. Если параметр λ есть случайная величина, распределенная по закону $F(x)$, то вероятность системы равенств $x(t_i)=k_i\,(1\leqslant i\leqslant n)$ будет (по формуле полной вероятности) равна

$$\mathsf{P}(t_i,\ k_i)=\int_0^\infty \mathsf{P}_\lambda(t_i,\ k_i)\,dF(\lambda). \tag{7.6}$$

Эта вероятность определена, таким образом, для любых $n,\ t_i,\ k_i$; а это, как мы знаем, означает задание определенного потока; этот поток будет, очевидно, стационарным; но, вообще говоря, он будет потоком с последействием. Если $w(t)$ и $v_k(t)$ имеют обычное значение для потока (7.6), а $w_\lambda(t)$ и $v_{k\lambda}(t)$ означают те же величины для простейшего потока с параметром λ, то мы, очевидно, имеем

$$w(t)=\int_0^\infty w_\lambda(t)\,dF(\lambda)$$

и

$$w_\lambda(t)=1-v_{0\lambda}(t)=1-e^{-\lambda t},$$

откуда

$$\frac{w(t)}{t}=\int_0^\infty \frac{1-e^{-\lambda t}}{t}\,dF(\lambda).$$

Если

$$\int_0^\infty \lambda\,dF(\lambda)<+\infty$$

[т. е. закон $F(x)$ имеет конечное математическое ожидание], то в силу $1-e^{-\lambda t}\leqslant\lambda t$ мы получаем

$$\frac{w(t)}{t}\leqslant\int_0^\infty \lambda\,dF(\lambda);$$

отношение $w(t)/t$ при $t \to 0$ ограничено, и рассматриваемый нами поток имеет конечный параметр. Но если интеграл

$$\int_0^\infty \lambda \, dF(\lambda) \tag{7.7}$$

расходится, то $w(t)/t \to +\infty$ при $t \to 0$. В самом деле, пусть $\varepsilon > 0$ произвольно мало и A столь велико, что

$$\int_0^A \lambda \, dF(\lambda) > \frac{1}{\varepsilon};$$

тогда в силу

$$\frac{w(t)}{t} \geqslant \int_0^A \frac{1 - e^{-\lambda t}}{t} \, dF(\lambda) \to \int_0^A \lambda \, dF(\lambda) \quad (t \to 0)$$

мы имеем

$$\lim_{t \to \infty} \frac{w(t)}{t} \geqslant \frac{1}{\varepsilon},$$

и значит, $w(t)/t \to +\infty$ $(t \to 0)$. Таким образом, в случае расходимости интеграла (7.7) наш поток имеет бесконечное значение параметра. Вместе с тем этот поток с вероятностью 1 дает конечное число вызовов в любом конечном промежутке времени; мы имеем

$$v_k(t) = \int_0^\infty v_{k\lambda}(t) \, dF(\lambda),$$

или, так как $v_{k\lambda}(t) = e^{-\lambda t} (\lambda t)^k / k!$,

$$v_k(t) = \frac{t^k}{k!} \int_0^\infty \lambda^k e^{-\lambda t} \, dF(\lambda) > 0,$$

$$\sum_{k=0}^\infty v_k(t) = \int_0^\infty e^{-\lambda t} \sum_{k=0}^\infty \frac{(\lambda t)^k}{k!} \, dF(\lambda) = \int_0^\infty dF(\lambda) = 1.$$

§ 8. Общая форма стационарного потока без последействия

Как мы видели в главе 1, стационарный поток без последействия, если он сверх того обладает еще свойством ординарности, есть простейший поток, общая структура которого легко может быть установлена. Теперь мы поставим себе задачу найти общий вид стационарного потока без последействия, отбрасывая требование ординарности.

Как мы видели в § 6, стационарный поток без последействия однозначно определяется заданием функций $v_k(t)$ $(k = 0, 1, 2, \ldots)$ [причем всегда $v_0(t) = e^{-\lambda t}$]. Поэтому наша задача сводится к определению общего вида семейства функций $v_k(t)$ для стационарных потоков без последействия. С этой целью мы прежде всего установим, что для любого такого потока при любом $k > 0$ отношение $v_k(t)/t$ при $t \to 0$ стремится к определенному пределу, который может быть либо нулем, либо положительным числом.

Обозначим через $\psi_k(t)$ $(k = 0, 1, 2, \ldots)$ вероятность того, что в промежутке длины t произойдет по меньшей мере k вызовов, так что для $k = 0, 1, 2, \ldots$

$$\psi_k(t) = \sum_{i=k}^{\infty} v_i(t), \qquad v_k(t) = \psi_k(t) - \psi_{k+1}(t),$$

$$\psi_0(t) = 1, \qquad \psi_1(t) = w(t), \qquad \psi_2(t) = \psi(t).$$

В § 7 мы доказали, что для любого стационарного потока отношение $\psi_1(t)/t$ при $t \to 0$ стремится к определенному пределу, конечному или бесконечному. В случае потока без последействия этот результат является тривиальным, так как $\psi_1(t) = w(t) = 1 - e^{-\lambda t}$, откуда $\psi_1(t)/t \to \lambda$ при $t \to 0$. Но зато мы теперь убедимся, что в случае стационарного потока без последействия предел отношения $\psi_k(t)/t$ при $t \to 0$ существует для любого $k > 0$. Так как $v_k(t) = \psi_k(t) - \psi_{k+1}(t)$, то отсюда будет следовать, что и предел отношения $v_k(t)/t$ при $t \to 0$ существует для любого $k > 0$. В случае простейшего потока мы имеем, конечно, $v_1(t)/t \to \lambda$, $v_k(t)/t \to 0$ $(k > 1)$.

Подобно рассуждению § 7, мы начнем с доказательства одного элементарного вспомогательного предложения из теории пределов, представляющего собой некоторое усиление леммы § 7.

Лемма. *Пусть функция $f(x)$ — неотрицательная и неубывающая в отрезке $0 < x \leqslant a$, отношение $f(x)/x$ ограничено в этом отрезке и*

$$f(nx) \leqslant nf(x) + cn^2x^2, \tag{8.1}$$

где $c > 0$ — постоянная, n — любое натуральное число, и $0 < nx \leqslant a$; тогда отношение $f(x)/x$ при $x \to 0$ стремится к некоторому пределу $l \geqslant 0$.

Доказательство. Полагая в (8.1) $x = x_0/n$, где $0 < x_0 \leqslant a$, получаем

$$f(x_0) \leqslant nf\left(\frac{x_0}{n}\right) + cx_0^2,$$

откуда

$$\frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n}} \geqslant \frac{f(x_0)}{x_0} - cx_0. \tag{8.2}$$

Положим

$$\overline{\lim_{x \to 0}}\, \frac{f(x)}{x} = l;$$

мы можем принять $l>0$, так как при $l=0$ утверждение леммы тривиально.

Пусть $\varepsilon>0$ задано произвольно. Выберем $x_0<\frac{\varepsilon}{c}$ и так, чтобы

$$\frac{f(x_0)}{x_0}>l-\varepsilon.$$

Пусть $0<x<x_0$ и натуральное число $n>1$ определяется неравенствами

$$\frac{x_0}{n}\leqslant x<\frac{x_0}{n-1};$$

тогда в силу (8.2)

$$\frac{f(x)}{x}\geqslant\frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n-1}}=\frac{n-1}{n}\,\frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n}}\geqslant\frac{n-1}{n}\,\frac{f(x_0)}{x_0}-cx_0\geqslant$$
$$\geqslant\left(1-\frac{1}{n}\right)(l-\varepsilon)-\varepsilon>l-3\varepsilon,$$

если x достаточно мало; так как, с другой стороны, $f(x)/x<l+\varepsilon$ для достаточно малых x, то

$$\frac{f(x)}{x}\to l\qquad(x\to 0),$$

и лемма доказана.

Чтобы установить существование предела отношения $\psi_k(t)/t$ при $t\to 0$, нам надо только показать, что функция $\psi_k(t)$ при $k>0$ удовлетворяет всем предпосылкам доказанной леммы. Неотрицательность и монотонность $\psi_k(t)$ в любом отрезке самоочевидны. Далее из $\psi_k(t)\leqslant\psi_1(t)=w(t)$ и $w(t)/t\to\lambda$ $(t\to 0)$ вытекает ограниченность $\psi_k(t)/t$ в любом отрезке. Поэтому нам остается только убедиться, что $\psi_k(t)$ удовлетворяет соотношению (8.1).

Обозначим с этой целью через g_l верхнюю грань отношения $\psi_l(t)/t$ в области $0<t<+\infty$ $(l\geqslant 1)$ и положим

$$A_k=\sum_{i=1}^{k-1}g_i g_{k-i}.$$

Мы утверждаем, что

$$\psi_k(nt)\leqslant n\psi_k(t)+A_k\frac{n(n-1)}{2}t^2, \tag{8.3}$$

откуда и будет следовать, что функция $\psi_k(t)$ удовлетворяет соотношению (8.1).

Неравенство (8.3) мы докажем с помощью индукции по n. При $n=1$ оно тривиально. Допустим, что для некоторого n оно имеет место. Для того, чтобы в отрезке $[0,(n+1)t]$ длины $(n+1)t=nt+t$ поступило не менее k вызовов [вероятность чего равна $\psi_k\{(n+1)t\}$], необходимо, чтобы при каком-либо l $(0\leqslant l\leqslant k)$ имелось не менее l

вызовов в отрезке $(0, t)$ и не менее $k-l$ вызовов в отрезке $[t, (n+1)t]$ (длины nt); поэтому

$$\begin{aligned}\psi_k[(n+1)t] &\leqslant \sum_{l=0}^{k} \psi_l(t)\,\psi_{k-l}(nt) = \\ &= \psi_0(t)\,\psi_k(nt) + \psi_k(t)\,\psi_0(nt) + \sum_{l=1}^{k-1} \psi_l(t)\,\psi_{k-l}(nt) \leqslant \\ &\leqslant \psi_k(nt) + \psi_k(t) + nt^2 \sum_{l=1}^{k-1} g_l g_{k-l} = \\ &= \psi_k(nt) + \psi_k(t) + A_k n t^2.\end{aligned}$$

В силу (8.3), отсюда

$$\psi_k[(n+1)t] \leqslant (n+1)\,\psi_k(t) + A_k \frac{n(n+1)}{2} t^2,$$

а это и есть соотношение (8.3) с $n+1$ вместо n.

Таким образом, функция $\psi_k(t)$ при $k>0$ удовлетворяет всем предпосылкам доказанной леммы, и следовательно, при $t \to 0$ отношение $\psi_k(t)/t$, а значит, и отношение $v_k(t)/t$ стремятся к определенному пределу. Так как $w(t)/t \to \lambda > 0$ при $t \to 0$, то и отношение $v_k(t)/w(t)$ при этом имеет определенный предел.

Положим

$$\lim_{t \to 0} [v_k(t)/w(t)] = p_k \qquad (k = 1, 2, \ldots).$$

Отношение $v_k(t)/w(t)$ есть вероятность получить k вызовов в отрезке длины t, *если известно, что в этом отрезке вызовы существуют*. Предел этого отношения при $t \to 0$, т. е. число p_k, можно поэтому рассматривать, как вероятность получения k вызовов в определенный момент, если известно, что в этот момент вообще вызовы происходят (такое истолкование величины p_k возможно, но, конечно, не обязательно).

Перейдем теперь к определению общего вида функций $v_k(t)$ для стационарного потока без последействия. В § 3 мы установили для такого потока [см. (3.1)] общее соотношение

$$v_k(t+\tau) = \sum_{l=0}^{k} v_l(\tau)\,v_{k-l}(t) \qquad (t>0, \tau>0, k=0, 1, 2, \ldots);$$

так как при $\tau \to 0$

$$v_0(\tau) = e^{-\lambda\tau} = 1 - \lambda\tau + o(\tau),$$

то отсюда при $k>0$

$$v_k(t+\tau) = (1-\lambda\tau)\,v_k(t) + \sum_{l=1}^{k} v_l(\tau)\,v_{k-l}(t) + o(\tau),$$

и, следовательно,

$$\frac{v_k(t+\tau)-v_k(t)}{\tau}=-\lambda v_k(t)+\sum_{l=1}^{k}\frac{v_l(\tau)}{\tau}v_{k-l}(t)+o(1).$$

Но при $l>0$ и $\tau\to 0$ по доказанному выше

$$\frac{v_l(\tau)}{\tau}=\frac{v_l(\tau)}{w(\tau)}\frac{w(\tau)}{\tau}\to\lambda p_l;$$

поэтому предельный переход доказывает существование $v'_k(t)$ и дает

$$v'_k(t)=-\lambda v_k(t)+\lambda\sum_{l=1}^{k}p_l v_{k-l}(t)\qquad(k=1,2,\ldots). \tag{8.4}$$

Добавляя сюда очевидное соотношение

$$v'_0(t)=-\lambda v_0(t), \tag{8.5}$$

мы получаем систему уравнений, позволяющих однозначно определить систему функций $v_k(t)$. В частности, к этому ведет путь замены неизвестных функций, которым мы пользовались в § 3. Полагая, как там,

$$v_k(t)=e^{-\lambda t}u_k(t)\qquad(k=0,1,2,\ldots),$$

мы легко приводим систему (8.4) к виду

$$u'_k(t)=\lambda[p_1u_{k-1}(t)+p_2u_{k-2}(t)+\cdots+p_ku_0(t)],$$

позволяющему рекуррентно определить все функции $u_k(t)$ [а значит, и $v_k(t)$]. Так, например, мы в силу $u_0(t)\equiv 1$ находим

$$u'_1(t)=\lambda p_1,$$

откуда

$$u_1(t)=\lambda p_1 t,\quad v_1(t)=e^{-\lambda t}\lambda p_1 t.$$

Мы не будем проводить здесь дальнейших выводов, так как значительно более простые и изящные результаты дает метод производящих функций, к применению которого мы теперь и переходим.

Положим

$$F(t,x)=\sum_{k=0}^{\infty}v_k(t)x^k;$$

отыскание системы функций $v_k(t)$ сводится к отысканию функции $F(t,x)$. Умножая соотношение (8.4) [для $k=0$ заменяемое соотношением (8.5)] на x^k и суммируя по k от 0 до ∞, мы находим

$$\frac{\partial F}{\partial t}=-\lambda F+\lambda\sum_{k=1}^{\infty}x^k\sum_{l=1}^{k}p_l v_{k-l}(t)=$$

$$=-\lambda F+\lambda\sum_{l=1}^{\infty}p_l\sum_{q=0}^{\infty}v_q(t)x^{q+l}=-\lambda F+\lambda\sum_{l=1}^{\infty}p_l x^l\sum_{q=0}^{\infty}v_q(t)x^q,$$

или, полагая

$$\sum_{l=1}^{\infty} p_l x^l = \Phi(x),$$

получаем

$$\frac{\partial F}{\partial t} = \lambda\,[\Phi(x) - 1]\,F,$$

$$\frac{\partial \ln F}{\partial t} = \lambda\,[\Phi(x) - 1],$$

а так как при любом x

$$F(0, x) = v_0(0) = 1,$$

то интегрированием по t находим

$$F(t, x) = e^{\lambda[\Phi(x)-1]t}, \tag{8.6}$$

и наша задача решена. Заметим еще, что при любом t

$$F(t, 1) = \sum_{k=0}^{\infty} v_k(t) = 1,$$

вследствие чего (8.6) дает

$$\Phi(1) = \sum_{l=1}^{\infty} p_l = 1.$$

Таким образом, производящая функция $F(t, x)$ для любого стационарного потока без последействия имеет вид (8.6), где $\lambda > 0$ и $\Phi(x) = \sum_{l=1}^{\infty} p_l x^l$, $p_l \geqslant 0$, $\sum_{l=1}^{\infty} p_l = 1$.

Убедимся теперь, что и обратно, если числа λ и p_l $(l = 1, 2, \ldots)$ подчиняются только что перечисленным требованиям, то существует стационарный поток без последействия, производящая функция которого дается формулой (8.6).

С этой целью мы допустим, что моменты времени, в которые происходят вызовы, образуют простейший поток с параметром λ, так что для вероятности $\overset{*}{v}_k(t)$ того, что за промежуток времени t произойдет k таких «вызывающих моментов», мы имеем обычное выражение

$$\overset{*}{v}_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

Однако поток самих вызовов не будет, вообще говоря, простейшим, так как мы допустим, что в каждый вызывающий момент может, вообще говоря, с вероятностью, отличной от нуля, поступить и более одного вызова. Примем вероятность поступления в данный вызывающий момент ровно l вызовов равной p_l $(l = 1, 2, \ldots)$, независимо от того, каков данный вызывающий момент и каково было течение этого потока до данного момента. Этим соглашением мы задаем некоторый определенный поток вызовов, который, очевидно, будет стационарным потоком без последействия. Покажем, что производящая функция $F(t, x)$ этого потока дается формулой (8.6).

Число вызовов, происходящих в любой вызывающий момент, мы определили как случайную величину, принимающую значение l с вероятностью p_l $(l = 1, 2, \ldots)$; производящая функция такой величины есть

$$\sum_{l=1}^{\infty} p_l x^l = \Phi(x).$$

Возьмем теперь r любых различных между собой вызывающих моментов и обозначим через $\mathsf{P}_r(k)$ вероятность того, что в эти r моментов в совокупности произойдет k вызовов, так что суммарное число вызовов за r вызывающих моментов есть случайная величина с производящей функцией

$$\sum_{k=0}^{\infty} \mathsf{P}_r(k)\, x^k$$

[где, разумеется, $\mathsf{P}_r(k) = 0$ при $k < r$]. Но эта случайная величина есть сумма r взаимно независимых случайных величин, каждая из которых имеет производящую функцию $\Phi(x)$. Так как при сложении взаимно независимых случайных величин их производящие функции перемножаются*, то поэтому

$$\sum_{k=0}^{\infty} \mathsf{P}_r(k)\, x^k = \{\Phi(x)\}^r.$$

А так как с другой стороны, очевидно, для рассматриваемого потока

$$v_k(t) = \sum_{r=0}^{\infty} \overset{*}{v}_r(t)\, \mathsf{P}_r(k),$$

то

$$F(t, x) = \sum_{k=0}^{\infty} v_k(t)\, x^k = \sum_{k=0}^{\infty} x^k \sum_{r=0}^{\infty} \overset{*}{v}_r(t)\, \mathsf{P}_r(k) =$$

$$= \sum_{r=0}^{\infty} \overset{*}{v}_r(t) \sum_{k=0}^{\infty} \mathsf{P}_r(k)\, x^k = \sum_{r=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^r}{r!} \{\Phi(x)\}^r =$$

$$= e^{-\lambda t} \sum_{r=0}^{\infty} \frac{\{\lambda t \Phi(x)\}^r}{r!} = e^{\lambda t[\Phi(x) - 1]},$$

что совпадает с формулой (8.6).

---

* Это непосредственно вытекает из того, что для случайной величины ξ с законом распределения $\mathsf{P}\{\xi = n\} = q_n$ $(n = 0, 1, 2, \ldots)$ производящая функция $f(x) = \sum_{n=0}^{\infty} q_n x^n$ есть, очевидно, математическое ожидание величины x^{ξ}.

Мы можем формулировать результат нашего исследования так: совокупность всех стационарных потоков без последействия совпадает с совокупностью всех потоков, даваемых формулой (8.6), где $\lambda > 0$:

$$\Phi(x) = \sum_{l=1}^{\infty} p_l x^l, \quad p_l \geqslant 0 \quad (l = 1, 2, \ldots), \quad \sum_{l=1}^{\infty} p_l = 1.$$

С предметной точки зрения мы убедились, что для каждого стационарного потока без последействия поток вызывающих моментов является простейшим и для полного описания данного потока вызовов надо, кроме параметра λ этого простейшего потока, задать еще закон распределения $(p_1, p_2, \ldots, p_n, \ldots)$ числа вызовов, поступающих в любой выбранный вызывающий момент. Очевидно, эти соображения делают совершенно прозрачной структуру самого общего стационарного потока без последействия.

Заметим еще, что в случае $p_1 = 1$, $p_k = 0$ $(k > 1)$ формула (8.6) дает нам производящую функцию простейшего потока с параметром λ:

$$F(t, x) = e^{\lambda t (x-1)} = \sum_{k=0}^{\infty} \left\{ e^{-\lambda t} \frac{(\lambda t)^k}{k!} \right\} x^k.$$

Глава 3

Функции Пальма

§ 9. Определение и доказательство существования

После того как в конце предыдущей главы мы полностью выяснили строение стационарных потоков без последействия, мы должны теперь обратиться к исследованию потоков более общего типа. Для довольно широких классов таких потоков весьма удобным орудием исследования оказалась одна функция, введенная Пальмом [8] и примененная им с успехом к решению ряда задач. Пальм определяет эту функцию $\varphi_0(t)$ (для любого стационарного потока) как *условную вероятность отсутствия вызовов в промежутке* $(t_0, t_0 + t)$, *если известно, что в момент* t_0 *произошел вызов*. Однако такое определение вряд ли можно считать достаточно удобным; то условие, при котором должна быть вычислена вероятность $\varphi_0(t)$, т. е. наличие вызова в некоторый момент t_0, само имеет во всех актуальных случаях вероятность 0, и это обстоятельство, как известно, не позволяет нам непосредственно определить функцию $\varphi_0(t)$ для заданного потока с помощью известных правил расчета условных вероятностей. Поэтому мы дадим этой функции другое, более сложное определение, которое зато позволит однозначно определять ее для любого стационарного потока. Вместе с тем мы определим не одну функцию, а целую последовательность $\varphi_k(t)$ $(k = 0, 1, 2, \ldots)$ функций, которые будем называть **функциями Пальма** и которые в дальнейшем окажутся нам полезными при решении ряда важных задач.

Пусть мы имеем два последовательных промежутка времени, из которых первый имеет длину τ, а второй — t (в дальнейшем мы будем для краткости называть самые эти промежутки соответственно «промежутком τ» и «промежутком t»). Обозначим для данного стационарного потока через $H_k(\tau, t)$ $(k \geqslant 0)$ вероятность следующего двойного события: 1) в промежутке τ произойдет по меньшей мере один вызов; 2) в промежутке t произойдет не более k вызовов. Эти два события, вообще говоря, будут взаимно зависимы; так как вероятность события 1) в наших старых обозначениях есть $w(\tau)$, то отношение

$$\frac{H_k(\tau, t)}{w(\tau)} \tag{9.1}$$

выражает собою условную вероятность события 2) при условии, что имело место событие 1), т. е. вероятность появления не более k вызовов в промежутке t при условии, что в промежутке τ появился по меньшей мере один вызов.

Если это отношение при $\tau \to 0$ (и при постоянном t) стремится к некоторому пределу, то этот предел естественно называть условной вероятностью появления не более чем k вызовов в промежутке t *при условии, что в начальный момент этого промежутка произошел вызов.*

Убедимся теперь, что предел отношения (9.1) при $\tau \to 0$ (и постоянном t) всегда существует, если только данный стационарный поток имеет конечный параметр λ. С этой целью рассмотрим сначала отношение $H_k(\tau, t)/t$. Чтобы доказать существование предела этого отношения при $\tau \to 0$, достаточно убедиться, что величина $H_k(\tau, t)$ как функция от τ удовлетворяет всем предпосылкам леммы § 7. Неотрицательность и монотонность этой функции самоочевидны. Пусть $\tau = \tau_1 + \tau_2$ и промежуток τ_1 предшествует промежутку τ_2. Тогда, если выполнено то двойное событие, вероятность которого мы обозначили $H_k(\tau, t)$, то, очевидно, выполняется по меньшей мере одно из следующих двух событий:

(A) В промежутке τ_2 имеется по меньшей мере один вызов, в промежутке t имеется не более k вызовов [вероятность события (A) равна $H_k(\tau_2, t)$].

(B) В промежутке τ_1 имеется по меньшей мере один вызов, в промежутке $\tau_2 + t$ имеется не более k вызовов [вероятность события (B) равна $H_k(\tau_1, \tau_2 + t) \leqslant H_k(\tau_1, t)$ (так как при фиксированном τ, $H_k(\tau, t)$, очевидно, есть невозрастающая функция от t)]. Таким образом, мы находим

$$H_k(\tau_1 + \tau_2, t) \leqslant H_k(\tau_1, t) + H_k(\tau_2, t),$$

т. е. функция $H_k(\tau, t)$ удовлетворяет (относительно τ) и последней предпосылке леммы § 7. Применяя эту лемму, мы находим, что отношение $H_k(\tau, t)/\tau$ при $\tau \to 0$ стремится к некоторому пределу или безгранично возрастает; однако последний случай исключается, так как, очевидно, всегда $H_k(\tau, t) \leqslant w(\tau)$, а отношение $w(\tau)/\tau$ по нашему предположению стремится к конечному пределу λ.

Наконец,

$$\frac{H_k(\tau, t)}{w(\tau)} = \frac{H_k(\tau, t)/\tau}{w(\tau)/\tau};$$

числитель и знаменатель этой дроби по доказанному стремятся при $\tau \to 0$ к определенным пределам; поэтому и

$$\lim_{\tau \to 0} \frac{H_k(\tau, t)}{w(\tau)} = \Phi_k(t) \tag{9.2}$$

существует; разумеется, этот предел является функцией от t.

Положим теперь

$$h_0(\tau, t) = H_0(\tau, t), \quad h_k(\tau, t) = H_k(\tau, t) - H_{k-1}(\tau, t) \quad (k > 0);$$

очевидно, $h_k(\tau, t)$ есть вероятность того, что 1) в промежутке τ имеется по меньшей мере один вызов и 2) в промежутке t имеется ровно k вызовов; отношение $h_k(\tau, t)/w(\tau)$ представляет собой условную вероятность иметь k вызовов в промежутке t при условии, что в промежутке τ имеется по меньшей мере один вызов. Из (9.2) следует

$$\lim_{\tau \to 0} \frac{h_0(\tau, t)}{w(\tau)} = \Phi_0(t), \quad \lim_{\tau \to 0} \frac{h_k(\tau, t)}{w(\tau)} = \Phi_k(t) - \Phi_{k-1}(t) \quad (k > 0).$$

Полагая

$$\varphi_0(t) = \Phi_0(t), \quad \varphi_k(t) = \Phi_k(t) - \Phi_{k-1}(t) \quad (k > 0),$$

мы имеем

$$\lim_{\tau \to 0} \frac{h_k(\tau, t)}{w(\tau)} = \varphi_k(t) \quad (k = 0, 1, 2, \ldots).$$

Функции $\varphi_k(t)$ мы и будем называть **функциями Пальма**. Функция $\varphi_k(t)$ может быть понимаема как *вероятность иметь k вызовов в промежутке длины t при условии, что в начальный момент этого промежутка произошел вызов.* Этим она отличается от функции $v_k(t)$, представляющей собою вероятность того же события при условии, что относительно начального момента ничего неизвестно. Проведенное нами рассуждение показывает, что вся совокупность функций Пальма однозначно определяется для любого стационарного потока с конечным параметром λ.

Заметим еще, что так как $H_k(\tau, t)$ относительно t есть функция невозрастающая, а $w(\tau)$ от t не зависит, то все функции $\Phi_k(t)$ [в частности, функция $\Phi_0(t) = \varphi_0(t)$] — невозрастающие в области $0 < t < +\infty$.

§ 10. Формулы Пальма

Функции Пальма связаны с основными функциями $v_k(t)$ данного стационарного потока простыми и важными формулами, которые нам предстоит теперь вывести.

Допустим, что данный стационарный поток — ординарный и имеет конечный параметр λ. Рассмотрим снова промежуток времени длины $\tau + t$, составленный из «промежутка τ» и непосредственно следующего за ним «промежутка t». Обозначим через n_1 и n_2 соответственно числа вызовов в промежутках τ и t (n_1, n_2 — случайные величины). Мы имеем, очевидно,

$$v_k(\tau + t) = \mathsf{P}\{n_1 + n_2 = k\} = \sum_{r=0}^{k} \mathsf{P}\{n_1 = r,\ n_2 = k - r\},$$

откуда в силу ординарности данного потока при $\tau \to 0$

$$v_k(\tau + t) = \mathsf{P}\{n_1 = 0, n_2 = k\} + \mathsf{P}\{n_1 = 1, n_2 = k - 1\} + o(\tau). \quad (10.1)$$

Но

$$\begin{aligned} \mathsf{P}\{n_1 = 0,\ n_2 = k\} &= \mathsf{P}\{n_2 = k\} - \mathsf{P}\{n_1 > 0,\ n_2 = k\} = \\ &= v_k(t) - h_k(\tau, t), \end{aligned} \quad (10.2)$$

где мы пользуемся обозначениями § 9. С другой стороны,

$$\begin{aligned} \mathsf{P}\{n_1 = 1,\ n_2 = k - 1\} &= \mathsf{P}\{n_1 > 0, n_2 = k - 1\} - \\ &- \mathsf{P}\{n_1 > 1,\ n_2 = k - 1\} = h_{k-1}(\tau, t) + o(\tau). \end{aligned} \quad (10.3)$$

Вставляя (10.2) и (10.3) в (10.1), находим

$$v_k(t + \tau) = v_k(t) - h_k(\tau, t) + h_{k-1}(\tau, t) + o(\tau),$$

откуда

$$\frac{v_k(\tau + t) - v_k(t)}{\tau} = \frac{h_{k-1}(\tau, t)}{w(\tau)} \frac{w(\tau)}{\tau} - \frac{h_k(\tau, t)}{w(\tau)} \frac{w(\tau)}{(\tau)} + o(1).$$

В силу результатов § 9 отсюда вытекает дифференцируемость функции $v_k(t)$ и соотношение

$$v_k'(t) = \lambda [\varphi_{k-1}(t) - \varphi_k(t)] \quad (k > 0); \tag{10.4}$$

при $k = 0$ это же рассуждение дает

$$v_0'(t) = -\lambda \varphi_0(t), \tag{10.5}$$

так что соотношение (10.4) имеет место и при $k = 0$, если положить $\varphi_{-1}(t) \equiv 0$. Складывая соотношения (10.4) для $k = 0, 1, \ldots, m$ и обозначая через

$$V_m(t) = \sum_{k=0}^{m} v\ \ (t)$$

вероятность иметь в промежутке длины t *не более* m вызовов, мы находим

$$V_m'(t) = -\lambda \varphi_m(t) \quad (m = 0, 1, 2, \ldots). \tag{10.6}$$

Формулы (10.4) и (10.6) и были целью нашего вывода [у Пальма имеется лишь формула (10.5)]. Эти формулы иногда удобнее применять в интегральной форме. Интегрируя обе части (10.6) от 0 до t мы находим

$$V_m(+0) - V_m(t) = \lambda \int_0^t \varphi_m(u)\, du \quad (m = 0, 1, 2, \ldots),$$

но

$$V_m(+0) = 1;$$

в самом деле, мы имеем

$$V_m(+0) \geqslant V_0(+0) = v_0(+0) = 1 - w(+0);$$

а так как, при $t \to 0$, $w(t)/t \to \lambda$, то $w(+0) = 0$. Итак, мы находим для любого $t > 0$

$$1 - V_m(t) = \lambda \int_0^t \varphi_m(u)\, du \quad (m = 0, 1, 2, \ldots), \tag{10.7}$$

откуда легко вытекает

$$\left.\begin{aligned} v_0(t) &= 1 - \lambda \int_0^t \varphi_0(u)\, du; \\ v_k(t) &= \lambda \int_0^t [\varphi_{k-1}(u) - \varphi_k(u)]\, du \quad (k = 1, 2, \ldots). \end{aligned}\right\} \tag{10.8}$$

Эти формулы просто и непосредственно выражают функции $v_k(t)$ данного потока через функции Пальма $\varphi_k(t)$.

§ 11. Интенсивность стационарного потока. Теорема Королюка

В § 4 мы условились называть *интенсивностью* μ данного стационарного потока математическое ожидание числа вызовов в единицу времени; в силу аддитивности математических ожиданий мы имеем тогда, что математическое ожидание числа вызовов в промежутке длины t для стационарного потока пропорционально t, т. е.

$$\sum_{k=0}^{\infty} k v_k(t) = \mu t.$$

Там же мы убедились, что всегда $\mu \geqslant \lambda$, а для простейшего потока $\mu = \lambda$.

В работах прикладного характера совпадение параметров μ и λ обычно принимается как самоочевидный факт, не требующий даже оговорки, при исследовании стационарных потоков самого общего типа. Ввиду практического значения этого допущения представляется важным разобраться в его предпосылках и дать ему строгое обоснование там, где это возможно.

Остановимся сперва на случае стационарного потока без последействия. В § 8 мы видели, что для потоков этого рода производящая функция

$$F(t, x) = \sum_{k=0}^{\infty} v_k(t)\, x^k$$

имеет вид

$$e^{\lambda t\,[\Phi(x) - 1]},$$

где $\lambda > 0$, $\Phi(x) = \sum_{l=1}^{\infty} p_l x^l$, $\quad p_l \geqslant 0$ $(l = 1, 2, \ldots)$, $\sum_{l=1}^{\infty} p_l = 1$. Так как, очевидно,

$$\mu t = \sum_{k=0}^{\infty} k v_k(t) = \left\{\frac{\partial F(t, x)}{\partial x}\right\}_{x=1}$$

и так как

$$\frac{\partial F(t,x)}{\partial x} = F(t, x)\,\lambda t\,\Phi'(x), \quad F(t, 1) = \sum_{k=0}^{\infty} v_k(t) = 1,$$

то

$$\mu t = \lambda t\,\Phi'(1) = \lambda t\,[p_1 + 2p_2 + 3p_3 + \ldots],$$

откуда

$$\mu = \lambda [p_1 + 2p_2 + 3p_3 + \ldots].$$

Так как

$$\sum_{l=1}^{\infty} p_l = 1$$

и

$$p_l \geqslant 0 \quad (l \geqslant 1),$$

то мы непосредственно видим, что для равенства $\mu = \lambda$ необходимо и достаточно иметь $p_1 = 1$. Но при $p_1 = 1$ данный поток, как мы видели в конце § 8, является простейшим. Таким образом, *среди стационарных потоков без последействия только простейшие потоки удовлетворяют требованию* $\mu = \lambda$; *для всех остальных* $\mu > \lambda$.

Так как для стационарного потока без последействия ординарность есть необходимое и достаточное условие для того, чтобы этот поток был простейшим, то можно еще сказать, что *для стационарного потока без последействия необходимым и достаточным условием равенства* $\mu = \lambda$ *является ординарность этого потока.*

Выведенные нами в § 10 формулы Пальма позволяют, как это показал В. С. Королюк, легко убедиться, что для любого стационарного потока ординарность влечет за собой равенство $\mu = \lambda$ (причем не исключается случай $\mu = \lambda = +\infty$).

В самом деле, мы имеем

$$\mu = \sum_{k=1}^{\infty} k v_k(1) = \sum_{k=1}^{\infty} [v_k(1) + v_{k+1}(1) + v_{k+2}(1) + \ldots] =$$

$$= \sum_{k=1}^{\infty} [1 - V_{k-1}(1)] = \sum_{k=0}^{\infty} [1 - V_k(1)],$$

откуда в силу формулы (10.7)

$$\mu = \lambda \sum_{k=0}^{\infty} \int_0^1 \varphi_k(u)\, du. \tag{11.1}$$

Но

$$\varphi_k(u) = \lim_{\tau \to 0} \frac{h_k(\tau, u)}{w(\tau)}, \tag{11.2}$$

а так как отношение $h_k(\tau, u) / w(\tau)$ есть условная вероятность иметь в промежутке u ровно k вызовов (при условии наличия вызовов в промежутке τ), то при любом $m>0$

$$\sum_{k=0}^{m} \frac{h_k(\tau, u)}{w(\tau)} \leqslant 1 \quad (0<u\leqslant 1),$$

а потому в силу (11.2)

$$\sum_{k=0}^{m} \varphi_k(u) \leqslant 1 \quad (0<u\leqslant 1);$$

следовательно, при любом $m>0$

$$\sum_{k=0}^{m} \int_0^1 \varphi_k(u)\,du = \int_0^1 \left\{\sum_{k=0}^{m} \varphi_k(u)\right\} du \leqslant 1;$$

а значит, и

$$\sum_{k=0}^{\infty} \int_0^1 \varphi_k(u)\,du \leqslant 1,$$

и (11.1) дает $\mu\leqslant\lambda$; а так как еще в § 4 мы видели, что всегда $\mu\geqslant\lambda$, то $\mu=\lambda$, и наше утверждение доказано.

Глава 4

Потоки с ограниченным последействием

§ 12. Другой способ описания потока

Способ задания потока вызовов, описанный нами в § 6, исходит из понимания потока как случайного процесса $x(t)$ и ничем не отличается от общего способа описания произвольного случайного процесса. Это автоматическое включение теории потоков в общую теорию случайных процессов, несомненно, имеет свои преимущества, так как позволяет применять к изучению потоков методы и результаты общей теории случайных процессов; однако, учитывая специфические свойства наших потоков как случайных процессов [и в первую очередь то, что величина $x(t)$ всегда монотонна и принимает лишь целые неотрицательные значения], мы можем надеяться найти для этих потоков хоть и менее общий, но зато более простой и удобный способ описания. Этим вопросом мы теперь и займемся.

Пусть снова начальный момент данного потока есть $t_0=0$; пусть t_i $(i=1, 2, \ldots)$ есть момент i-го вызова, так что $t_{i-1}\leqslant t_i$ $(i=1, 2, \ldots)$; положим, наконец,

$$t_i - t_{i-1} = z_i \quad (i=1, 2, \ldots),$$

так что $z_1 = t_1$, а при $i > 1$ z означает величину промежутка времени между $(i-1)$-м и i-м вызовами. Очевидно, все t_i и z_i $(i = 1, 2, \ldots)$ представляют собой случайные величины, способные принимать лишь неотрицательные значения.

Условимся теперь считать поток заданным, если для любого $n > 0$ задан n-мерный закон распределения вектора $(z_1, z_2, \ldots, z_n)$. Очевидно, этот способ описания потока более элементарен, чем выбранный нами в § 6, так как в основе его лежит представление о потоке не как о случайном процессе общего вида, а как о последовательности случайных величин. Убедимся теперь, что оба способа задания потока равносильны, т. е. что поток, заданный с помощью какого-либо одного из этих двух способов, будет тем самым однозначно описанным и в смысле другого способа.

Пусть, как в § 6, $x(t)$ означает число вызовов, предшествующих моменту t; очевидно, неравенства $t_k < u$ и $x(u) \geqslant k$ выражают собой одно и то же событие; то же самое имеет место и для неравенств $t_k \geqslant u$, $x(u) < k$, а значит, и для неравенств $u \leqslant t_k < v$, $x(u) < k \leqslant x(v)$. Отсюда далее следует, что система неравенств

$$u_k \leqslant t_k < v_k \quad (1 \leqslant k \leqslant n) \tag{12.1}$$

выражает то же событие, что и система неравенств

$$x(u_k) < k \leqslant x(v_k) \quad (1 \leqslant k \leqslant n), \tag{12.2}$$

каковы бы ни были вещественные числа u_k, v_k $(1 \leqslant k \leqslant n)$.

Если поток вызовов задан в смысле § 6, то при любом n и при любых u_k, v_k $(1 \leqslant k \leqslant n)$ нам задан закон распределения $2n$-мерного вектора $[x(u_k), x(v_k); k = 1, 2, \ldots, n]$ и, значит, однозначно определена вероятность системы (12.2), а следовательно, и равносильной ей системы (12.1). Но ввиду произвольности чисел u_k, v_k последнее означает, что однозначно задан закон распределения вектора $(t_1, t_2, \ldots, t_n)$, а значит, и вектора

$$(z_1, z_2, \ldots, z_n) = (t_1, t_2 - t_1, \ldots, t_n - t_{n-1});$$

таким образом, данный поток однозначно определен и в новом смысле.

Обратно, если при любом n нам задан закон распределения вектора $(z_1, z_2, \ldots, z_n)$, то в силу

$$t_k = \sum_{i=1}^{k} z_i \quad (1 \leqslant k \leqslant n)$$

тем самым однозначно определен и закон распределения вектора

$(t_1, t_2, \ldots, t_n)$. Но при любых $u_1, u_2, \ldots, u_n$ и любых целых $k_1, k_2, \ldots$ $\ldots, k_n$ система неравенств

$$t_{k_i} < u_i, \quad t_{k_{i+1}} \geqslant u_i \quad (1 \leqslant i \leqslant n) \tag{12.3}$$

равносильна системе равенств

$$x(u_i) = k_i \quad (1 \leqslant i \leqslant n); \tag{12.4}$$

так как вероятность системы (12.3) однозначно определена, то то же имеет место и для системы (12.4). А это означает, что при любом n и любых u_i $(1 \leqslant i \leqslant n)$ однозначно определен закон распределения вектора $[x(u_1), x(u_2), \ldots, x(u_n)]$, т. е. что наш процесс однозначно определен в смысле § 6.

Таким образом, указанный нами новый способ задания потока вызовов действительно равносилен принятому в § 6.

§ 13. Потоки с ограниченным последействием

Если данный поток — без последействия, то величины $z_1, z_2, \ldots, z_n$, очевидно, взаимно независимы. Однако обратное заключение, как мы узнаем в дальнейшем, было бы неверным. Взаимная независимость величин z_k в значительной степени ограничивает явление последействия, но не исключает его полностью. В части II мы узнаем, что как раз потоки с взаимно независимыми z_k, но с наличием последействия играют важнейшую роль в теории обслуживания вызовов полнодоступным пучком линий. Мы должны поэтому заняться теперь установлением некоторых основных свойств таких потоков.

Условимся (следуя Пальму) называть *потоком с ограниченным последействием* всякий поток, у которого $z_1, z_2, \ldots, z_n, \ldots$ есть последовательность взаимно независимых случайных величин. Очевидно, для однозначного описания такого потока достаточно задать законы распределения всех величин z_k $(k = 1, 2, \ldots)$. Мы будем в дальнейшем обозначать эти законы через $F_k(x)$ $(k = 1, 2, \ldots)$.

Мы знаем, что если данный поток — стационарный и ординарный, то полное отсутствие последействия влечет за собой простейший характер потока, подробно изученный нами в главе 1. Поэтому стационарный и ординарный поток с ограниченным последействием мы можем рассматривать как некоторое обобщение простейшего потока*. Именно

* Интересно заметить, что ограниченность последействия следует из отсутствия последействия лишь для ординарных потоков. Неординарный поток без последействия может не быть потоком с ограниченным последействием. Примером может служить поток, рассмотренный нами в конце § 8 при $p_1 = p_2 = \frac{1}{2}$. (Этим замечанием автор обязан П. И. Васильеву.)

такого рода потоки представляют значительный интерес для теории обслуживания в случае систем с потерями (см. далее гл. 8). Стационарный ординарный поток с ограниченным последействием мы будем для краткости называть потоком типа P (или потоком Пальма).

В § 9 мы ввели для любого стационарного потока систему «функций Пальма» $\varphi_k(t)$ $(k=0,1,2,\ldots)$. Функция $\varphi_0(t)$, как мы теперь увидим, играет основную роль в теории потоков типа P. *Заданием этой функции поток типа P однозначно определяется.* В самом деле, так как в случае потока типа P величины $z_1, z_2, \ldots, z_n, \ldots$ между собою независимы, то для однозначного определения закона распределения каждого вектора $(z_1, z_2, \ldots, z_n)$ $(n=1,2,\ldots)$, а значит, и для однозначного описания потока достаточно задать законы распределения $F_k(x)$ величин z_k $(k=1,2,\ldots)$. Но эти законы однозначно определяются заданием функции Пальма $\varphi_0(t)$, как показывает следующее предложение.

Теорема. *Для потока типа P*

$$F_1(x)=\lambda\int_0^x \varphi_0(u)\,du,\quad F_k(x)=1-\varphi_0(x)\quad (k\geqslant 2).$$

Для лучшей обозримости мы разобьем доказательство на несколько этапов.

1. Так как $F_1(x)=\mathsf{P}\{t_1<x\}$ есть вероятность наличия вызовов в промежутке $(0,x)$ и, следовательно, в наших старых обозначениях равна

$$w(x)=1-v_0(x)=1-V_0(x),$$

то утверждаемое выражение для $F_1(x)$ непосредственно вытекает из формулы (10.7) при $m=0$; при этом параметр λ данного потока определяется через функцию $\varphi_0(t)$ с помощью соотношения

$$\lambda\int_0^\infty \varphi_0(u)\,du=F_1(+\infty)=1.$$

Нам остается, таким образом, рассмотреть случай $k>1$.

2. Положим, как в § 8,

$$\psi_k(t)=1-V_{k-1}(t)=v_k(t)+v_{k+1}(t)+\cdots$$

Тогда имеет место

Лемма. *Для любого потока типа P и любого $r>0$*

$$\frac{\psi_{r+1}(u)}{\psi_r(u)}\to 0\quad (u\to 0).$$

Доказательство. Обозначая через t_k момент k-го вызова и полагая, как прежде, $t_k - t_{k-1} = z_k$ $(k = 1, 2, \ldots)$, мы очевидно имеем

$$\psi_{r+1}(u) = \mathsf{P}\{t_{r+1} < u\} \leqslant \mathsf{P}\{t_r < u, z_{r+1} < u\},$$

и следовательно, в силу независимости z_{r+1} от t_r

$$\psi_{r+1}(u) \leqslant \psi_r(u) F_{r+1}(u).$$

Для доказательства леммы достаточно поэтому убедиться, что $F_{r+1}(u) \to 0$ при $u \to 0$.

Пусть $a > 0$ столь велико, что $\psi_r(a) > 0$. Пусть $x > 0$ произвольно мало и n таково, что $(n-1)\ x < a \leqslant nx$. Условимся называть «ячейками» отрезки $[(k-1)x, kx]$ $(1 \leqslant k \leqslant n)$. Если $t_r < nx$ и $z_{r+1} < x$, то моменты t_r и t_{r+1} лежат либо в одной ячейке, либо в двух соседних ячейках, так что по меньшей мере один из n отрезков

$$[(l-1)x, (l+1)x] \quad (l = 1, 2, \ldots, n)$$

длины $2\,x$ содержит более одного вызова. Поэтому

$$\begin{aligned}\mathsf{P}\{t_r < nx,\ z_{r+1} < x\} = \psi_r(nx) F_{r+1}(x) \leqslant n\psi_2(2x) = \\ = 2nx\,\frac{\psi_2(2x)}{2x} < 2(a+x)\frac{\psi_2(2x)}{2x},\end{aligned}$$

откуда при $x < a$

$$F_{r+1}(x) \leqslant \frac{2(a+x)}{\psi_r(nx)}\,\frac{\psi_2(2x)}{2x} \leqslant \frac{4a}{\psi_r(a)}\,\frac{\psi_2(2x)}{2x} \to 0$$

при $x \to 0$ в силу ординарности данного потока. Этим завершено доказательство нашей леммы.

3. Переходя теперь к доказательству теоремы, убедимся прежде всего, что $F_2(t) = 1 - \varphi_0(t)$. С этой целью рассмотрим введенную нами в главе 3 вероятность $h_0(\tau, t)$ того, что в некотором промежутке длины τ вызовы имеются, а в последующем за ним промежутке длины t вызовов нет. Если число вызовов в промежутке τ равно k, то из отсутствия вызовов в промежутке t следует $z_{k+1} > t$, вероятность чего есть $1 - F_{k+1}(t)$; поэтому

$$h_0(\tau, t) \leqslant \sum_{k=1}^{\infty} v_k(\tau)[1 - F_{k+1}(t)] \leqslant v_1(\tau)[1 - F_2(t)] + \psi_2(\tau).$$

С другой стороны, при том же условии (k вызовов в промежутке τ) из $z_{k+1} > t + \tau$ следует, что в промежутке t вызовов нет, поэтому

$$h_0(\tau, t) \geqslant \sum_{k=1}^{\infty} v_k(\tau)[1 - F_{k+1}(t+\tau)] \geqslant v_1(\tau)[1 - F_2(t+\tau)].$$

Таким образом,

$$v_1(\tau)[1-F_2(t+\tau)]\leqslant h_0(\tau,t)\leqslant v_1(\tau)[1-F_2(t)]+\psi_2(\tau).$$

Деля все части этих неравенств на $w(\tau)$ и замечая, что при $\tau\to 0$

$$\frac{v_1(\tau)}{w(\tau)}\to 1,\quad \frac{\psi_2(\tau)}{w(\tau)}\to 0,\quad \frac{h_0(\tau,t)}{w(\tau)}\to\varphi_0(t),$$

мы в пределе находим

$$1-F_2(t+0)\leqslant\varphi_0(t)\leqslant 1-F_2(t),$$

откуда $F_2(t)=1-\varphi_0(t)$ во всех точках непрерывности закона распределения $F_2(t)$.

4. Теперь мы с помощью индукции убедимся, что $F_{r+2}(t)=1-\varphi_0(t)$ для всех $r\geqslant 0$.

В силу стационарности данного потока закон распределения расстояния между двумя первыми вызовами, следующими за каким-либо моментом $a>0$, совпадает с законом распределения $F_2(t)$ расстояния z_2 между первыми двумя вызовами, следующими за моментом 0. Но если в промежутке $(0,a)$ имеется r вызовов, то расстояние между первыми двумя вызовами, следующими за моментом a, есть z_{r+2}, и закон распределения его, равный $F_{r+2}(t)$, не зависит от предшествующего течения потока. Таким образом,

$$F_2(t)=\sum_{r=0}^{\infty}v_r(a)F_{r+2}(t). \tag{13.1}$$

Пусть теперь уже установлено, что

$$F_2(t)=F_3(t)=\cdots=F_{r+1}(t)=1-\varphi_0(t).$$

Тогда (13.1) дает

$$1-\varphi_0(t)=[1-\varphi_0(t)]\sum_{k=0}^{r-1}v_k(a)+v_r(a)F_{r+2}(t)+\sum_{k>r}v_k(a)F_{k+2}(t),$$

откуда в силу $v_r(a)=\psi_r(a)-\psi_{r+1}(a)$

$$[1-\varphi_0(t)]\psi_r(a)-F_{r+2}(t)\psi_r(a)=-\psi_{r+1}(a)F_{r+2}(t)+\sum_{k>r}v_k(a)F_{k+2}(t).$$

Так как последняя сумма правой части не превосходит

$$\sum_{k>r}v_k(a)=\psi_{r+1}(a),$$

то мы получаем

$$\psi_r(a)|1-\varphi_0(t)-F_{r+2}(t)|\leqslant\psi_{r+1}(a);$$

$$|1-\varphi_0(t)-F_{r+2}(t)|\leqslant\frac{\psi_{r+1}(a)}{\psi_r(a)}.$$

Это неравенство имеет место при любых $t>0$ и $a>0$; но при $a\to 0$ правая часть по доказанной лемме стремится к нулю; а так как левая часть от a не зависит, то при любом $t>0$

$$F_{r+2}(t)=1-\varphi_0(t),$$

что и требовалось доказать.

Глава 5

Предельная теорема

§ 14. Постановка задачи. Теорема Пальма

Как мы уже говорили в главе 1, значительное большинство исследований прикладного характера основывается на предположении, что первичный поток поступающих на данную установку вызовов является простейшим потоком. Однако давно уже известен целый ряд принципиальных соображений, заставляющих сомневаться в том, что предпосылки, которые составляют собой определение простейшего потока, с достаточной степенью точности выполняются в большинстве практически встречающихся случаев (в особенности это относится к требованию отсутствия последействия). Если поэтому наблюдения и опыт констатируют некоторое небольшое отклонение реально встречающихся потоков от простейших, то этому не следует удивляться; более того, удивление может вызвать тот факт, что отклонения такого рода в большинстве случаев бывают менее значительными, чем этого можно было бы ожидать из теоретических соображений. Таким образом, если обычно при сопоставлении выводов теории с опытными данными перед исследователем встает задача — объяснить причины отклонения реально протекающих явлений от теоретически предсказанного их течения, то в данном случае дело обстоит как раз наоборот: опытные данные согласуются с выводами построенной теории, как правило, лучше, чем этого можно было бы ожидать по принципиальным соображениям, и именно это «слишком хорошее» согласие требует объяснения.

Пальмом [8] сделана заслуживающая внимания попытка объяснения фактов этого рода, исходя из предположения, что данный поток представляет собой простую сумму (суперпозицию) большого числа взаимно независимых потоков малой интенсивности, причем каждый из слагаемых потоков является стационарным и ординарным, в отношении же последействия эти потоки могут вести себя произвольным образом. При этом оказывается, что в весьма широких предположениях суммарный поток по своему характеру должен быть близок к простейшему. Такая постановка задачи, повидимому, во многих случаях близка к реальной ситуации. Так, если к данной установке прикреплено большое число абонентов, то общий поток вызовов слагается из потоков (сравнительно весьма малой интенсивности), исходящих от отдельных абонентов, причем эти слагаемые потоки можно в первом приближении считать стационарными, ординарными и взаимно независимыми.

Мы приходим на этом пути к ряду своеобразных предельных теорем, которые способны в значительной степени объяснить исследуемое явление. Этим вопросом мы и займемся в настоящей главе.

Пусть исследуемый поток представляет собой суперпозицию n стационарных, ординарных и взаимно независимых потоков. Обозначим через λ_r интенсивность r-го потока, через $\varphi_r(t)$ его функцию Пальма [которую в гл. 3 мы обозначали через $\varphi_0(t)$] и через $v_{kr}(t)$ — вероятность поступления в промежутке $(0, t)$ k вызовов r-го потока. Те же величины для суммарного потока обозначим соответственно через Λ, $\Phi(t)$ и $V_k(t)$ (так что в частности $\Lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$). Мы будем исходить из следующих предпосылок:

1°. При $n \to \infty$, Λ остается постоянным, в то время как числа $\lambda_1, \lambda_2, \ldots, \lambda_n$ равномерно стремятся к нулю, так что для любого $\varepsilon > 0$ мы имеем $\lambda_r < \varepsilon$ $(r = 1, 2, \ldots, n)$, если n достаточно велико.

2°. При любом постоянном $t > 0$ и при $n \to \infty$ числа $\varphi_r(t)$ $(r = 1, 2, \ldots, n)$ равномерно стремятся к единице, так что для любого $\varepsilon > 0$ мы имеем $1 - \varphi_r(t) < \varepsilon$ $(1 \leqslant r \leqslant n)$, если n достаточно велико.

Предпосылка 2° требует некоторого пояснения. Ближайший анализ показывает, что одного равномерного уменьшения интенсивности слагаемых потоков, выражаемого предпосылкой 1°, еще недостаточно для того, чтобы суммарный поток приближался к простейшему; этому могут помешать скопления большого числа вызовов одного и того же потока на небольших участках — скопления, возможность которых создается тем, что последействие в каждом из слагаемых потоков мы не подвергали до сих пор никаким ограничениям. Предпосылка 2° имеет целью как раз уменьшить шансы такого рода скоплений. Она говорит, что при сколь угодно большом t вероятность не получить после некоторого вызова за время t ни одного нового вызова *того же потока* должна при $n \to \infty$ стремиться к единице равномерно по всем слагаемым потокам.

Прежде всего мы покажем, что при сделанных предпосылках вероятность $V_0(t)$ отсутствия вызовов суммарного потока в промежутке $(0, t)$ приближается при $n \to \infty$ к соответствующей вероятности для простейшего потока с параметром Λ.

Теорема Пальма. *При постоянном $t > 0$ и при $n \to \infty$*

$$V_0(t) \to e^{-\Lambda t}.$$

Доказательство. Формула (10.5) дает для r-го потока

$$1 - v_{0r}(t) = \lambda_r \int_0^t \varphi_r(u)\,du = \lambda_r t - \lambda_r \int_0^t [1 - \varphi_r(u)]\,du,$$

или

$$v_{0r}(t) = 1 - \lambda_r t + \lambda_r \int_0^t [1 - \varphi_r(u)]\,du;$$

поэтому в силу 2° при достаточно большом n

$$v_{0r}(t) = 1 - \lambda_r t + \varepsilon\theta_r\lambda_r t, \quad |\theta_r| < 1, \quad 1 \leqslant r \leqslant n, \tag{14.1}$$

отсюда легко находим

$$|\ln v_{0r}(t) + \lambda_r t| < c(t)\varepsilon\lambda_r \quad (1 \leqslant r \leqslant n),$$

где $c(t) > 0$ зависит только от t. Так как в силу взаимной независимости потоков

$$V_0(t) = \prod_{r=1}^{n} v_{0r}(t),$$

то отсюда

$$|\ln V_0(t) + \Lambda t| = \left|\sum_{r=1}^{n} [\ln v_{0r}(t) + \lambda_r t]\right| \leqslant$$
$$\leqslant \sum_{r=1}^{n} |\ln v_{0r}(t) + \lambda_r t| < c(t)\varepsilon\Lambda.$$

Так как $\varepsilon > 0$ сколь угодно мало при достаточно большом n, то при $n \to \infty$

$$\ln V_0(t) \to -\Lambda t, \quad V_0(t) \to e^{-\Lambda t},$$

что и требовалось доказать.

Без достаточных оснований Пальм полагает, что доказанная теорема уже влечет за собой приближенно простейший характер суммарного потока*. Разумеется, этой теоремы еще далеко не достаточно, и мы должны теперь перейти к дальнейшему исследованию вопроса.

---

* Мы должны при этом отметить, что доказательство Пальма, с одной стороны, налагает на функции $\varphi_r(t)$ излишние требования, а с другой — содержит пробелы.

§ 15. Предельное поведение функций $V_k(t)$

Мы должны теперь в первую очередь убедиться, что и при любом $k>0$ функция $V_k(t)$ нашего суммарного потока при $n\to\infty$ стремится к соответствующей функции простейшего потока с параметром Λ, т. е. к $e^{-\Lambda t}(\Lambda t)^k/k!$ С этой целью нам понадобится следующая общая

Лемма. *Пусть мы имеем стационарный и ординарный поток с интенсивностью λ и функцией Пальма $\varphi(t)$ и пусть $\psi(t)$, как прежде, означает вероятность поступления не менее двух вызовов за время t. Тогда при любом $t>0$*

$$\psi(t)\leqslant\lambda t\,[1-\varphi(t)].$$

Доказательство. Разобьем отрезок $(0,t)$ на m равных между собой частей (ячеек)

$$\Delta_k=\left(\frac{k-1}{m}t,\ \frac{k}{m}t\right)\quad(1\leqslant k\leqslant m).$$

Поступление в отрезке $(0,t)$ по меньшей мере двух вызовов, очевидно, влечет за собой наступление по меньшей мере одного из следующих двух событий:

(A) Существует по меньшей мере одна ячейка Δ_k, содержащая не менее двух вызовов.

(B) Существует такая ячейка $\Delta_k\,(k<m)$, что как в Δ_k, так и в отрезке $\left(\frac{k}{m}t,\ t\right)$ содержатся вызовы.

Поэтому мы имеем

$$\psi(t)\leqslant\mathsf{P}(A)+\mathsf{P}(B). \tag{15.1}$$

Прежде всего мы имеем в силу ординарности данного потока

$$\mathsf{P}(A)\leqslant m\psi\left(\frac{t}{m}\right)=t\,\frac{\psi\left(\frac{t}{m}\right)}{\frac{t}{m}}\to 0\quad(m\to\infty). \tag{15.2}$$

Далее в § 9 мы обозначали через $h_0(\tau,t)$ вероятность того, что в промежутке длины τ имеются вызовы, а в следующем за ним промежутке длины t вызовов нет. $w(\tau)-h_0(\tau,t)$ есть поэтому вероятность того, что вызовы имеются как в τ, так и в t, поэтому

$$\begin{aligned}\mathsf{P}(B)&\leqslant\sum_{k=1}^{m-1}\left[w\left(\frac{t}{m}\right)-h_0\left(\frac{t}{m},\ \frac{m-k}{m}t\right)\right]\leqslant\\&\leqslant m\left[w\left(\frac{t}{m}\right)-h_0\left(\frac{t}{m},t\right)\right]=\\&=mw\left(\frac{t}{m}\right)\left[1-\frac{h_0\left(\frac{t}{m};\ t\right)}{w\left(\frac{t}{m}\right)}\right].\end{aligned} \tag{15.3}$$

Так как при $m \to \infty$

$$mw\left(\frac{t}{m}\right) = t\,\frac{w\left(\frac{t}{m}\right)}{\frac{t}{m}} \to \lambda t,$$

то правая часть последних неравенств при $m \to \infty$ стремится к $\lambda t\,[1 - \varphi(t)]$. А так как $\psi(t)$ от m не зависит, то из (15.1), (15.2) и (15.3) вытекает в пределе при $m \to \infty$

$$\psi(t) \leqslant \lambda t\,[1 - \varphi(t)],$$

что и требовалось доказать.

Введем теперь следующие обозначения для событий:

A_k — в промежутке $(0, t)$ поступает k вызовов суммарного потока;

H_1 — ни один из слагаемых потоков не дает в $(0, t)$ более одного вызова;

H_2 — по меньшей мере один из слагаемых потоков дает в $(0, t)$ более одного вызова.

Нашей целью является исследование асимптотического поведения величины $V_k(t) = \mathsf{P}(A\ )$. Но

$$\mathsf{P}(A_k) = \mathsf{P}(H_1 A_k) + \mathsf{P}(H_2 A_k),$$

и, обозначая функцию $\psi(t)$ для r-го слагаемого потока через $\psi_r(t)$, в силу доказанной леммы

$$[\mathsf{P}(H_2 A_k) \leqslant \mathsf{P}(H_2) \leqslant \sum_{r=1}^{n} \psi_r(t) \leqslant \sum_{r=1}^{n} \lambda_r t\,[1 - \varphi_r(t)];$$

а так как в силу предпосылки 2° мы имеем при достаточно большом n

$$1 - \varphi_r(t) < \varepsilon \quad (r = 1, 2, \ldots, n),$$

то при достаточно большом n

$$\mathsf{P}(H_2 A_k) \leqslant \varepsilon t\,\Lambda,$$

и следовательно, при $n \to \infty$

$$\mathsf{P}(H_2 A_k) \to 0, \quad V_k(t) = \mathsf{P}(H_1 A_k) + o(1). \tag{15.4}$$

Но событие $H_1 A_k$ состоит, очевидно, в том, что из n слагаемых потоков какие-то k дают в промежутке $(0, t)$ по одному вызову, тогда как остальные $n - k$ в этом промежутке вызовов не дают. Поэтому, если $C(r_1, r_2, \ldots, r_k)$ означает произвольное сочетание из k различных между собой чисел ряда $1, 2, \ldots, n$, то

$$\mathsf{P}(H_1 A_k) = \sum_{C} \frac{v_{1r_1}(t)\, v_{1r_2}(t) \ldots v_{1r_k}(t)}{v_{0r_1}(t)\, v_{0r_2}(t) \ldots v_{0r_k}(t)} \prod_{l=1}^{n} v_{0l}(t) =$$

$$= V_0(t) \sum_{C} \prod_{p=1}^{k} \frac{v_{1r_p}(t)}{v_{0r_p}(t)}, \tag{15.5}$$

где суммирование производится по всем сочетаниям описанного типа.

Теперь мы можем приступить к доказательству нашего основного утверждения.

Теорема. *При* $n \to \infty$

$$V_k(t) \to e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} \quad (k = 0, 1, 2, \ldots).$$

Доказательство. Из (14.1) следует, что при достаточно большом n

$$| w_r(t) - \lambda_r t | = | 1 - v_{0r}(t) - \lambda_r t | < \varepsilon \lambda_r t \quad (1 \leqslant r \leqslant n);$$

а так как $v_{1r}(t) = w_r(t) - \psi_r(t)$ и в силу доказанной леммы, при достаточно большом n, $\psi_r(t) < \varepsilon\lambda_r t$, то мы можем при постоянном t писать

$$\left. \begin{aligned} v_{1r_p}(t) &= \lambda_{r_p} t + q_1 \varepsilon \lambda_{r_p} t, \\ v_{0r_p}(t) &= 1 - \lambda_{r_p} t + q_2 \varepsilon \lambda_{r_p} t = 1 + q_3 \varepsilon \end{aligned} \right\} \quad (p = 1, 2, \ldots, k),$$

где q_1, q_2, q_3 (как и q_4, $q_5, \ldots$ в дальнейшем) ограничены при $n \to \infty$. Отсюда

$$\frac{v_{1r_p}(t)}{v_{0r_p}(t)} = \frac{\lambda_{r_p} t (1 + q_1 \varepsilon)}{1 + q_3 \varepsilon} \lambda_{r_p} t (1 + q_4 \varepsilon) \quad (p = 1, 2, \ldots k),$$

и, следовательно,

$$\prod_{p=1}^{k} \frac{v_{1r_p}(t)}{v_{0r_p}(t)} = \lambda_{r_1} \lambda_{r_2} \ldots \lambda_{r_k} t^k (1 + q_5 \varepsilon).$$

В силу (15.5) и теоремы Пальма (§ 14) поэтому

$$\mathsf{P}(H_1 A_k) = e^{-\Lambda t} t^k (1 + q_6 \varepsilon) \sum_{C} \lambda_{r_1} \lambda_{r_2} \ldots \lambda_{r_k}.$$

Так как ε произвольно мало при достаточно большом n, то для доказательства нашей теоремы нам в силу (15.4) достаточно убедиться, что при $n \to \infty$ *

$$S_k = \sum_{C_k} \lambda_{r_1} \lambda_{r_2} \ldots \lambda_{r_k} \to \frac{\Lambda^k}{k!}. \tag{15.6}$$

Это мы теперь и сделаем. При $k = 1$ соотношение (15.6) тривиально. Пусть поэтому для некоторого $k > 1$ при $n \to \infty$

$$S_{k-1} = \sum_{C_{k-1}} \lambda_{r_1} \lambda_{r_2} \ldots \lambda_{r_{k-1}} = \frac{\Lambda^{k-1}}{(k-1)!} + o(1). \tag{15.7}$$

---

* Мы здесь для большей отчетливости обозначаем различные сочетания по k из чисел $1, 2, \ldots, n$ через C_k (вместо прежнего обозначения C).

Умножим каждый член суммы S_{k-1} на сумму всех λ_i, не входящих в него, т. е. на величину

$$\Lambda - \lambda_{r_1} - \lambda_{r_2} - \cdots - \lambda_{r_{k-1}}.$$

Тогда после раскрытия всех скобок мы получим сумму произведений вида

$$\lambda_{r_1} \lambda_{r_2} \ldots \lambda_{r_k},$$

где индексы $r_1, r_2, \ldots, r_k$ попарно различны между собой.

Каждое такое произведение есть один из членов суммы S_k; обратно, любой член суммы S_k будет, очевидно, получен при этой операции и притом в точности k раз [член $\lambda_{r_1} \lambda_{r_2} \ldots \lambda_{r_{k-1}} \lambda_{r_k}$ получается как $(\lambda_{r_1} \lambda_{r_2} \ldots \lambda_{r_{k-1}}) \lambda_{r_k}$, как $(\lambda_{r_1} \lambda_{r_2} \ldots \lambda_{r_{k-2}} \lambda_{r_k}) \lambda_{r_{k-1}}$ и т. д., наконец, как $(\lambda_{r_2}, \ldots, \lambda_{r_k}) \lambda_{r_1}$]. Так как, очевидно, при достаточно большом n и при любой комбинации C_{k-1}

$$\Lambda - (k-1)\varepsilon \leqslant \Lambda - \lambda_{r_1} - \lambda_{r_2} - \cdots - \lambda_{r_{k-1}} \leqslant \Lambda,$$

то из нашего подсчета следует

$$(\Lambda - k\varepsilon) S_{k-1} \leqslant kS_k \leqslant \Lambda S_{k-1},$$

и следовательно, из (15.7) при $n \to \infty$

$$(\Lambda - k\varepsilon)\left[\frac{\Lambda^{k-1}}{(k-1)!} + o(1)\right] \leqslant kS_k \leqslant \Lambda\left[\frac{\Lambda^{k-1}}{(k-1)!} + o(1)\right];$$

а так как ε при достаточно большом n как угодно мало, то $kS_k \to \Lambda^k / (k-1)!$ при $n \to \infty$, т. е.

$$\lim_{n \to \infty} S_k = \frac{\Lambda^k}{k!},$$

и наша теорема доказана индукцией.

§ 16. Предельная теорема

Только что доказанная нами теорема устанавливает, что в рассматриваемых нами условиях функции $V_k(t)$ для суммарного потока стремятся при $n \to \infty$ к соответствующим функциям простейшего потока с параметром Λ. Это, однако, еще не означает, что наш суммарный поток сам приближается к этому простейшему потоку. Дело в том, что, как мы видели в § 6, совокупность функций $V_k(t)$ $(k = 0, 1, 2, \ldots)$ однозначно определяет собой данный поток лишь при условии, что это есть поток без последействия; мы же пока не рассматривали вопроса о последействии в нашем суммарном потоке. Поэтому вопрос о приближении этого суммарного потока к простейшему потоку с параметром Λ требует дальнейшего исследования.

Как показывает заключительная формула § 6, для простейшего потока с параметром Λ определяющей является формула

$$\mathsf{P}\{x(t_i)=k_i,\ 1\leqslant i\leqslant m\}=e^{-\Lambda t_m}\Lambda^{k_m}\prod_{i=1}^{m}\frac{(t_i-t_{i-1})^{k_i-k_{i-1}}}{(k_i-k_{i-1})!},\qquad (16.1)$$

где $t_0=k_0=0$, m — любое натуральное число, $0<t_1<t_2<\ldots<<t_m$, $0\leqslant k_1\leqslant k_2\leqslant\ldots\leqslant k_m$ и все k_i — неотрицательные целые числа. Мы можем поэтому считать наш суммарный поток стремящимся к простейшему потоку с параметром Λ, если для этого суммарного потока вероятность, стоящая в левой части равенства (16.1), при любых m, t_i, k_i $(1\leqslant i\leqslant m)$ и при безгранично возрастающем n имеет своим пределом правую часть этого равенства. Это мы теперь и установим.

Введем сначала более удобные для данной цели обозначения. Положим $t_i-t_{i-1}=u_i$, $k_i-k_{i-1}=l_i$, $t_m=\sum\limits_{i=1}^{m}u_i=u$, $k_m=\sum\limits_{i=1}^{m}l_i=k$ и обозначим через $n(u_i)$ число вызовов, поступающих в промежутке $u_i=(t_{i-1},\ t_i)$. Тогда, очевидно, равенство (16.1) равносильно равенству

$$\mathsf{P}\{n(u_i)=l_i,\ 1\leqslant i\leqslant m\}=e^{-\Lambda u}\Lambda^{k}\prod_{i=1}^{m}\frac{u_i^{l_i}}{l_i!}.\qquad (16.2)$$

В § 15 мы рассматривали событие H_1A_k, состоящее в том, что за некоторый промежуток времени $(0, u)$ происходит k вызовов и что все эти вызовы принадлежат различным слагающим потокам. Пусть отрезок $(0, u)$ разбит на m частей $u_1, u_2, \ldots, u_m$, длины которых мы будем обозначать теми же буквами, и пусть событие H_1A_k совершилось. В силу стационарности и взаимной независимости слагающих потоков для любого из поступивших в промежутке $(0, u)$ k вызовов вероятность попасть в отрезок $u_i\,(1\leqslant i\leqslant m)$ тогда равна u_i/u, каково бы ни было положение этого отрезка и каковы бы ни были моменты остальных поступивших вызовов. Обозначим через B событие

$$n(u_i)=l_i\,(1\leqslant i\leqslant m);$$

тогда из только что сказанного следует, что

$$\mathsf{P}_{H_1A_k}(B)=\frac{k!}{l_1!l_2!\ldots l_m!}\left(\frac{u_1}{u}\right)^{l_1}\left(\frac{u_2}{u}\right)^{l_2}\ldots\left(\frac{u_m}{u}\right)^{l_m}.$$

Но в § 15 мы доказали, что при $n\to\infty$

$$V_k(u)=\mathsf{P}(A_k)\to e^{-\Lambda u}\frac{(\Lambda u)^k}{k!},$$

$$\mathsf{P}(H_1A_k)\to e^{-\Lambda u}\frac{(\Lambda u)^k}{k!},\quad \mathsf{P}(H_2A_k)\to 0.$$

Поэтому, учитывая, что в силу $\sum_{i=1}^{m} l_i = k$ событие A_k есть следствие события B, мы находим при $n \to \infty$

$$\mathsf{P}(B) = \mathsf{P}(A_k B) = \mathsf{P}(H_1 A_k B) + \mathsf{P}(H_2 A_k B) =$$
$$= \mathsf{P}(H_1 A_k)\,\mathsf{P}_{H_1 A_k}(B) + \mathsf{P}(H_2 A_k)\,\mathsf{P}_{H_2 A_k}(B) \to$$
$$\to e^{-\Lambda u}\frac{(\Lambda u)^k}{k!}\cdot\frac{k!}{l_1! l_2! \ldots l_m!}\left(\frac{u_1}{u}\right)^{l_1}\left(\frac{u_2}{u}\right)^{l_2}\ldots\left(\frac{u_m}{u}\right)^{l_m} = e^{-\Lambda u}\Lambda^k\prod_{i=1}^{m}\frac{u_i^{l_i}}{l_i!}.$$

Так как правая часть этого соотношения совпадает с правой частью равенства (16.2), то наша предельная теорема доказана. Мы можем, таким образом, утверждать, что при рассматриваемых нами условиях суммарный поток действительно стремится к простейшему потоку с параметром Λ.

ЧАСТЬ II

СИСТЕМЫ С ПОТЕРЯМИ

§ 17. Вводные замечания

После того как в первой части книги нами подробно изучены свойства потока поступающих вызовов, мы в последующих разделах должны рассмотреть основные вопросы, связанные с обслуживанием этого потока. Каждая станция (пункт, в который поступают вызовы) снабжена несколькими приборами, предназначенными для обслуживания этих вызовов; эти приборы могут быть весьма различного типа соответственно возлагаемым на них функциям; в частности, таким «прибором» может быть и человек (телефонистка, продавец в магазине, врач в амбулатории и т. п.); мы условимся ради единства терминологии во всех случаях называть обслуживающие приборы *линиями*. Процесс обслуживания протекает так, что всякий поступающий вызов *занимает* на некоторое время одну из имеющихся в момент его поступления свободных (незанятых) линий; пока линия занята каким-либо вызовом, она недоступна для вновь поступающих вызовов; период занятия какой-либо линии одним вызовом мы будем (опять-таки только ради единства терминологии) называть *разговором*.

Для задач обслуживания основное значение имеет различение двух типов устройства станций. Если в момент поступления какого-либо вызова имеются свободные линии, то вызов при любом устройстве станции занимает какую-либо одну из них и приступает к разговору. Различия возникают лишь в том случае, когда поступающий вызов застает все линии занятыми. При одном устройстве (*системы с потерями*) такой вызов просто получает отказ (или, как говорят, «теряется») и все дальнейшее течение процесса обслуживания идет так, как если бы этот вызов вообще не поступал; при другом устройстве (*системы с ожиданием*) вызов, заставший все линии занятыми, сохраняется как претендент на будущий разговор и в дальнейшем занимает одну из освободившихся линий. Эти два типа устройств отличаются друг от друга не только способом решения основных задач, но и самой их постановкой: дело в том, что уже показатели качества обслуживания в этих двух случаях совершенно различны. Для систем с потерями основным показателем является, очевидно, вероятность отказа (потери) — понятие, лишенное смысла для систем с ожиданием. Напротив, для систем с ожиданием центральной задачей служит изучение времени ожидания как случайной величины; эта задача, очевидно, не имеет никакого смысла для систем с потерями.

Ввиду столь существенных различий мы, естественно, должны рассматривать эти два типа устройств обособленно друг от друга. Системам с потерями мы посвятим вторую, а системам с ожиданием — третью часть книги. Здесь же заметим еще только, что мыслимы и системы промежуточного типа; так, например, можно заставлять неудачные вызовы ожидать, покуда число ожидающих не превзойдет определенной границы, после чего вновь поступающие вызовы уже получают отказ. Теория таких имеющих несомненное практическое значение «смешанных» систем в настоящее время еще почти не разработана.

Кроме указанного основного различия, устройства станций отличаются друг от друга еще многими другими чертами, вследствие чего число различных между собой употребительных устройств становится очень большим. Само собой понятно, что в нашей краткой монографии мы имеем возможность остановиться лишь на небольшом числе типов таких устройств и что поэтому при решении каждой задачи мы вынуждены исходить из некоторых определенных предпосылок, которые в действительности отнюдь не являются единственно возможными. Мы считаем полезным здесь же рассмотреть еще некоторые важные черты, которыми могут отличаться друг от друга различные системы обслуживания вызовов.

1°. Мы всегда будем предполагать, что все имеющиеся линии в равной степени доступны всем поступающим вызовам (такую систему называют «полнодоступным пучком линий»). В противоположность этому в действительности нередко приходится встречать случаи, когда определенные категории вызовов «прикреплены» к определенным линиям и не могут занимать других линий.

2°. Мы почти всегда будем предполагать поток поступающих вызовов простейшим. В части I мы сделали достаточно указаний относительно того, в какой мере это допущение соответствует действительности.

3°. Мы будем называть полнодоступный пучок упорядоченным, если его линии перенумерованы так, что поступающий вызов всегда занимает линию с наименьшим номером из числа тех, которые свободны в момент его поступления (т. е. первую, если она свободна; вторую, если первая занята, а вторая свободна, и т. д.). В неупорядоченном пучке линии занимаются в случайном порядке. На практике встречаются устройства обоих типов. Следует отметить, что для ряда основных задач вопрос об упорядоченности или неупорядоченности пучка не играет никакой роли; такова, например, вся проблема Эрланга, которой будут посвящены главы 6 и 7. Однако есть и такие задачи, которые, напротив, получают смысл только для упорядоченного пучка. такова, например, задача Пальма, которая будет рассмотрена в главе 8.

4°. Для систем с ожиданием имеет во многих задачах существенное значение вопрос о порядке обслуживания ожидающих вызовов. Это обслуживание может проводиться либо в порядке очереди, либо в случайном порядке, причем в ряде задач подсчеты приводят в этих двух случаях к различным результатам.

5°. Наконец, весьма важное значение для большинства задач теории обслуживания имеет вопрос о законе распределения длительности занятий (разговоров). В значительном большинстве исследований этот закон предполагается *показательным* (т. е. вероятность того, что длительность разговора будет больше t, принимается равной $e^{-\beta t}$ где $\beta > 0$ — постоянная). Этот выбор обусловлен главным образом тем, что он значительно облегчает необходимые расчеты. Можно без преувеличения сказать, что заметное большинство задач теории обслуживания решается сравнительно просто при показательном распределении длительности разговоров и, напротив, приводит к неодолимым трудностям при почти всяком ином предположении о форме этого закона распределения.

Такое исключительное положение показательного распределения длительности разговоров обусловлено главным образом одним его важным свойством, которым мы не раз будем пользоваться в дальнейшем. Пусть $f_a(t)$ есть вероятность того, что разговор, длящийся уже a секунд, продлится еще не менее t секунд, так что при показательном распределении $f_0(t) = e^{-\beta t}$. Так как, очевидно, всегда

$$f_0(a+t) = f_0(a) f_a(t),$$

то при показательном распределении

$$e^{-\beta(a+t)} = e^{-\beta a} f_a(t),$$

откуда

$$f_a(t) = e^{-\beta t}.$$

Это означает, что *при показательном распределении длительности разговоров закон распределения оставшейся части разговора не зависит от его «возраста», т. е. от того, сколько времени он уже длится.* Именно это свойство показательного распределения в большинстве случаев и упрощает производимые расчеты. Вместе с тем оно же заставляет думать, что в практических ситуациях гипотеза показательного распределения длительности разговоров вряд ли может рассчитывать на точное осуществление и в лучшем случае способна служить лишь более или менее хорошим приближением к действительности.

Глава 6

Задача Эрланга для конечного пучка

§ 18. Постановка задачи

В этой главе мы будем иметь дело с полнодоступным пучком (упорядоченным или нет — безразлично) из n линий, на который поступает простейший поток вызовов с параметром λ; мы допустим, что длительность разговоров подчиняется показательному закону распределения $1-e^{-x}$. Так как в случае общего показательного закона $1-e^{-\beta x}$ средняя длительность разговора равна $1/\beta$, то выбор $\beta=1$ означает просто, что мы принимаем эту среднюю длительность разговора за единицу времени, что, конечно, ни в какой мере не ограничивает общности исследования.

Если известно, что в некоторый момент 0 было занято ровно k линий данного пучка $(0 \leqslant k \leqslant n)$, то число $N(t)$ занятых линий в какой-либо последующий момент t есть случайная величина, значение которой определяется рядом случайных факторов: моментами окончания тех k разговоров, которые ведутся в момент 0, моментами поступления новых вызовов между 0 и t и длинами тех разговоров, которые ведутся этими вызовами. Число $N(t)$ представляет собой, таким образом, однопараметрическое семейство случайных величин, или, как говорят, *случайный процесс*. Этот процесс обладает, при сделанных нами предпосылках, одним важным свойством, позволяющим применить к его изучению хорошо разработанные методы.

Пусть $N(t_0)=i$, т. е. в момент t_0 занято i линий. Тогда последующее течение процесса в вероятностном смысле независимо от всего, что происходило до момента t_0. В самом деле, это дальнейшее течение, как мы уже отметили, однозначно определяется следующими тремя факторами:

1. Моментами окончания тех i разговоров, которые ведутся в момент t_0.

2. Моментами появления новых вызовов после t_0.

3. Длительностями разговоров для вызовов, упомянутых в 2.

Но легко видеть, что ни один из этих трех случайных факторов не зависит от того, что происходило до момента t_0. Для фактора 1 это вытекает из принятого нами показательного распределения длительности разговоров, при котором, как мы видели в § 17, длительность остающейся части разговора не зависит от его возраста. Для фактора 2 это следует из того, что поступающий поток вызовов — простейший и, следовательно, не обладает последействием. Наконец, для фактора 3 это очевидно само собой. Таким образом, действительно все три перечисленных фактора не зависят от «прошлого» нашей системы, т. е. от течения процесса до момента t_0, а следовательно, не зависит от прошлого и течение процесса после момента t_0, ибо оно однозначно определяется указанными тремя факторами.

Таким образом, случайный процесс $N(t)$ обладает следующим свойством: если известно $N(t_0)$, то течение процесса после момента t_0 в вероятностном смысле независимо от его течения до момента t_0 (коротко: если известно настоящее, то будущее не зависит от прошедшего). Случайные процессы, обладающие этим свойством, называют **процессами Маркова**.

Если в некоторый момент t занято i линий пучка [т. е. $N(t)=i$], то мы будем говорить, что в этот момент система находится в «состоянии i» ($i=0, 1, \ldots, n$); всего, таким образом, для системы возможно $n+1$ различных состояний. Обозначим через $\mathsf{P}_{ik}(t)$ ($t>0$, $0\leqslant i\leqslant n$, $0\leqslant k\leqslant n$) условную вероятность того, что система, находившаяся в некоторый момент в состоянии i, по истечении t единиц времени перейдет в состояние k [вероятность $N(a+t)=k$ при условии $N(a)=i$]. Эти «переходные» вероятности играют основную роль в исследовании процессов Маркова. Очевидно, всегда

$$\mathsf{P}_{ik}(t)\geqslant 0, \quad \sum_{k=0}^{n}\mathsf{P}_{ik}(t)=1 \quad (0\leqslant i\leqslant n,\ 0\leqslant k\leqslant n).$$

Если $t_1>0$, $t_2>0$, $0\leqslant i\leqslant n$, $0\leqslant k\leqslant n$, то имеет место соотношение

$$\mathsf{P}_{ik}(t_1+t_2)=\sum_{r=0}^{n}\mathsf{P}_{ir}(t_1)\,\mathsf{P}_{rk}(t_2). \tag{18.1}$$

В самом деле, для того, чтобы за время t_1+t_2 перейти из состояния i в состояние k, система должна сперва перейти за время t_1 из состояния i в некоторое состояние r, а потом за время t_2 перейти из состояния r в состояние k, так что соотношение (18.1) есть результат простого применения «формулы полной вероятности». Очень важно отметить, что эта формула имеет место только для процессов Маркова; в самом деле, только для процессов Маркова переходную вероятность $\mathsf{P}_{rk}(t_2)$ можно считать независимой от i; если бы наш процесс не был процессом Маркова, то на месте $\mathsf{P}_{rk}(t_2)$ должна была бы стоять вероятность перехода за время t_2 из состояния r в состояние k *при дополнительном условии, что система предварительно за время t_1 перешла в состояние r из состояния i*. Для процессов же Маркова вероятность $\mathsf{P}_{rk}(t_2)$ не зависит от того, что происходило до этого перехода, благодаря чему и имеет место формула (18.1).

Формула (18.1), иногда называемая уравнением Чэпмана-Колмогорова, лежит в основании всех исследований о процессах Маркова; в нашем изложении она также будет играть значительную роль.

Если в начальный момент 0 система находится в данном состоянии i, то вероятность застать ее в момент $t>0$ в состоянии k равна $\mathsf{P}_{ik}(t)$. Мы можем, однако, сделать относительно начального момента допущение более общего характера: в момент 0 мы можем считать известным не состояние системы, а лишь «начальные вероятности» $\mathsf{P}_i(0)$ $(0 \leqslant i \leqslant n)$ различных состояний; этот общий случай, конечно, сводится к упомянутому нами частному случаю, когда из чисел $\mathsf{P}_i(0)$ какое-нибудь одно равно единице (а остальные равны нулю). Вероятность $\mathsf{P}_k(t)$ застать систему в момент t в состоянии k по формуле полной вероятности равна

$$\mathsf{P}_k(t) = \sum_{i=0}^{n} \mathsf{P}_i(0)\,\mathsf{P}_{ik}(t); \tag{18.2}$$

эта вероятность зависит как от t, так и от начальных данных $\mathsf{P}_i(0)$ $(0 \leqslant i \leqslant n)$.

Если в уравнении (18.1) помножить обе части на $\mathsf{P}_i(0)$ и просуммировать по i от 0 до n, то в силу (18.2) мы получим

$$\mathsf{P}_k(t_1+t_2) = \sum_{r=0}^{n} \mathsf{P}_r(t_1)\,\mathsf{P}_{rk}(t_2) \quad (0 \leqslant k \leqslant n). \tag{18.3}$$

Задача Эрланга, которой мы посвятим настоящую главу, состоит в отыскании вероятности застать систему в том или ином данном состоянии. В свете изложенного нами выше такая постановка вопроса требует пояснений; случайный процесс $N(t)$ нестационарен, вероятности

$$\mathsf{P}\{N(t)=k\} = \mathsf{P}_k(t) \quad (0 \leqslant k \leqslant n)$$

меняются с течением времени и, кроме того, зависят еще от начальных данных, т. е. от чисел $\mathsf{P}_i(0)$ $(0 \leqslant i \leqslant n)$; представляется поэтому, что искомые в задаче Эрланга вероятности $\mathsf{P}_k(t)$ могут быть определены лишь при данных t и $\mathsf{P}_i(0)$ $(0 \leqslant i \leqslant n)$. В приложениях, однако, обычно считают возможным говорить о вероятности p_k застать систему в состоянии k, независимо от выбранного момента времени и от начальных данных. Чтобы теоретически оправдать такую практику, можно попытаться установить, что процесс $N(t)$ при $t \to \infty$ безгранично приближается к некоторому стационарному процессу, не зависящему от начальных данных; говоря более конкретно, надо установить, что вероятности $\mathsf{P}_k(t)$ при $t \to \infty$ стремятся к некоторым постоянным числам p_k $(0 \leqslant k \leqslant n)$, не зависящим от начальных данных. Эти числа p_k мы тогда, естественно, и принимаем за искомые в задаче Эрланга вероятности застать систему в том или другом определенном состоянии, ибо число p_k, с одной стороны, не зависит от начальных данных задачи, а с другой — становится сколь угодно близким к реальной вероятности $\mathsf{P}_k(t)$, если процесс продолжается достаточно долгое время.

Итак, нашей задачей является показать, что *при* $t \to \infty$ *функции* $\mathsf{P}_k(t)$ *стремятся к числам* p_k $(0 \leqslant k \leqslant n)$, не зависящим от начальных данных. Разумеется, при этом числа p_k должны быть нами найдены. Для практики особо важное значение имеет число p_n — вероятность застать все линии занятыми. Это есть вероятность «потери» (отказа), являющаяся для систем с потерями важнейшим показателем качества обслуживания.

Мы прежде всего редуцируем поставленную нами задачу к другой, более удобной для применения уравнения Чэпмана-Колмогорова, с помощью следующего вспомогательного предложения.

Лемма. *Для того, чтобы вероятности* $\mathsf{P}_k(t)$ *при* $t \to \infty$ *стремились к не зависящим от начальных данных числам* p_k $(0 \leqslant k \leqslant n)$, *необходимо и достаточно, чтобы к тем же пределам стремились соответственно переходные вероятности* $\mathsf{P}_{ik}(t)$ $(0 \leqslant k \leqslant n)$ *при любом значении* i.

Доказательство. Оба утверждения леммы почти очевидны в силу (18.2).

1. Пусть $\mathsf{P}_k(t) \to p_k$ $(t \to \infty,\ 0 \leqslant k \leqslant n)$, где p_k не зависит от начальных данных; выбирая тогда $\mathsf{P}_i(0) = 1$, $\mathsf{P}_l(0) = 0$ $(l \neq i)$, мы в силу (18.2) имеем $\mathsf{P}_k(t) = \mathsf{P}_{ik}(t)$, и, следовательно, $\mathsf{P}_{ik}(t) \to p_k$ $(t \to \infty,\ 0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$.

2. Пусть, обратно, $\mathsf{P}_{ik}(t) \to p_k$ $(t \to \infty,\ 0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$; тогда в силу (18.2) при любом выборе вероятностей $\mathsf{P}_i(0)$ мы имеем

$$\mathsf{P}_k(t) \to \sum_{i=0}^{n} \mathsf{P}_i(0)\, p_k = p_k \quad (t \to \infty,\ 0 \leqslant k \leqslant n),$$

так как $\sum\limits_{i=0}^{n} \mathsf{P}_i(0) = 1$.

В силу этой леммы наша ближайшая задача сводится к доказательству того, что при $t \to \infty$ переходные вероятности $\mathsf{P}_{ik}(t)$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$ стремятся к пределам p_k, не зависящим от i.

§ 19. Теорема Маркова

Задачу, упомянутую в конце предыдущего параграфа, можно было бы пытаться решить, найдя выражения переходных вероятностей $\mathsf{P}_{ik}(t)$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$ для специального интересующего нас процесса $N(t)$ и стремясь затем анализом этих выражений установить существование нужных нам пределов и одновременно найти эти пределы. Однако мы предпочтем другой путь, при котором трудная задача отыскания функций $P_{ik}(t)$ может быть обойдена. В настоящем параграфе мы, не пытаясь найти переходных вероятностей $P_{ik}(t)$ и их пределов, установим только самый факт существования этих пределов; такой путь становится возможным потому, что эта теорема существования представляет собой свойство очень широкого класса процессов Маркова, отнюдь не являясь характеристикой нашего специального процесса $N(t)$. После того как это будет сделано, мы в следующем параграфе, опираясь на доказанное уже существование пределов, сможем найти эти пределы для специально интересующего нас процесса, снова минуя явные выражения функций $P_{ik}(t)$.

Условимся называть процесс Маркова, характеризуемый переходными вероятностями $P_{ik}(t)$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$, транзитивным, если существует такое $t > 0$, что $P_{ik}(t) > 0$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$. Таким образом, для транзитивного процесса существует такой промежуток времени, в течение которого возможен переход системы из любого состояния в любое другое. Непосредственно ясно, что интересующий нас процесс $N(t)$ транзитивен, причем в качестве t может быть выбрано любое положительное число.

Теорема Маркова*. *Для любого транзитивного процесса Маркова* $P_{ik}(t)$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$ *предел*

$$\lim_{t \to \infty} P_{ik}(t) = p_k \quad (0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$$

существует и не зависит от i.

Доказательство. Во всем дальнейшем k будет означать произвольное закрепленное число ряда $0, 1, \ldots, n$. Положим

$$\operatorname*{Max}_{0 \leqslant i \leqslant n} P_{ik}(t) = M_k(t), \quad \operatorname*{Min}_{0 \leqslant i \leqslant n} P_{ik}(t) = m_k(t).$$

В силу (18.1) для любого i $(0 \leqslant i \leqslant n)$ и любых $t > 0$, $\tau > 0$

$$P_{ik}(t+\tau) = \sum_{r=0}^{n} P_{ir}(\tau) P_{rk}(t) \leqslant M_k(t) \sum_{r=0}^{n} P_{ir}(\tau) = M_k(t),$$

а следовательно,

$$M_k(t+\tau) \leqslant M_k(t),$$

т. е. $M_k(t)$ есть невозрастающая функция от t. Подобным же образом легко убедиться, что $m_k(t)$ есть неубывающая функция от t. Отсюда следует, что $M_k(t)$ и $m_k(t)$ при $t \to \infty$ стремятся к определенным пределам. Теорема, очевидно, будет доказана, если мы пока-

* У Маркова доказана для «цепей», т. е. процессов с дискретным временем; однако доказательство переносится без всяких изменений на интересующий нас случай непрерывного времени.

жем, что эти пределы совпадают между собой, а для этого необходимо и достаточно иметь

$$\Delta_k(t) = M_k(t) - m_k(t) \to 0 \quad (t \to \infty).$$

В дальнейшем все суммы по всем индексам будут распространяться на значения $0, 1, \ldots, n$ этих индексов, вследствие чего мы можем не указывать области суммирования. Пусть $\mathsf{P}_{ir}(t_0) > 0$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$ (такое t_0 найдется в силу транзитивности процесса). Положим,

$$d_{il}^{(r)} = \mathsf{P}_{ir}(t_0) - \mathsf{P}_{lr}(t_0) \quad (0 \leqslant i, l, r \leqslant n),$$

и будем в дальнейшем обозначать через $\sum'$ (соответственно $\sum''$) суммы, распространенные лишь на область положительных (соответственно неположительных) $d_{il}^{(r)}$. Тогда в силу

$$\sum_r \mathsf{P}_{ir}(t_0) = \sum_r \mathsf{P}_{lr}(t_0) = 1 \quad (0 \leqslant i, l \leqslant n)$$

мы имеем

$$0 = \sum_r d_{il}^{(r)} = \sum_r{}' |d_{il}^{(r)}| - \sum_r{}'' |d_{il}^{(r)}| \quad (0 \leqslant i, l \leqslant n),$$

или

$$\sum_r{}' |d_{il}^{(r)}| = \sum_r{}'' |d_{il}^{(r)}| = h_{il} \quad (0 \leqslant i,\ l \leqslant n);$$

при этом в силу $\mathsf{P}_{lr}(t_0) > 0$ $(0 \leqslant l,\ r \leqslant n)$

$$h_{il} = \sum_r{}' d_{il}^{(r)} = \sum_r{}' [P_{ir}(t_0) - \mathsf{P}_{lr}(t_0)] < \sum_r{}' \mathsf{P}_{ir}(t_0) \leqslant \sum_r \mathsf{P}_{ir}(t_0) = 1.$$

Это неравенство имеет место для любых i и l, вследствие чего и

$$h = \max_{0 \leqslant i, l \leqslant n} h_{il} < 1.$$

Пусть теперь q — любое натуральное число. Тогда при $0 \leqslant i \leqslant n$, $0 \leqslant l \leqslant n$ в силу (18.1)

$$\mathsf{P}_{ik}(qt_0 + t_0) - \mathsf{P}_{lk}(qt_0 + t_0) = \sum_r \mathsf{P}_{ir}(t_0)\,\mathsf{P}_{rk}(qt_0) -$$

$$- \sum_r \mathsf{P}_{lr}(t_0)\,\mathsf{P}_{rk}(qt_0) = \sum_r [\mathsf{P}_{ir}(t_0) - \mathsf{P}_{lr}(t_0)]\,\mathsf{P}_{rk}(qt_0) =$$

$$= \sum_r d_{il}^{(r)}\,\mathsf{P}_{rk}(qt_0) = \sum_r{}' d_{il}^{(r)}\,\mathsf{P}_{rk}(qt_0) - \sum_r{}'' |d_{il}^{(r)}|\,\mathsf{P}_{rk}(qt_0) \leqslant$$

$$\leqslant M_k(qt_0)\,h_{il} - m_k(qt_0)\,h_{il} = h_{il}\,\Delta_k(qt_0) \leqslant h\Delta_k(qt_0).$$

Так как это неравенство имеет место для любых i, l, то можно принять

$$\mathsf{P}_{ik}(qt_0 + t_0) = M_k(qt_0 + t_0), \quad \mathsf{P}_{lk}(qt_0 + t_0) = m_k(qt_0 + t_0);$$

тогда мы получаем

$$\Delta_k(qt_0 + t_0) \leqslant h\Delta_k(qt_0);$$

рекуррентное же применение этого неравенства дает

$$\Delta_k(qt_0) \leqslant h^{q-1}\,\Delta_k(t_0) \leqslant h^{q-1} \to 0 \quad (q \to \infty).$$

В силу монотонности функции $\Delta_k(t)$ отсюда очевидно следует, что

$$\Delta_k(t) \to 0 \quad (t \to \infty).$$

Этим теорема Маркова доказана.

§ 20. Уравнения и формулы Эрланга

Теперь мы переходим к ставшему классическим методу Эрланга определения величин p_k, существование которых нами только что доказано. В отличие от предыдущего параграфа, мы будем при этом иметь в виду исключительно наш конкретный процесс $N(t)$.

Во всем дальнейшем мы будем иметь дело с промежутком времени бесконечно малой длины τ. Мы условимся для краткости обозначать через $o(\tau)$ всякую бесконечно малую порядка выше τ и соединять знаком $\approx$ всякие две величины, разность которых есть величина вида $o(\tau)$.

Согласно принятым нами в § 18 предпосылкам вероятность $w(\tau)$ поступления по меньшей мере одного вызова за промежуток времени τ есть величина $\approx\lambda\tau$, а вероятность $\psi(\tau)$ поступления более одного вызова — величина вида $o(\tau)$. С другой стороны, если какая-либо линия в данный момент занята, то вероятность оставаться занятой еще в течение τ секунд (или более) для нее равна $e^{-\tau}$; если занято k линий, то вероятность того, что все они останутся занятыми в течение промежутка времени τ, равна поэтому $e^{-k\tau}$; вероятность же того, что в течение промежутка времени τ по меньшей мере одна из этих линий освободится, равна

$$1 - e^{-k\tau} \approx k\tau.$$

Поступление вызовов и освобождение линий представляют собой элементарные события, в моменты которых скачкообразно меняется величина $N(t)$. Из того, что мы до сих пор установили для вероятностей таких элементарных событий, с очевидностью следует, что вероятность наступления в промежутке длины τ по меньшей мере одного элементарного события (того или другого типа) при $\tau \to 0$ асимптотически пропорциональна τ; вероятность же наступления в промежутке длины двух или более элементарных событий (все равно, каких типов) есть величина вида $o(\tau)$ (или, что то же, ≈ 0).

Эти замечания позволяют легко найти асимптотические выражения переходных вероятностей $\mathsf{P}_{ik}(\tau)$ при $\tau \to 0$. Прежде всего, если

$|i-k|>1$, то переход из состояния i в состояние k требует, очевидно, наступления по меньшей мере двух элементарных событий; поэтому в силу вышесказанного при $\tau \to 0$

$$\mathsf{P}_{ik}(\tau) \approx 0 \quad (|i-k|>1). \tag{20.1}$$

Далее для перехода из состояния $k<n$ в состояние $k+1$ требуется либо наступление одного вызова, либо наступление более чем одного элементарного события; поэтому в силу вышесказанного при $\tau \to 0$

$$\mathsf{P}_{k,\,k+1}(\tau) \approx \lambda\tau \quad (0 \leqslant k < n).$$

Чтобы система перешла из состояния $k>0$ в состояние $k-1$, требуется либо освобождение одной из линий, либо наступление более чем одного элементарного события; так как вероятность освобождения одной из k занятых линий за время τ при $\tau \to 0$, как мы видели выше, $\approx k\tau$, то мы находим

$$\mathsf{P}_{k,\,k-1}(\tau) \approx k\tau \quad (0 < k \leqslant n).$$

Наконец, в силу (20.1) мы имеем при $\tau \to 0$

$$\mathsf{P}_{kk}(\tau) \approx 1 - \mathsf{P}_{k,\,k+1}(\tau) - \mathsf{P}_{k,\,k-1}(\tau) \quad (0 \leqslant k \leqslant n),$$

где только при $k=n$ второй, а при $k=0$ третий член правой части надо заменить нулем; это дает

$$\begin{aligned} &\mathsf{P}_{00}(\tau) \approx 1 - \lambda\tau;\\ &\mathsf{P}_{kk}(\tau) \approx 1 - \lambda\tau - k\tau \quad (1 \leqslant k \leqslant n-1);\\ &\mathsf{P}_{nn}(\tau) \approx 1 - n\tau. \end{aligned}$$

Таким образом, для всех вероятностей $\mathsf{P}_{ik}(\tau)$ нами установлены очень простые асимптотические выражения, дающие их с точностью до величины вида $o(\tau)$.

Теперь мы обратимся к уравнению (18.3), в силу которого при любом постоянном $t>0$

$$\mathsf{P}_k(t+\tau) = \sum_r \mathsf{P}_r(t)\,\mathsf{P}_{rk}(\tau).$$

Применяя к вероятностям $\mathsf{P}_{rk}(\tau)$ в правой части этого равенства найденные нами асимптотические оценки, мы находим

$$\begin{aligned} \mathsf{P}_0(t+\tau) &= \mathsf{P}_0(t)\,\mathsf{P}_{00}(\tau) + \mathsf{P}_1(t)\,\mathsf{P}_{10}(\tau) + o(\tau) =\\ &= (1-\lambda\tau)\,\mathsf{P}_0(t) + \tau P_1(t) + o(\tau);\\ \mathsf{P}_k(t+\tau) &= \mathsf{P}_{k-1}(t)\,\mathsf{P}_{k-1,\,k}(\tau) + \mathsf{P}_k(t)\,\mathsf{P}_{kk}(\tau) + \mathsf{P}_{k+1}(t)\,\mathsf{P}_{k+1,\,k}(\tau) + o(\tau) =\\ &= \lambda\tau\,\mathsf{P}_{k-1}(t) + (1-\lambda\tau-k\tau)\,\mathsf{P}_k(t) + (k+1)\,\tau\,\mathsf{P}_{k+1}(t) + o(\tau) \quad (0<k<n);\\ \mathsf{P}_n(t+\tau) &= \mathsf{P}_{n-1}(t)\,\mathsf{P}_{n-1,\,n}(\tau) + \mathsf{P}_n(t)\,\mathsf{P}_{nn}(\tau) + o(\tau) =\\ &= \lambda\tau\,\mathsf{P}_{n-1}(t) + (1-n\tau)\,\mathsf{P}_n(t) + o(\tau). \end{aligned}$$

Отсюда

$$\frac{\mathsf{P}_0(t+\tau)-P_0(t)}{\tau}=-\lambda\,\mathsf{P}_0(t)+\mathsf{P}_1(t)+o(1);$$

$$\frac{\mathsf{P}_k(t+\tau)-\mathsf{P}_k(t)}{\tau}=\lambda\mathsf{P}_{k-1}(t)-(\lambda+k)\,\mathsf{P}_k(t)+(k+1)\,\mathsf{P}_{k+1}(t)+o(1)\ (0<k<n);$$

$$\frac{\mathsf{P}_n(t+\tau)-\mathsf{P}_n(t)}{\tau}=\lambda\mathsf{P}_{n-1}(t)-n\,\mathsf{P}_n(t)+o(1).$$

Если мы заставим теперь τ стремиться к нулю (сохраняя t постоянным), то убеждаемся в существовании производных всех функций $\mathsf{P}_k(t)$ $(k=0, 1, \ldots, n)$ и находим в пределе

$$\left.\begin{aligned}&\mathsf{P}_0'(t)=-\lambda\mathsf{P}_0(t)+\mathsf{P}_1(t);\\&\mathsf{P}_k'(t)=\lambda\mathsf{P}_{k-1}(t)-(\lambda+k)\,\mathsf{P}_k(t)+(k+1)\,\mathsf{P}_{k+1}(t)\ \ (0<k<n);\\&\mathsf{P}_n'(t)=\lambda\mathsf{P}_{n-1}(t)-n\mathsf{P}_n(t).\end{aligned}\right\}\quad(\mathscr{E})$$

Система $n+1$ уравнений $(\mathscr{E})$ с $n+1$ неизвестными функциями $\mathsf{P}_k(t)$ $(k=0, 1, \ldots, n)$ называется системой Эрланга. Так как все уравнения этой системы однородны, то искомые функции содержат произвольный постоянный множитель, который может быть определен из очевидного «нормировочного» условия

$$\sum_{k=0}^{n}\mathsf{P}_k(t)=1.$$

Как мы уже говорили, нам нет надобности искать решения системы дифференциальных уравнений $(\mathscr{E})$. В § 19 мы доказали, что для любого k $(0\leqslant k\leqslant n)$ существует предел

$$\lim_{t\to\infty}\mathsf{P}_k(t)=p_k.$$

Отсюда следует, что правые части всех уравнений $(\mathscr{E})$ при $t\to\infty$ имеют пределы. Переходя к левым частям, мы видим, что все производные $\mathsf{P}_k'(t)$ при $t\to\infty$ стремятся к пределам; но такой предел может быть только нулем, так как, если бы какое-нибудь $\mathsf{P}_k'(t)$ стремилось к числу, отличному от нуля, соответствующее $|\mathsf{P}_k(t)|$ при $t\to\infty$ возрастало бы безгранично, что [независимо от реального смысла величин $\mathsf{P}_k(t)$ как вероятностей] невозможно уже в силу теоремы Маркова. Таким образом, мы приходим к выводу, что

$$\mathsf{P}_k'(t)\to 0\quad(t\to\infty)\quad(0\leqslant k\leqslant n),$$

вследствие чего система $(\mathscr{E})$ в пределе при $t\to\infty$ дает

$$\left.\begin{aligned}-\lambda p_0+p_1&=0;\\\lambda p_{k-1}-(\lambda+k)\,p_k+(k+1)\,p_{k+1}&=0\quad(0<k<n);\\\lambda p_{n-1}-np_n&=0.\end{aligned}\right\}\qquad(20.2)$$

Эта простая система линейных уравнений вместе с нормировочным условием $\sum_{k=0}^{n} p_k = 1$ может служить для однозначного определения искомых чисел p_k.

Если положить

$$\lambda p_{k-1} - kp_k = z_k \quad (1 \leqslant k \leqslant n),$$

система (20.2) может быть записана в виде

$$z_1 = 0, \quad z_k - z_{k+1} = 0 \quad (0 < k < n), \quad z_n = 0,$$

откуда $z_k = 0 \quad (1 \leqslant k \leqslant n)$; это же дает

$$p_k = \frac{\lambda}{k} p_{k-1} \quad (1 \leqslant k \leqslant n),$$

и, следовательно,

$$p_k = \frac{\lambda^k}{k!} p_0 \quad (1 \leqslant k \leqslant n).$$

Применяя нормировочное условие, находим

$$p_0 = \frac{1}{\sum_{l=0}^{n} \frac{\lambda^l}{l!}},$$

и следовательно,

$$p_k = \frac{\frac{\lambda^k}{k!}}{\sum_{l=0}^{n} \frac{\lambda^l}{l!}} \quad (k = 0, 1, \ldots, n). \tag{20.3}$$

Формулы (20.3), называемые обычно **формулами Эрланга**, полностью решают поставленную нами задачу. В частности, вероятность «потери» дается формулой

$$p_n = \frac{\frac{\lambda^n}{n!}}{\sum_{l=0}^{n} \frac{\lambda^l}{l!}}.$$

Полезно отметить, насколько решающую роль во всем проведенном исследовании играла предпосылка о показательном распределении длительности разговоров; только при этом допущении процесс $N(t)$ становится процессом Маркова; при отказе от этого допущения все развитые нами в § 19 и 20 методы становятся поэтому принципиально неприменимыми. В специальной литературе имеется целый ряд попыток доказать, что формулы Эрланга остаются в силе и при любом другом распределении длительности разговоров. Однако, насколько мы видим, эти попытки не привели до сих пор к сколько-нибудь законченным результатам.

§ 21. Эргодическая теорема

Любая вероятность в любой теории получает реальный смысл лишь в том случае, если известна реальная совокупность объектов, в которой эта вероятность интерпретирует собой долю того или другого признака. Каков же реальный смысл тех вероятностей p_k, которыми мы занимались в последних параграфах? Что означает, в частности, «вероятность потери» p_n?

В подавляющем большинстве специальных исследований реальная интерпретация этих вероятностей носит один и тот же вполне определенный характер: их истолковывают, как «средние относительные времена пребывания» системы в соответствующих состояниях. Это означает следующее. Обозначим через $x_k(t)$ величину, равную 1, если система в момент t находится в состоянии k, и равную 0 в противном случае. Тогда интеграл

$$\int_0^T x_k(t)\,dt$$

представляет собой суммарную длину тех промежутков времени (между 0 и T), в течение которых система находится в состоянии k а отношение

$$\frac{1}{T}\int_0^T x_k(t)\,dt$$

дает нам «среднее относительное время пребывания» системы в состоянии k [за промежуток $(0, T)$]. Под вероятностью p_k застать систему в состоянии k понимают тогда предел

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T x_k(t)\,dt. \tag{21.1}$$

Представляется, однако, очевидным, что определенные нами в предшествующих параграфах величины p_k непосредственно не допускают подобного истолкования. Это видно уже из того, что величина $x_k(t)$ представляет собой с принятой нами точки зрения случайную функцию (случайный процесс), а следовательно, предел (21.1) (если он существует) — случайную величину, которая поэтому не может отождествляться с вероятностью p_k, по сущности своей, не зависящей от случая. С другой стороны, вероятности p_k нами определены как пределы при $t\to\infty$ вероятностей $\mathsf{P}_k(t)$; если все вероятности понимать как средние времена пребывания системы в том или ином состоянии, то величины $\mathsf{P}_k(t)$ не допускают, как легко видеть, никакой разумной интерпретации.

Таким образом, избранное нами определение вероятностей p_k, а также и развитый нами метод их вычисления непосредственно не дают никаких оснований для отождествления их с пределами вида (21.1) вопреки установившейся во всей прикладной литературе практике. Если это отождествление невозможно, то все же с практической точки зрения представляется весьма желательным найти соображения, позволяющие с известным основанием считать интегралы

$$X_k(t) = \frac{1}{T}\int_0^T x_k(t)\,dt$$

при больших значениях T хотя бы приближенно совпадающими с определенными нами вероятностями p_k; если бы это удалось, то такое сближение в значительной степени оправдало бы общепринятую в специальной литературе практику понимания вероятностей p_k как «средних относительных времен пребывания» — практику, очень удобную в прикладных задачах.

Так как $X_k(t)$ есть с нашей точки зрения случайная величина, то близость ее к не зависящей от случая величине p_k при самых благоприятных обстоятельствах может утверждаться лишь с некоторой (достаточно большой) вероятностью. В настоящем параграфе мы докажем предложение, идущее в указанном направлении так далеко, как только можно было надеяться; именно, имеет место

Теорема. *Как бы мало ни было* $\varepsilon > 0$,

$$\lim_{T\to\infty} \mathsf{P}\{|X_k(t) - p_k| > \varepsilon\} = 0.$$

Предварительное замечание. Предложения подобного рода, в которых вероятность некоторого состояния системы, первоначально определенная как доля в большой совокупности систем одинакового строения, сближается затем со средним временем пребывания в данном состоянии для какой-либо одной системы за большой промежуток времени, в теоретической физике называются обычно эргодическими теоремами. Доказываемое нами предложение представляет собой весьма типичный пример эргодической теоремы.

Доказательство. Во всем дальнейшем начальные данные [вероятности $\mathsf{P}_k(0)$, $0 \leqslant k \leqslant n$] предполагаются произвольными, но твердо установленными. Математическое ожидание случайной величины ξ мы будем обозначать через $\mathsf{M}\xi$ или $\mathsf{M}\{\xi\}$. Так как величина $x_k(t)$ может принимать значения 1 и 0 с соответственными вероятностями $\mathsf{P}_k(t)$ и $1 - \mathsf{P}_k(t)$, то

$$\mathsf{M}x_k(t) = \mathsf{P}_k(t). \tag{21.2}$$

Так как $\mathsf{P}_k(t) \to p_k$ при $t \to \infty$, то

$$\int_0^T [\mathsf{P}_k(t) - p_k]\,dt = o(T) \quad (T \to \infty).$$

Поэтому при достаточно большом T

$$\mathsf{P}\{|X_k(T) - p_k| > \varepsilon\} = \mathsf{P}\left\{\left|\int_0^T [x_k(t) - p_k]\,dt\right| > \varepsilon T\right\} <$$

$$< \mathsf{P}\left\{\left|\int_0^T [x_k(t) - \mathsf{P}_k(t)]\,dt\right| > \frac{\varepsilon}{2} T\right\}.$$

Отсюда, применяя неравенство Чебышева, находим в силу (21.2)

$$\mathsf{P}\{|X_k(T) - p_k| > \varepsilon\} < \frac{4}{\varepsilon^2 T^2} \mathsf{M}\left\{\left[\int_0^T [x_k(t) - \mathsf{P}_k(t)]\,dt\right]^2\right\} =$$

$$= \frac{4}{\varepsilon^2 T^2} \mathsf{M}\left\{\int_0^T\int_0^T [x_k(u) - \mathsf{P}_k(u)]\,[x_k(v) - \mathsf{P}_k(v)]\,du\,dv\right. =$$

$$= \frac{4}{\varepsilon^2 T^2} \int_0^T\int_0^T [\mathsf{M}\{x_k(u)\,x_k(v)\} - \mathsf{P}_k(u)\,\mathsf{P}_k(v)]\,du\,dv =$$

$$= \frac{8}{\varepsilon^2 T^2} \iint\limits_{0 \leqslant u \leqslant v \leqslant T} [\mathsf{M}\{x_k(u)\,x_k(v)\} - \mathsf{P}_k(u)\,\mathsf{P}_k(v)]\,du\,dv.$$

Так как величина $x_k(u)\,x_k(v)$ может принимать лишь значения 1 и 0, то $\mathsf{M}\{x_k(u)\,x_k(v)\} = \mathsf{P}\{x_k(u)\,x_k(v) = 1\}$ есть вероятность застать систему в состоянии k как в момент u, так и в момент v; поэтому при $v > u$

$$\mathsf{M}\{x_k(u)\,x_k(v)\} = \mathsf{P}_k(u)\,\mathsf{P}_{kk}(v - u),$$

и мы находим

$$\mathsf{P}\{|X_k(T) - p_k| > \varepsilon\} < \frac{8}{\varepsilon^2 T^2} \iint\limits_{0 \leqslant u \leqslant v \leqslant T} \mathsf{P}_k(u)\,[\mathsf{P}_{kk}(v - u) - \mathsf{P}_k(v)]\,du\,dv =$$

$$= \frac{8}{\varepsilon^2 T^2} \int_0^T \mathsf{P}_k(u)\,du \int_u^T [\mathsf{P}_{kk}(v - u) - \mathsf{P}_k(v)]\,dv =$$

$$= \frac{8}{\varepsilon^2 T^2} \int_0^T \mathsf{P}_k(u)\,du \int_0^{T-u} [\mathsf{P}_{kk}(z) - \mathsf{P}_h(z + u)]\,dz.$$

Пусть теперь $\delta > 0$ произвольно мало и A настолько велико, что при $z > A$

$$|\mathsf{P}_{ik}(z) - p_k| < \frac{\delta}{2} \quad (0 \leqslant i \leqslant n).$$

Тогда при $z > A$, $u > 0$

$$|\mathsf{P}_{kk}(z) - p_k| < \frac{\delta}{2},$$

$$|\mathsf{P}_k(z+u) - p_k| = \left|\sum_i \mathsf{P}_i(0)\,\mathsf{P}_{ik}(z+u) - p_k\right| =$$

$$= \left|\sum_i \mathsf{P}_k(0)\,[\mathsf{P}_{ik}(z+u) - p_k]\right| \leqslant \frac{\delta}{2}\sum_i \mathsf{P}_i(0) = \frac{\delta}{2},$$

и следовательно,

$$|\mathsf{P}_{kk}(z) - \mathsf{P}_k(z+u)| < \delta.$$

Поэтому при $T > A$

$$\mathsf{P}\{|X_k(T) - p_k| > \varepsilon\} \leqslant \frac{8}{\varepsilon^2 T^2}\int_0^T du \int_0^T |\mathsf{P}_{kk}(z) - \mathsf{P}_k(z+u)|\,dz \leqslant$$

$$\leqslant \frac{8}{\varepsilon^2 T^2}\int_0^T du \left\{\int_0^A dz + \int_A^T \delta\,dz\right\} \leqslant \frac{8}{\varepsilon^2 T^2}\{AT + \delta T^2\} = \frac{8A}{\varepsilon^2 T} + \frac{8\delta}{\varepsilon^2}.$$

Беря сперва δ достаточно малым, выбирая затем A описанным выше образом и беря, наконец, T достаточно большим, мы видим, что правая часть последнего неравенства сколь угодно мала при достаточно большом T. Этим наша теорема доказана.

Глава 7

Задача Эрланга для бесконечного пучка

§ 22. Уравнение для производящей функции

Если число линий в пучке бесконечно, то рассматриваемую установку нельзя уже причислять к «системам с потерями», так как потери становятся невозможными. Расчет вероятностей различных состояний сохраняет, однако, практическое значение и в этом случае, так как на практике встречаются такие положения, когда потери недопустимы и число линий должно быть достаточно большим для того, чтобы вероятность потери оказалась пренебрегаемо малой; в таких случаях вероятности различных состояний дают возможность оценить степень использования системы, что в свою очередь имеет значение для расчета быстроты износа и других экономических показателей.

Если в формулах Эрланга (20.3) перейти к пределу при $n \to \infty$, то мы получаем

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}\ (k = 0, 1, 2, \ldots);$$

можно поэтому предвидеть, что это пуассоновское распределение и даст нам вероятности различных состояний в случае бесконечного пучка. Однако метод, которым мы пришли к формулам (20.3), в случае бесконечного пучка оказывается неприменимым, так как теорема Маркова, на которой он основан, существенным образом предполагает число состояний конечным. Мы покажем, что случай бесконечного пучка может быть очень просто изучен методом производящих функций.

Те рассуждения, которые привели нас в § 20 к системе ($\mathcal{E}$) уравнений Эрланга, сохраняются почти полностью и в случае бесконечного пучка. Различие состоит, очевидно, лишь в том, что группа уравнений системы ($\mathcal{E}$), построенная нами в § 20 для $0<k<n$, теперь имеет место для любого $k>0$, последнее же из уравнений системы ($\mathcal{E}$) отпадает совсем. Таким образом, мы получаем для вероятностей различных состояний в случае бесконечного пучка систему уравнений

$$\left.\begin{aligned} &\mathsf{P}_0'(t) = -\lambda\mathsf{P}_0(t) + \mathsf{P}_1(t);\\ &\mathsf{P}_k'(t) = \lambda\mathsf{P}_{k-1}(t) - (\lambda + k)\,\mathsf{P}_k(t) + (k+1)\,\mathsf{P}_{k+1}(t) \quad (k>0). \end{aligned}\right\} \qquad (\mathcal{E}^*)$$

Эту систему можно считать более простой, чем система ($\mathcal{E}$), так как здесь для всех $k>0$ мы имеем уравнения одинакового типа; именно это обстоятельство и позволяет применить к решению системы ($\mathcal{E}^*$) метод производящих функций [к системе ($\mathcal{E}$) он непосредственно применен быть не может].

Положим

$$\Phi(t, x) = \sum_{k=0}^{\infty} \mathsf{P}_k(t)\, x^k$$

[ряд, очевидно, сходится при $|x|\leqslant 1$ и любом t]; положим еще для $k>0$

$$\lambda\mathsf{P}_{k-1}(t) - k\mathsf{P}_k(t) = Q_k(t);$$

тогда система уравнений ($\mathcal{E}^*$) получает более краткий вид:

$$\left.\begin{aligned} &\mathsf{P}_0'(t) = -Q_1(t);\\ &\mathsf{P}_k'(t) = Q_k(t) - Q_{k+1}(t) \quad (k>0). \end{aligned}\right\}$$

Отсюда

$$\frac{\partial\Phi}{\partial t} = \sum_{k=0}^{\infty} \mathsf{P}_k'(t)\, x^k = -Q_1(t) + \sum_{k=1}^{\infty} [Q_k(t) - Q_{k+1}(t)]\, x^k =$$

$$= (x-1)\sum_{k=1}^{\infty} Q_k(t)\, x^{k-1} =$$

$$= (x-1)\left\{\lambda\sum_{k=1}^{\infty} \mathsf{P}_{k-1}(t)\, x^{k-1} - \sum_{k=1}^{\infty} k\mathsf{P}_k(t)\, x^{k-1}\right\} = (x-1)\left\{\lambda\Phi - \frac{\partial\Phi}{\partial x}\right\},$$

или

$$\frac{\partial\Phi}{\partial t} + (x-1)\frac{\partial\Phi}{\partial x} - \lambda(x-1)\,\Phi = 0. \qquad (22.1)$$

Это простое уравнение с частными производными первого порядка и послужит нам для определения функции $\Phi(t, x)$. Прежде всего мы его еще несколько упростим преобразованием неизвестной функции. Положим

$$e^{\lambda(x-1)(1-e^{-t})} = G(t, x)$$

и

$$\Phi(t, x) = G(t, x) F(t, x),$$

где $F(t, x)$ — новая неизвестная функция. Мы будем иметь

$$\frac{\partial \Phi}{\partial t} = G\frac{\partial F}{\partial t} + FG\lambda(x-1)e^{-t};$$

$$\frac{\partial \Phi}{\partial x} = G\frac{\partial F}{\partial x} + FG\lambda(1-e^{-t}),$$

вследствие чего

$$\frac{\partial \Phi}{\partial t} + (x-1)\frac{\partial \Phi}{\partial x} - \lambda(x-1)\Phi = G\left\{\frac{\partial F}{\partial t} + \lambda(x-1)e^{-t}F + (x-1)\frac{\partial F}{\partial x} + \right.$$
$$\left. + \lambda(x-1)(1-e^{-t})F - \lambda(x-1)F\right\} = G\left\{\frac{\partial F}{\partial t} + (x-1)\frac{\partial F}{\partial x}\right\},$$

и уравнение (22.1) равносильно уравнению

$$\frac{\partial F}{\partial t} + (x-1)\frac{\partial F}{\partial x} = 0. \tag{22.2}$$

Положим теперь

$$(x-1)e^{-t} = L(t, x)$$

и составим функциональный определитель

$$\frac{D(F, L)}{D(t, x)} = e^{-t}\left\{\frac{\partial F}{\partial t} + (x-1)\frac{\partial F}{\partial x}\right\};$$

уравнение (22.2) равносильно уравнению

$$\frac{D(F, L)}{D(t, x)} = 0,$$

и значит, общим решением его служит соотношение

$$F = R(L) = R[(x-1)e^{-t}],$$

где R — произвольная дифференцируемая функция своего аргумента. Для искомой функции Φ мы отсюда находим выражение

$$\Phi(t, x) = e^{\lambda(x-1)(1-e^{-t})} R(x-1)e^{-t}]. \tag{22.3}$$

Это — общее решение уравнения (22.1) Для решения нашей задачи мы должны с помощью начальных данных определить вид функции R.

§ 23. Решение задачи

Пусть в начальный момент $t=0$ мы имеем

$$\mathsf{P}_k(0)=a_k \quad (k=0, 1, \ldots);$$

в частности, если известно, что при $t=0$ система находится в состоянии i, то $a_i=1$, $a_k=0$ $(k \neq i)$; во всех случаях $\sum_{k=0}^{\infty} a_k=1$. Полагая тогда $t=0$ в выражении (22.3) функции $\Phi(t, x)$, мы находим

$$\Phi(0, x)=\sum_{k=0}^{\infty} \mathsf{P}_k(0) x^k=\sum_{k=0}^{\infty} a_k x^k=R(x-1),$$

откуда

$$R(z)=\sum_{k=0}^{\infty} a_k (z+1)^k.$$

В частности,

$$R[(x-1) e^{-t}]=\sum_{k=0}^{\infty} a_k [1+(x-1) e^{-t}]^k, \tag{23.1}$$

и для производящей функции мы получаем уже окончательное выражение

$$\Phi(t, x)=e^{\lambda(x-1)(1-e^{-t})} \sum_{k=0}^{\infty} a_k [1+(x-1) e^{-t}]^k\,{}^*.$$

Наша задача состоит в том, чтобы, пользуясь этим выражением, найти пределы вероятностей $\mathsf{P}_k(t)$ при $t \to \infty$.

Как показывает соотношение (23.1), величина $R[(x-1)e^{-t}]$ может быть представлена в виде (сходящегося при $t \geqslant 0$, $|x| \leqslant 1$) ряда по степеням x:

$$R[(x-1) e^{-t}]=\sum_{k=0}^{\infty} b_k(t) x^k.$$

Убедимся теперь, что при $t \to \infty$ мы имеем

$$b_0(t) \to 1, \quad b_k(t) \to 0 \quad (k>0).$$

В самом деле, в силу (23.1)

$$\sum_{k=0}^{\infty} b_k(t) x^k=R[(x-1) e^{-t}]=\sum_{k=0}^{\infty} a_k (x e^{-t}+1-e^{-t})^k; \tag{23.2}$$

---

* Этот ряд заведомо сходится при $t \geqslant 0$, $|x| \leqslant 1$. В самом деле, легко убедиться, что при этих условиях и $|1+(x-1) e^{-t}| \leqslant 1$.

так как $a_k \geqslant 0$, $e^{-t} > 0$ и $1 - e^{-t} \geqslant 0$, то сравнение левой части с правой прежде всего показывает, что все $b_k(t) \geqslant 0$. Далее, полагая в этом равенстве $x = 0$, мы находим

$$b_0(t) = \sum_{k=0}^{\infty} a_k (1 - e^{-t})^k,$$

откуда при $t \to \infty$

$$b_0(t) \to \sum_{k=0}^{\infty} a_k = 1;$$

наконец, полагая в (23.2) $x = 1$, мы находим при любом $t \geqslant 0$

$$\sum_{k=0}^{\infty} b_k(t) = \sum_{k=0}^{\infty} a_k = 1,$$

откуда в силу предыдущего $b_k(t) \to 0$ $(k > 0,\ t \to \infty)$, и наше утверждение доказано.

В силу (22.3) мы теперь имеем

$$\Phi(t, x) = \sum_{n=0}^{\infty} P_n(t) x^n = R[(x-1)e^{-t}]\, e^{\lambda(x-1)(1-e^{-t})} =$$
$$= \sum_{k=0}^{\infty} b_k(t) x^k e^{-\lambda(1-e^{-t})} \sum_{r=0}^{\infty} \frac{\lambda^r (1-e^{-t})^r}{r!} x^r,$$

откуда

$$\mathsf{P}_n(t) = e^{-\lambda(1-e^{-t})} \sum_{k=0}^{n} b_k(t) \frac{\lambda^{n-k} (1-e^{-t})^{n-k}}{(1-k)!} \quad (n \geqslant 0).$$

Если теперь при фиксированном n заставить t безгранично возрастать, то, как мы видели выше, $b_0(t) \to 1$, $b_k(t) \to 0$ $(k > 0)$, а потому

$$\mathsf{P}_n(t) \to e^{-\lambda} \frac{\lambda^n}{n!} (t \to \infty,\ n = 0, 1, 2, \ldots),$$

что полностью решает поставленную задачу. Мы приходим при этом к тем самым формулам, которые мы в начале § 22 в порядке гипотезы получили предельным переходом из формул Эрланга для конечного пучка.

§ 24. Поток с переменным параметром

Методы, изложенные нами в § 22 и 23, позволяют легко решить задачу Эрланга для бесконечного пучка и в том случае, когда параметр поступающего потока вызовов меняется с течением времени. Мы уже имели дело с этим случаем в § 5. Мы будем в дальнейшем и здесь обозначать через $\lambda(t)$ «мгновенное значение» параметра в момент t, определяемое формулой (5.1).

Так как уравнения системы ($\mathcal{E}^*$) § 22 имеют чисто локальную природу (относятся к некоторому определенному моменту времени t), то проведенный нами вывод этих уравнений остается в полной силе и в случае переменного параметра, причем только на место постоянного λ становится «мгновенное значение» $\lambda(t)$ этого параметра, вообще говоря, различное в различные моменты времени t. Таким образом, в качестве исходной системы уравнений для определения функций $P_k(t)$ мы имеем

$$\left.\begin{aligned} P_0'(t) &= -\lambda(t)P_0(t) + P_1(t); \\ P_k'(t) &= \lambda(t)P_{k-1}(t) - [\lambda(t)+k]P_k(t) + (k+1)P_{k+1}(t) \quad (k>0). \end{aligned}\right\} \tag{24.1}$$

Полагая

$$\Phi(t, x) = \sum_{k=0}^{\infty} P_k(t)x^k,$$

мы из системы (24.1) в точности как в § 22 находим для производящей функции $\Phi(t, x)$ уравнение с частными производными

$$\frac{\partial\Phi}{\partial t} + (x-1)\frac{\partial\Phi}{\partial x} - \lambda(t)(x-1)\Phi = 0, \tag{24.2}$$

отличающееся от уравнения (22.1) только тем, что на месте постоянного параметра λ стоит теперь функция $\lambda(t)$. Это отличие, несущественное для вывода уравнения (24.2), в значительной мере влияет на его решение, заставляя нас обратиться к другой замене искомой функции.

Положим во всем дальнейшем

$$e^{-t}\int_0^t e^u\lambda(u)\,du = \Lambda(t);$$

$$e^{(x-1)\Lambda(t)} = G(t, x);$$

$$\Phi(t, x) = G(t, x)F(t, x),$$

где $F(t, x)$ — новая неизвестная функция. Производя эту замену в (24.2), мы легко находим для функции $F(t, x)$ уравнение

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

в точности совпадающее с уравнением (22.2). В § 22 мы видели, что общим решением этого уравнения служит

$$F(t, x) = R[(x-1)e^{-t}],$$

где R — произвольная дифференцируемая функция своего аргумента. Отсюда

$$\Phi(t, x) = e^{(x-1)\Lambda(t)}R[(x-1)e^{-t}]. \tag{24.3}$$

Теперь мы должны определить вид функции R с помощью начальных данных $\mathsf{P}_k(0) = a_k$ $(k = 0, 1, 2, \ldots)$. Мы находим, как в § 23,

$$R(z) = \sum_{k=0}^{\infty} a_k (z+1)^k,$$

откуда, в частности,

$$R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} a_k [1 + (x-1)e^{-t}]^k, \tag{24.4}$$

где ряд заведомо сходится при $t \geqslant 0$, $|x| \leqslant 1$. Отсюда с помощью (24.3) мы получаем окончательное выражение производящей функции

$$\Phi(t, x) = e^{(x-1)\Lambda(t)} \sum_{k=0}^{\infty} a_k [1 + (x-1)e^{-t}]^k. \tag{24.5}$$

В § 23 мы полагали

$$R[(x-1)e^{-t} = \sum_{k=0}^{\infty} b_k(t) x^k, \tag{24.6}$$

и убедились, что при $t \to \infty$

$$b_0(t) \to 1, \quad b_k(t) \to 0, \quad (k > 0).$$

Эти выводы сохраняют силу и в нашем новом случае, так как новая функция $R(z)$ ничем не отличается от прежней.

Мы имеем в силу (24.5), (24.4) и (24.6)

$$\Phi(t, x) = \sum_{n=0}^{\infty} \mathsf{P}_n(t) x^n = e^{(x-1)\Lambda(t)} \sum_{k=0}^{\infty} b_k(t) x^k = e^{-\Lambda(t)} \sum_{r=0}^{\infty} \frac{[\Lambda(t)]^r}{r!} x^r \sum_{k=0}^{\infty} b_k(t) x^k =$$
$$= e^{-\Lambda(t)} \sum_{n=0}^{\infty} \left\{ \sum_{r=0}^{n} \frac{[\Lambda(t)]^r}{r!} b_{n-r}(t) \right\} x^n,$$

и следовательно,

$$\mathsf{P}_n(t) = e^{-\Lambda(t)} \sum_{r=0}^{n} \frac{[\Lambda(t)]^r}{r!} b_{n-r}(t) \quad (n = 0, 1, 2, \ldots).$$

Допустим теперь, что при $t \to \infty$ параметр $\lambda(t)$ остается ограниченным;

$$\lambda(t) < \Lambda \qquad (t \geqslant 0).$$

Тогда и

$$\Lambda(t) = e^{-t} \int_0^t e^u \lambda(u)\, du < \Lambda \qquad (t \geqslant 0);$$

поэтому для любого постоянного n при $t \to \infty$ в сумме

$$\sum_{r=0}^{n} \frac{[\Lambda(t)]^r}{r!} b_{n-r}(t)$$

все члены, кроме последнего ($r = n$), стремятся к нулю, в то время как последний бесконечно мало отличается от $[\Lambda(t)]^n / n!$. Поэтому при $t \to \infty$ и постоянном n

$$\mathsf{P}_n(t) - e^{-\Lambda(t)} \frac{[\Lambda(t)]^n}{n!} \to 0,$$

т. е. закон распределения $\mathsf{P}_n(t)$ безгранично приближается к закону Пуассона с (переменным) параметром

$$\Lambda(t) = e^{-t} \int_0^t e^u \lambda(u)\, du.$$

Этим поставленная задача решена.

§ 25. Бесконечный пучок при произвольном законе распределения длин разговоров

В § 22—24 мы убедились, что для задачи Эрланга бесконечный пучок в известном смысле представляет собой более простой объект исследования, чем конечный (в частности, в § 24 мы до конца рассмотрели случай входящего потока с переменным параметром — задача, которая, насколько нам известно, еще не решена для конечного пучка). В нашем изложении сравнительная простота трактовки случая бесконечного пучка все время связывалась с возможностью применения метода производящих функций; при этом мы ради соблюдения возможно тесной аналогии с теорией конечного пучка исходили всегда из системы дифференциальных уравнений Эрланга. На самом деле, задача Эрланга для бесконечного пучка представляет собой чрезвычайно простую проблему, которая может быть решена и вполне элементарными средствами; при этом оказывается, что показательный закон распределения длин разговоров, служивший важной предпосылкой в методе уравнений Эрланга, при элементарной трактовке задачи без существенных усложнений может быть заменен любым другим законом. Ради этого важного обобщения мы и остановимся в настоящем параграфе на элементарном выводе формул Эрланга для бесконечного пучка.

Мы сохраним все предпосылки § 22 с той разницей, что вероятность $F(x)$ для наудачу выбранного разговора иметь длину $> x$ мы будем предполагать произвольной невозрастающей функцией, подчиненной только требованиям

$$F(0) = 1, \;\; F(+\infty) = 0, \;\; -\int_0^\infty x\, dF(x) = \int_0^\infty F(x)\, dx = 1,$$

из которых последнее выражает собой соглашение принимать за единицу времени среднюю длительность разговора. Будем, как прежде, обозначать через $\mathsf{P}_k(t)$ вероятность того, что в момент $t>0$ ведется k разговоров (или, что то же, занято k линий). Наша задача — показать, что

$$\lim_{t\to\infty} \mathsf{P}_k(t) = e^{-\lambda}\frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \ldots).$$

Пусть $X = (x_1, x_2, \ldots, x_r)$ — произвольный r-мерный вектор, принадлежащий области D_r $(0 < x_1 < x_2 < \ldots < x_r < t)$ r-мерного пространства. Условимся называть «гипотезой (r, X)» событие, состоящее в том, что в промежутке $(0, t)$ происходит r вызовов и моменты t_i $(0 < t_1 < < t_2 < \ldots < t_r < t)$ этих вызовов удовлетворяют неравенствам

$$x_i < t_i < x_i + dx_i \quad (1 \leqslant i \leqslant r).$$

Так как поток вызовов — простейший с параметром λ, то вероятность гипотезы (r, X) с точностью до бесконечно малых высших порядков выражается при малых dx_i формулой

$$\mathsf{P}(r, X) = \lambda e^{-\lambda x_1} dx_1 \lambda e^{-\lambda(x_2 - x_1)} dx_2 \ldots \lambda e^{-\lambda(x_r - x_{r-1})} dx_r e^{-\lambda(t - x_r)},$$

где последний множитель есть вероятность того, что между моментами x_r и t вызовов не поступает. Отсюда

$$\mathsf{P}(r, X) = e^{-\lambda t}\lambda^r dx_1 dx_2 \ldots dx_r \tag{25.1}$$

(и, в частности, не зависит от вектора X). Пусть $\mathsf{P}_{k,(r,X)}(t)$ — условная вероятность застать k разговоров в момент t, если имеет место гипотеза (r, X). Очевидно, $\mathsf{P}_{k,(r,X)}(t)$ может быть отличной от нуля лишь при $r \geqslant k$. Так как для разговора, начавшегося в момент x $(0 < x < t)$, вероятность не закончиться к моменту t равна $F(t - x)$, то

$$\mathsf{P}_{k,(r,X)}(t) = \sum_C \prod_{i=1}^{k} F(t - x_{s_i}) \prod_{\substack{s=1\\ s \neq s_i}}^{r} [1 - F(t - x_s)], \tag{25.2}$$

где C — любое сочетание $(s_1, s_2, \ldots, s_k)$ по k чисел ряда $(1, 2, \ldots, r)$ и где суммирование производится по всем таким сочетаниям.

По формуле полной вероятности мы имеем

$$\mathsf{P}_k(t) = \sum_{r, X} \mathsf{P}_r(r, X)\mathsf{P}_{k,(r,X)}(t),$$

где по r мы должны суммировать от k до ∞, а по X интегрировать по области D_r. Поэтому мы находим в силу (25.1) и (25.2)

$$\mathsf{P}_k(t) = e^{-\lambda t}\sum_{r=k}^{\infty}\lambda^r \int\limits_{D_r} \sum_C \prod_{i=1}^{k} F(t - x_{s_i}) \prod_{\substack{s=1\\ s\neq s_i}}^{r} [1 - F(t - x_s)]\, dx_1 \ldots dx_r.$$

Подинтегральная функция здесь, очевидно, симметрична относительно переменных $x_1, x_2, \ldots, x_r$. Интеграл поэтому не изменится, если в определении области D_r мы заменим порядок $x_1 < x_2 < \ldots < x_r$ переменных интеграции любым другим порядком; а так как всех таких порядков возможно $r!$ и так как соединение всех получаемых таким образом областей D_r дает r-мерный куб $K_r\,[0 < x_i < t,\ i = 1, 2, \ldots, r]$, то мы можем интегрировать по всему этому кубу с последующим делением на $r!$. Таким образом, мы получаем

$$\mathsf{P}_k(t) = e^{-\lambda t}\sum_{r=k}^{\infty}\frac{\lambda^r}{r!}\sum_{C}\int\limits_{K_r}\prod_{i=1}^{k}F(t - x_{s_i})\prod_{\substack{s=1\\ s\neq s_i}}^{r}[1 - F(t - x_s)]\,dx_1\ldots dx_r.$$

Здесь интеграл по кубу K_r распадается на произведение r простых интегралов и равен

$$\left\{\int\limits_0^t F(t-u)\,du\right\}^k\left\{\int\limits_0^t [1 - F(t-u)]\,du\right\}^{r-k} = \left\{\int\limits_0^t F(z)\,dz\right\}^k\left\{\int\limits_0^t [1 - F(z)]\,dz\right\}^{r-k}.$$

Полагая

$$\int\limits_t^{\infty} F(z)\,dz = \varepsilon(t),$$

мы имеем

$$\int\limits_0^t F(z)\,dz = 1 - \varepsilon(t), \quad \int\limits_0^t [1 - F(z)]\,dz = t - 1 + \varepsilon(t),$$

и следовательно [так как число сочетаний C равно $\binom{r}{k}$],

$$\mathsf{P}_k(t) = e^{-\lambda t}\sum_{r=k}^{\infty}\frac{\lambda^r}{r!}\binom{r}{k}[1 - \varepsilon(t)]^k[t - 1 + \varepsilon(t)]^{r-k} =$$

$$= e^{-\lambda t}\frac{\lambda^k}{k!}[1 - \varepsilon(t)]^k\sum_{r=k}^{\infty}\frac{\lambda^{r-k}}{(r-k)!}[t - 1 + \varepsilon(t)]^{r-k} =$$

$$= e^{-\lambda t}\frac{\lambda^k}{k!}[1 - \varepsilon(t)]^k e^{\lambda[t-1+\varepsilon(t)]} =$$

$$= e^{-\lambda}\frac{\lambda^k}{k!}[1 - \varepsilon(t)]^k e^{\lambda\varepsilon(t)}.$$

Так как $\varepsilon(t) \to 0$ при $t \to \infty$, то отсюда

$$\lim_{t\to\infty}\mathsf{P}_k(t) = e^{-\lambda}\frac{\lambda^k}{k!}\ (k = 0, 1, 2, \ldots),$$

что и требовалось доказать*.

* В этом доказательстве мы неявно предполагали, что поток вызовов начинается с момента 0, т. е. что $\mathsf{P}_0(0) = 1$, $\mathsf{P}_k(0) = 0$ $(k > 0)$. Доказательство лишь немного усложнилось бы при произвольных начальных данных.

Глава 8

Задача Пальма

§ 26. Постановка задачи

Для решения задачи Эрланга в гл. 6 и 7 нам было безразлично, рассматривается ли данный пучок линий как упорядоченный или нет. В настоящей главе, напротив, мы будем иметь дело с задачей, которая имеет смысл лишь для упорядоченного пучка. Мы будем, таким образом, все время исходить из предположения, что линии данного пучка перенумерованы и что каждый поступающий вызов занимает линию с наименьшим номером из числа тех, которые свободны в момент его поступления. Очевидно, что при таком порядке средняя загруженность будет различной для различных линий: наиболее загруженной окажется первая линия, за ней вторая, и т. д.

С другой стороны, задачи, рассматриваемые в настоящей главе, таковы, что уже по смыслу их постановки безразлично, имеем ли мы дело с конечным или бесконечным пучком; таким образом, это различение, столь важное, как мы видели, для задачи Эрланга, для целей настоящей главы совершенно несущественно.

Во всем остальном мы сохраняем предпосылки предшествующих глав. Входящий поток вызовов мы предполагаем простейшим с параметром λ. Длительности разговоров предполагаются не зависящими ни друг от друга, ни от каких-либо данных о поступлении вызовов и распределенными по показательному закону со средним значением 1.

Условимся обозначать через L_r линию с номером r. Для всех рассуждений настоящей главы имеет основное значение тот простой и самоочевидный факт, что совокупность линий $L_1, L_2, \ldots, L_r$ (при любом r, не превосходящем общего числа линий пучка) мы можем рассматривать как самостоятельный полнодоступный пучок. Каждый вызов, «потерянный» на этом пучке (т. е. заставший первые r линий занятыми), поступает на линию L_{r+1} (если, конечно, таковая имеется), и обратно — для того, чтобы вызов поступил на L_{r+1}, необходимо, чтобы он был потерян на пучке $(L_1, L_2, \ldots, L_r)$. Вероятность потери на этом пучке есть доля времени, в течение которой все линии $L_1, L_2, \ldots, L_r$ заняты; она, очевидно, совершенно не зависит от того, существуют ли еще линии с более высокими номерами и сколько таких линий; она может быть вычислена по формуле Эрланга для вероятности потери на пучке из r линий и равна

$$\frac{\frac{\lambda^r}{r!}}{\sum_{i=1}^{r} \frac{\lambda^i}{i!}}.$$

В частности, при $r=1$ мы получаем для вероятности потери вызова на линии L_1 выражение

$$\frac{\lambda}{1+\lambda}, \tag{26.1}$$

а при $r=2$ — для вероятности потери на пучке (L_1, L_2) выражение

$$\frac{\frac{\lambda^2}{2}}{1+\lambda+\frac{\lambda^2}{2}} = \frac{\lambda^2}{\lambda^2+2\lambda+2}. \tag{26.2}$$

Но потерю вызова на пучке (L_1, L_2) мы можем рассматривать как двойное событие: 1) потеря на L_1 и 2) потеря на L_2. Вероятность первого из этих событий равна $\lambda/(1+\lambda)$. Чтобы найти условную вероятность второго события при условии, что первое состоялось, заметим, что это есть вероятность потери на L_2 для вызова, потерянного на L_1 (или, что то же, поступившего на L_2). Если обозначить через λ' интенсивность потока вызовов, поступающих на L_2, то искомая условная вероятность второго события будет поэтому по формуле (26.1) равна $\lambda'/(1+\lambda')$ (так как для этого потока L_2 служит первой линией). Но λ' есть число вызовов, теряющихся на L_1 в единицу времени; так как поступает на L_1 в единицу времени λ вызовов, а доля потерь составляет $\lambda/(1+\lambda)$, то

$$\lambda' = \lambda\frac{\lambda}{1+\lambda} = \frac{\lambda^2}{1+\lambda},$$

и искомая условная вероятность второго события равна

$$\frac{\lambda'}{1+\lambda'} = \frac{\lambda^2}{1+\lambda+\lambda^2}.$$

Для вероятности потери на пучке (L_1, L_2) мы таким образом получаем выражение

$$\frac{\lambda}{1+\lambda}\,\frac{\lambda^2}{1+\lambda+\lambda^2} = \frac{\lambda^3}{(1+\lambda)(1+\lambda+\lambda^2)}, \tag{26.3}$$

отличное от непосредственно даваемого формулами Эрланга выражения (26.2). Следовательно, формула (26.3) для вероятности потери на пучке (L_1, L_2) является ошибочной*. Анализируя цепь рассуждений, приведших нас к этой формуле, мы легко находим источник ошибки. Обозначив через λ' интенсивность потока вызовов, поступающих на L_2, мы приняли, следуя формуле Эрланга (26.1), вероятность потери на L_2 для поступающего на эту линию вызова равной $\lambda'/(1+\lambda')$; тем самым мы неявно допустили, что поток вызовов, поступающих на L_2, является *простейшим,* так как формулы Эрланга установлены нами лишь в этом предположении. То, что мы пришли к неверному

* Указанием на этот поучительный «парадокс» я обязан В. К. Лезерсону.

результату, доказывает, что это предположение было ошибочным. Мы можем, таким образом, считать установленным, что ***поток вызовов, поступающих на*** L_2 (*или, что то же, теряемых на* L_1), ***не есть простейший поток.*** Тем более, конечно, у нас нет оснований ожидать, чтобы простейшими оказались потоки вызовов, поступающих на L_3, $L_4, \ldots$ Этот отрицательного характера вывод поучителен тем, что ясно показывает, насколько важно уже для решения самых элементарных задач не ограничиваться изучением одних только простейших поступающих потоков.

Детальное, до конца идущее исследование природы потока вызовов, поступающих на любую линию L_r данного упорядоченного полнодоступного пучка, составляет собой теоретически интересную и практически важную задачу, решению которой и будет посвящена настоящая глава. Все основные результаты в этом направлении были получены Пальмом [8] в 1943 г.

§ 27. Элементарные расчеты

Как мы уже указали в предыдущем параграфе, для вызова, поступающего на данный пучок (или, что то же, на линию L_1), вероятность оказаться потерянным на пучке $(L_1, L_2, \ldots, L_r)$ равна по формуле Эрланга

$$E_r = \frac{\frac{\lambda^r}{r!}}{\sum\limits_{k=0}^{r} \frac{\lambda^k}{k!}} \qquad (r = 1, 2, \ldots).$$

Очевидно, что вызов теряется на линии L_r тогда и только тогда, если он теряется на пучке $(L_1, L_2, \ldots, L_r)$. Поэтому можно также сказать, что число E_r выражает собой вероятность потери на линии L_r; однако при этом необходимо отчетливо иметь в виду, что речь идет о вероятности потери на L_r ***для вызова, поступающего на*** L_1; вероятность же Π_r потери на L_r ***для вызова, поступающего на*** L_r, имеет другую величину, которую мы теперь должны постараться найти. Мы можем при этом допустить, что $r > 1$, так как, очевидно, $\Pi_1 = E_1$.

Вероятность E_r того, что вызов, поступивший на L_1, будет потерян на L_r, очевидно, может быть представлена в виде произведения двух множителей: вероятности E_{r-1} того, что он будет потерян на L_{r-1}, и условной вероятности его потери на L_r, если известно, что он потерян на L_{r-1} (или, что то же, поступил на L_r). Но эта условная вероятность и есть Π_r; поэтому

$$E_r = E_{r-1}\Pi_r,$$

или

$$\Pi_r = \frac{E_r}{E_{r-1}} \qquad (r > 1).$$

Несмотря на свою кажущуюся простоту, эта формула для расчетов неудобна тем, что содержит одновременно E_r и E_{r-1}. Поэтому удобнее заменить ее формулой

$$\Pi_r = \frac{\lambda}{r + \lambda E_{r-1}} \qquad (r > 1), \tag{27.1}$$

которую мы сейчас докажем. Положим для краткости

$$\sum_{k=0}^{q} \frac{\lambda^k}{k!} = S_q \qquad (q \geqslant 0),$$

так что

$$E_r = \frac{1}{S_r} \frac{\lambda^r}{r!} \qquad (r \geqslant 1);$$

следовательно, при $r > 1$

$$\frac{1}{\Pi_r} = \frac{E_{r-1}}{E_r} = \frac{r}{\lambda} \frac{S_r}{S_{r-1}} = \frac{r}{\lambda}\left[1 + \frac{\lambda^r/r!}{S_{r-1}}\right] = \frac{1}{\lambda}(r + \lambda E_{r-1}),$$

а это и есть (27.1). При этом необходимо иметь в виду, что λ в формуле (27.1) означает интенсивность первичного потока вызовов, поступающих на L_1.

Представляет существенный интерес сравнить между собой вероятности потери на различных линиях при одинаковой интенсивности поступающих на них потоков. Произведенные подсчеты во всех случаях показывают, что эта вероятность возрастает с номером линии. В своей цитированной нами работе Пальм утверждает, что это непосредственно вытекает из формулы (27.1). Мы не видим, однако, как можно было бы это показать. Более того, нам вообще неизвестно, верно ли это утверждение в его общей формулировке. Нам удалось доказать в этом направлении только следующее значительно более скромное предложение.

Теорема. *Вероятность потери на L_r при $r > 1$ всегда больше, чем вероятность потери на L_1, если поступающие на эти линии потоки имеют одинаковую интенсивность.*

Доказательство. Если интенсивность поступающего на L_1 потока равна λ, то вероятность поступить на L_r для вызова, поступившего на L_1, равна E_{r-1} (так как это есть вероятность потерпеть потерю на линиях $L_1, L_2, \ldots, L_{r-1}$). Поэтому среди λ вызовов, поступающих в среднем на L_1 в единицу времени, на L_r будет в среднем поступать λE_{r-1} вызовов, т. е. поток вызовов, поступающих на L_r, будет иметь интенсивность λE_{r-1}. При этом, как мы знаем, вероятность потери на L_r для вызовов этого потока равна

$$\Pi_r = \frac{\lambda}{r + \lambda E_{r-1}} \qquad (r > 1).$$

Если бы поток той же интенсивности λE_{r-1} падал на L_1, то вероятность потери на L_1 для вызовов этого потока согласно формуле (26.1) была бы

$$\frac{\lambda E_{r-1}}{1+\lambda E_{r-1}} \quad (r>1).$$

Для доказательства нашей теоремы достаточно поэтому убедиться, что при $r>1$ и $\lambda>0$ мы всегда имеем

$$\frac{\lambda E_{r-1}}{1+\lambda E_{r-1}} < \frac{\lambda}{r+\lambda E_{r-1}},$$

или

$$r+\lambda E_{r-1} < \lambda + \frac{1}{E_{r-1}},$$

или, наконец,

$$\lambda E_{r-1} + r - \lambda < \frac{1}{E_{r-1}} \quad (r>1). \tag{27.2}$$

Пусть сначала $r=2$, $E_{r-1}=E_1=\frac{\lambda}{1+\lambda}$. Тогда (27.2) получает вид:

$$\frac{\lambda^2}{1+\lambda}+2-\lambda<\frac{1+\lambda}{\lambda},$$

что равносильно

$$\frac{\lambda^2}{1+\lambda}+1-\lambda<\frac{1}{\lambda},$$

или

$$\frac{1}{1+\lambda}<\frac{1}{\lambda},$$

и, следовательно, выполняется при любом $\lambda>0$.

Пусть теперь неравенство (27.2) справедливо при каком-нибудь $r>1$ (и любом $\lambda>0$); покажем, что в таком случае оно останется верным и при замене r на $r+1$, т. е. что при любом $\lambda>0$ будет иметь место неравенство

$$\lambda E_r + r + 1 - \lambda < \frac{1}{E_r}.$$

В силу (27.1) [помня, что $\Pi_r = E_r / E_{r-1}$] мы имеем

$$E_r = \frac{\lambda E_{r-1}}{r+\lambda E_{r-1}}.$$

Отсюда легко находим

$$\lambda E_r + r + 1 - \lambda - \frac{1}{E_r} = r\left\{\frac{\lambda E_{r-1}+r-\lambda}{\lambda E_{r-1}+r} - \frac{1}{\lambda E_{r-1}}\right\} = \frac{rH_r(\lambda)}{\lambda E_{r-1}(\lambda E_{r-1}+r)},$$

где

$$H_r(\lambda) = \lambda E_{r-1}(\lambda E_{r-1} + r - \lambda) - (\lambda E_{r-1} + r) =$$
$$= \lambda E_{r-1}\left(\lambda E_{r-1} + r - \lambda - \frac{1}{E_{r-1}}\right) - (\lambda E_{r-1} + r - \lambda) =$$
$$= (\lambda E_{r-1} - 1)(\lambda E_{r-1} + r - \lambda) - \lambda.$$

Наша теорема будет доказана, если мы убедимся, что $H_r(\lambda) < 0$ при любом $\lambda > 0$. Положим для краткости

$$\lambda E_{r-1} - 1 = A, \quad \lambda E_{r-1} + r - \lambda = B,$$

так что

$$H_r(\lambda) = AB - \lambda.$$

Мы можем допустить, что $AB > 0$, так как в противном случае тривиальным образом $H_r(\lambda) < 0$. Если $A > 0$ и $B > 0$, то, так как в силу (27.2) $B < 1/E_{r-1}$, мы имеем

$$AB < \frac{\lambda E_{r-1} - 1}{E_{r-1}} < \lambda;$$

если же $A < 0$, $B < 0$, то, очевидно,

$$|A| < 1, \quad |B| = -B = \lambda - r - \lambda E_{r-1} < \lambda, \quad AB = |A||B| < \lambda.$$

Таким образом, $H_r(\lambda) < 0$ во всех случаях, и наша теорема доказана.

§ 28. Основная теорема Пальма

Мы убедились в § 26, что если на линию L_1 поступает простейший поток вызовов, то поток, поступающий на линию L_r ($r > 1$), уже не будет простейшим. Важнейшая основная теорема теории Пальма состоит в том, что при этом *на любую линию L_r поступает поток типа* P, *т. е. стационарный, ординарный и с ограниченным последействием.* От простейшего такой поток отличается, следовательно, только тем, что требование отсутствия последействия заменяется более общим требованием ограниченности последействия.

Для доказательства этой теоремы нам не понадобится никаких расчетов; достаточно лишь более внимательно вглядеться в картину происходящего. Так как для L_1 утверждение теоремы тривиально, то надо только показать, что, если оно верно для L_r, оно остается верным и для L_{r+1} ($r = 1, 2, \ldots$); а так как вызовы, поступающие на L_{r+1}, совпадают с вызовами, теряющимися на L_r, то мы должны доказать следующее: *если на линию L_r поступает поток вызовов типа* P, *то потерянные на L_r вызовы также образуют поток типа* P. При такой постановке задачи, очевидно, самое существование линии L_{r+1} является несущественным.

Обозначим для краткости через A поток вызовов, поступающих на L_r, и через B — поток вызовов, теряющихся на L_r. Течение потока

B после произвольно выбранного момента t_0 будет однозначно определено, если станет известно, сколько времени будет еще длиться занимающий линию L_r в момент t_0 разговор, а также каковы моменты поступающих после t_0 вызовов потока A и какова длительность начинаемых этими вызовами разговоров. Но все эти факторы в свою очередь однозначно определяются течением потока A, который есть поток типа Р и, следовательно, стационарен. Поэтому все перечисленные факторы не зависят от выбранного момента t_0, а вместе с ними не зависит от t_0 и дальнейшее течение потока B; другими словами, поток B также стационарен.

Ординарность потока B с самоочевидностью вытекает из того, что он составляет собой часть (ординарного) потока A.

Покажем, наконец, что поток B — с ограниченным последействием. Пусть $t_0 = \tau_0 = 0$; обозначим через t_i $(i = 1, 2, \ldots)$ моменты следующих за t_0 вызовов потока A и через $\tau_1, \tau_2, \ldots$ — моменты следующих за t_0 вызовов потока B и положим

$$\tau_k - \tau_{k-1} = \zeta_k \quad (k = 1, 2, \ldots).$$

Наша задача — показать, что закон распределения величины ζ_{k+1} не зависит от значений величин $\zeta_1, \zeta_2, \ldots, \zeta_k$, или, что то же, от значений величин $\tau_1, \tau_2, \ldots, \tau_k$. Но величина ζ_{k+1} будет однозначно определена, если станут известны:

1) расстояния от τ_k до поступающих на L_r после момента τ_k дальнейших вызовов;

2) остаточная длительность разговора, занимающего линию L_r в момент τ_k;

3) длительности разговоров, занимающих линию L_r после момента τ_k.

Все эти три фактора независимы от значений, принимаемых величинами $\tau_1, \tau_2, \ldots, \tau_k$. Для первого это следует из того, что поток A — с ограниченным последействием; для второго — из показательного закона распределения длин разговоров, а для третьего это самоочевидно. Но так как этими тремя факторами значение случайной величины ζ_{k+1} определяется однозначно, то и эта случайная величина не зависит от величин $\tau_1, \tau_2, \ldots, \tau_k$, или, что то же, от величин $\zeta_1, \zeta_2, \ldots, \zeta_k$. Но это и означает, что поток B — с ограниченным последействием. Таким образом, основная теорема Пальма доказана.

Отсюда, очевидно, следует, что для упорядоченного полнодоступного пучка с простейшим входящим потоком и показательным распределением длин разговоров поток вызовов, поступающих на любую линию L_r этого пучка, представляет собой поток типа Р. Но такой поток (см. § 13) однозначно определяется заданием функции Пальма $\varphi(t)$. Условимся в дальнейшем обозначать через $\varphi_r(t)$ $(r = 0, 1, 2, \ldots)$ функцию Пальма для потока вызовов, поступающих на линию L_{r+1} (так что, в частности, $\varphi_0(t) = e^{-\lambda t}$). Мы видим, что наша задача сводится к определению функции $\varphi_r(t)$ для любого $r > 0$; этому и будет посвящено все дальнейшее.

§ 29. Вывод основной системы уравнений

Пусть в момент t_0 теряется вызов на линии L_r (или, что то же, поступает вызов на линию L_{r+1}). Тогда $\varphi_r(t)$ есть вероятность того, что в промежутке $(t_0, t_0 + t)$ ни один вызов не будет потерян на L (не поступит на L_{r+1}). Но это событие может произойти двумя способами:

(A) В промежутке $(t_0, t_0 + t)$ на L_r не поступит ни одного вызова.

(В) В промежутке $(t_0, t_0 + t)$ на L_r будут поступать вызовы, но ни один из них не будет потерян.

Мы имеем поэтому

$$\varphi_r(t) = \mathsf{P}(A) + \mathsf{P}(B);$$

при этом, в случае $r > 0$, $\mathsf{P}(A) = \varphi_{r-1}(t)$, так как, с одной стороны, вызов, потерянный в момент t_0 на L_r, был потерян и на L_{r-1}, а с другой — поток вызовов, поступающих на L_r, совпадает с потоком вызовов, теряемых на L_{r-1}.

Переходим к определению $\mathsf{P}(B)$. Пусть первый вызов, поступающий на L_r после момента t_0, происходит в промежутке $(t_0 + x, t_0 + x + dx)$; вероятность этого события равна $\varphi_{r-1}(x) - \varphi_{r-1}(x + dx) = -d\varphi_{r-1}(x)$. Для того, чтобы этот вызов не был потерян на L_r, очевидно, необходимо и достаточно, чтобы линия L_r до момента $t_0 + x$ освободилась от того разговора, которым она была занята в момент t_0; вероятность этого события равна $1 - e^{-x}$. Таким образом, вероятность того, что первый после момента t_0 вызов поступит на L_r в промежутке $(t_0 + x, t_0 + x + dx)$ и что этот вызов не будет потерян, равна

$$-(1 - e^{-x})\, d\varphi_{r-1}(x).$$

Мы утверждаем теперь, что если наступило все описанное и если $x < t$, то вероятность того, что в остающемся промежутке $(t_0 + x, t_0 + t)$ не будет потеряно на L_r ни одного вызова, равна $\varphi_r(t - x)$. Это вытекало бы непосредственно из определения функции $\varphi_r(t)$, если бы вызов, поступивший на L_r в промежутке $(t_0 + x, t_0 + x + dx)$, был потерян на этой линии; но на самом деле этот вызов по нашему допущению не теряется, так что наше утверждение требует обоснования. Будет ли вызов, поступивший на L_r в момент $t_0 + x$ [точнее: в промежутке $(t_0 + x, t_0 + x + dx)$], потерян или нет, во всяком случае, раз этот вызов произошел, линия L_r с момента $t_0 + x$ будет занята. Когда она освободится — это в силу показательного закона распределения длин разговоров совершенно не зависит от того, была ли она занята поступившим в момент t_0+x вызовом или была занята ранее (и поступивший в момент t_0+x вызов был потерян). С другой стороны, моменты дальнейших (после t_0+x) поступающих на L_r вызовов зависят от того факта, что такой вызов поступил в момент t_0+x, но совершенно не зависят от судьбы этого вызова (от того, был ли он потерян или нет); не зависят, конечно, от этой судьбы и длины тех разговоров, которые начинаются этими последующими вызовами. Таким образом, ни один из факторов, определяющих собой наличие или отсутствие потерь на L_r в промежутке $(t_0+x,\ t_0+t)$, не зависит от того, какая судьба постигла вызов, поступивший в момент t_0+x. И хотя этот вызов по нашему предположению не был потерян, вероятность того, что в промежутке $(t_0+x,\ t_0+t)$ потерь на линии L_r не будет, такова же, как если бы он был потерян, т. е. равна $\varphi_r(t-x)$. Сопоставляя это с тем, что было установлено ранее, мы приходим к следующему выводу: при $x<t$ вероятность того, что первый после момента t_0 вызов поступит на L_r в промежутке $(t_0+x,\ t_0+x+dx)$ и что между t_0 и t_0+t ни один вызов не будет потерян на L_r, равна

$$-(1-e^{-x})\,d\varphi_{r-1}(x)\,\varphi_r(t-x).$$

Но чтобы получить вероятность события B, мы, очевидно, должны просуммировать все такие вероятности по x от 0 до t. Это дает

$$\mathsf{P}(B)=-\int_0^t(1-e^{-x})\,\varphi_r(t-x)\,d\varphi_{r-1}(x),$$

и следовательно,

$$\varphi_r(t)=\varphi_{r-1}(t)-\int_0^t(1-e^{-x})\,\varphi_r(t-x)\,d\varphi_{r-1}(x)\qquad(r\geqslant 1).\quad(29.1)$$

Это и есть исходная система уравнений теории Пальма. Из (29.1) непосредственно видно, что при любом $t>0$

$$\varphi_r(t)\geqslant\varphi_{r-1}(t)$$

— неравенство, которое является очевидным и само по себе.

§ 30. Преобразование Лапласа

С целью решения основной системы уравнений (29.1) мы теперь заменим искомые функции $\varphi_r(t)$ их преобразованиями Лапласа. Положим

$$\psi_r(t)=\int_0^\infty e^{-tx}\,\varphi_r(x)\,dx\quad(r\geqslant 0).$$

Напишем уравнение (29.1) в виде

$$\varphi_r(x)=\varphi_{r-1}(x)-\int_0^x (1-e^{-z})\,\varphi_r(x-z)\,d\varphi_{r-1}(z),$$

умножим обе части на e^{-tx} и проинтегрируем по x от 0 до ∞. Это дает

$$\psi_r(t)=\psi_{r-1}(t)-\int_0^\infty e^{-tx}\,dx\int_0^x (1-e^{-z})\,\varphi_r(x-z)\,d\varphi_{r-1}(z)=$$

$$=\psi_{r-1}(t)-\int_0^\infty (1-e^{-z})\,e^{-zt}\,d\varphi_{r-1}(z)\int_z^\infty e^{-t(x-z)}\varphi_r(x-z)\,dx=$$

$$=\psi_{r-1}(t)-\int_0^\infty (1-e^{-z})\,e^{-zt}\,d\varphi_{r-1}(z)\int_0^\infty e^{-ty}\varphi_r(y)\,dy=$$

$$=\psi_{r-1}(t)-\psi_r(t)\int_0^\infty (1-e^{-z})\,e^{-zt}\,d\varphi_{r-1}(z). \qquad (30.1)$$

Интеграция по частям легко дает

$$\int_0^\infty e^{-zt}\,d\varphi_{r-1}(z)=-1+t\psi_{r-1}(t),$$

и следовательно,

$$\int_0^\infty e^{-z(t+1)}\,d\varphi_{r-1}(z)=-1+(t+1)\,\psi_{r-1}(t+1);$$

поэтому из (30.1) мы получаем

$$\psi_r(t)=\psi_{r-1}(t)-\psi_r(t)\,[t\psi_{r-1}(t)-(t+1)\,\psi_{r-1}(t+1)],$$

откуда

$$\psi_r(t)=\frac{\psi_{r-1}(t)}{1+t\,\psi_{r-1}(t)-(t+1)\,\psi_{r-1}(t+1)}. \qquad (30.2)$$

Таким образом, для определения функций $\psi_r(t)$ мы получаем простую рекуррентную формулу. Так как $\varphi_0(x)=e^{-\lambda x}$, то

$$\psi_0(t)=\int_0^\infty e^{-(\lambda+t)x}\,dx=\frac{1}{t+\lambda},$$

и соотношение (30.2) позволяет последовательно определить все функции $\psi_r(t)$. В частности, мы непосредственно видим, что все эти функции рациональны. Однако мы не можем удовлетвориться этим, так как для обратного перехода к функциям $\varphi_r(x)$ нам важно более детально знать свойства рациональных дробей $\psi_r(t)$. В частности, для этого обратного перехода существенное значение имеет разложение функций $\psi_r(t)$ на простые дроби и, следовательно, природа и расположение корней их знаменателей. Этими вопросами мы и должны будем теперь заняться.

Заметим еще, что простым преобразованием искомых функций

$$t\psi_r(t) = \chi_r(t) \quad (r = 0, 1, 2, \ldots)$$

мы можем привести рекуррентные формулы (30.2) к более краткому виду:

$$\chi_r(t) = \frac{\chi_{r-1}(t)}{1 + \chi_{r-1}(t) - \chi_{r-1}(t+1)};$$

впрочем в дальнейшем мы этим замечанием пользоваться не будем.

§ 31. Определение функций $\psi_r(t)$

Обозначим через $B_r(t)$ $(r = -1, 0, 1, 2, \ldots)$ многочлен степени $r+1$:

$$B_r(t) = \lambda^{r+1} + \sum_{l=0}^{r} \binom{r+1}{l} t(t+1)\ldots(t+r-l)\lambda^l,$$

где $\sum\limits_{l=0}^{-1} = 0$, и следовательно, $B_{-1}(t) = \lambda^0 = 1$.

Лемма. *Многочлены $B_r(t)$ связаны рекуррентной формулой*

$$B_r(t) = t\,B_{r-1}(t+1) + \lambda\,B_{r-1}(t) \quad (r = 0, 1, \ldots) \tag{31.1}$$

Доказательство. Мы имеем

$$\lambda B_{r-1}(t) = \lambda^{r+1} + \sum_{l=0}^{r-1} \binom{r}{l} t(t+1)\ldots(t+r-1-l)\lambda^{l+1};$$

$$tB_{r-1}(t+1) = t\lambda^r + \sum_{l=0}^{r-1} \binom{r}{l} t(t+1)\ldots(t+r-l)\lambda^l =$$

$$= \sum_{l=0}^{r} \binom{r}{l} t(t+1)\ldots(t+r-l)\lambda^l.$$

Так как $\binom{r}{0} = \binom{r+1}{0}$ и, при $l > 0$, $\binom{r}{l-1} + \binom{r}{l} = \binom{r+1}{l}$, то отсюда

$$\lambda\,B_{r-1}(t) + t\,B_{r-1}(t+1) = \lambda^{r+1} + \sum_{l=1}^{r} \binom{r+1}{l} t(t+1)\ldots(t+r-l)\lambda^l +$$

$$+ \binom{r}{0} t(t+1)\ldots(t+r) =$$

$$= \lambda^{r+1} + \sum_{l=0}^{r} \binom{r+1}{l} t(t+1)\ldots(t+r-l)\lambda^l = B_r(t),$$

что и доказывает лемму.

Теперь мы можем найти явное выражение для рациональных функций $\psi_r(t)$.

Теорема.

$$\psi_r(t) = \frac{B_{r-1}(t+1)}{B_r(t)} \quad (t = 0, 1, 2, \ldots). \tag{31.2}$$

Доказательство. Так как $B_{-1}(t) = 1$ и, как легко видеть, $B_0(t) = t + \lambda$, то при $r = 0$ доказываемое соотношение (31.2) имеет вид:

$$\psi_0(t) = \frac{1}{t+\lambda},$$

и было нами установлено уже в § 30. Допустим поэтому, что $r > 0$ и соотношение (31.2) уже установлено для $\psi_{r-1}(t)$. Убедимся, что оно остается справедливым и для $\psi_r(t)$; этим, очевидно, теорема будет доказана.

В силу принятого допущения мы имеем

$$\begin{aligned} 1 + t\psi_{r-1}(t) - (t+1)\psi_{r-1}(t+1) &= \\ = 1 + \frac{t B_{r-2}(t+1)}{B_{r-1}(t)} - \frac{(t+1) B_{r-2}(t+2)}{B_{r-1}(t+1)} &= \\ = \frac{K_r(t)}{B_{r-1}(t) B_{r-1}(t+1)}, \end{aligned}$$

где

$$\begin{aligned} K_r(t) = B_{r-1}(t+1) B_{r-1}(t) + t B_{r-2}(t+1) B_{r-1}(t+1) - \\ - (t+1) B_{r-2}(t+2) B_{r-1}(t). \end{aligned} \tag{31.3}$$

Поэтому в силу (30.2)

$$\begin{aligned} \psi_r(t) &= \frac{B_{r-2}(t+1)}{B_{r-1}(t)} \, \frac{B_{r-1}(t) B_{r-1}(t+1)}{K_r(t)} = \\ &= \frac{B_{r-2}(t+1) B_{r-1}(t+1)}{K_r(t)}. \end{aligned}$$

Подлежащее доказательству соотношение (31.2) поэтому равносильно соотношению

$$K_r(t) = B_r(t) B_{r-2}(t+1),$$

которое в свою очередь в силу (31.1) равносильно соотношению

$$K_r(t) = t B_{r-1}(t+1) B_{r-2}(t+1) + \lambda B_{r-1}(t) B_{r-2}(t+1);$$

подставляя вместо $K_r(t)$ его выражение (31.3), мы находим, что доказать надо следующее равенство:

$$\begin{aligned} B_{r-1}(t+1) B_{r-1}(t) - (t+1) B_{r-2}(t+2) B_{r-1}(t) = \\ = \lambda B_{r-1}(t) B_{r-2}(t+1), \end{aligned}$$

или

$$B_{r-1}(t+1) = (t+1)\,B_{r-2}(t+2) + \lambda\,B_{r-2}(t+1).$$

Но это соотношение мы непосредственно получаем, заменяя в рекуррентной формуле (31.1) r на $r-1$ и t на $t+1$. Таким образом, наша теорема доказана.

Мы видим, что каждая функция $\psi_r(t)$ представляет собой правильную рациональную дробь, числитель которой есть многочлен степени r, а знаменатель — степени $r+1$.

§ 32. Разложение функций $\psi_r(t)$ на простые дроби

Чтобы определить функции $\varphi_r(x)$, лапласовскими преобразованиями которых служат функции $\psi_r(t)$, мы должны теперь посмотреть, как рациональные функции $\psi_r(t)$ разлагаются на простые дроби, и с этой целью исследовать корни многочленов $B_r(t)$, служащих знаменателями этих рациональных функций.

Многочлен $B_r(t)$ степени $r+1$ имеет, как непосредственно видно из его определения, положительные коэффициенты, причем коэффициент при t^{r+1} равен 1. Отсюда уже следует, что все вещественные корни этого многочлена отрицательны и что он может быть представлен в виде:

$$B_r(t) = (t + a_{r0})(t + a_{r1}) \ldots (t + a_{rr}),$$

где числа a_{ri} $(0 \leqslant i \leqslant r)$ положительны или мнимы. Мы утверждаем, что *числа a_{ri} все положительны и что, если они расположены в порядке возрастания, то*

$$a_{ri} > a_{r,\,i-1} + 1 \quad (1 \leqslant i \leqslant r),$$

т. е. расстояние между двумя соседними корнями превосходит единицу.

Мы докажем это утверждение посредством индукции. Мы уже видели, что $B_0(t) = t + \lambda$, так что для $r = 0$ наше утверждение верно. Допустим, что оно верно для $B_r(t)$ $(r \geqslant 0)$, и покажем, что в таком случае оно справедливо и для $B_{r+1}(t)$. Для этого мы воспользуемся рекуррентной формулой (31.1), в силу которой

$$B_{r+1}(t) = t\,B_r(t+1) + \lambda\,B_r(t) =$$
$$= t\,(t + 1 + a_{r0}) \ldots (t + 1 + a_{rr}) + \lambda\,(t + a_{r0}) \ldots (t + a_{rr});$$

при этом в силу нашего предположения все $a_{rk} > 0$ и $a_{rk} > a_{r,k-1} + 1$ $(1 \leqslant k \leqslant r)$. Эта формула (как впрочем и непосредственное определение) показывает, что $B_{r+1}(0) > 0$. С другой стороны, она дает

$$B_{r+1}(-a_{r0}) = -\,a_{r0}\,(a_{r1} + 1 - a_{r0}) \ldots (a_{rr} + 1 - a_{r0}) < 0.$$

Это показывает, что $B_{r+1}(t)$ имеет корень между 0 и $-a_{r0}$. Пусть теперь k — одно из чисел ряда $0,1,\ldots,r-1$. Тогда

$$B_{r+1}(-a_{rk}-1)=\lambda(a_{r0}-a_{rk}-1)\ldots(a_{r,k-1}-a_{rk}-1)\times \\ \times(a_{rk}-a_{rk}-1)(a_{r,k+1}-a_{rk}-1)\ldots(a_{rr}-a_{rk}-1)$$

имеет, как легко подсчитать, $k+1$ отрицательных множителей, и следовательно, знак $(-1)^{k+1}$; напротив

$$B_{r+1}(-a_{r,k+1})=-a_{r,k+1}(-a_{r,k+1}+1+a_{r0})\ldots(-a_{r,k+1}+1+a_{rk})\times \\ \times(-a_{r,k+1}+1+a_{r,k+1})\ldots(-a_{r,k+1}+1+a_{rr})$$

имеет $k+2$ отрицательных множителей и, следовательно, знак $(-1)^k$. Таким образом, многочлен $B_{r+1}(t)$ для любого k $(0\leqslant k\leqslant r-1)$ имеет корень в промежутке между $-a_{rk}-1$ и $-a_{r,k+1}$. Наконец, так как старший член многочлена $B_{r+1}(t)$ есть t^{r+2}, то при отрицательных t, достаточно больших по абсолютной величине, $B_{r+1}(t)$ имеет знак $(-1)^{r+2}$; в то же время в выражении

$$B_{r+1}(-a_{rr}-1)=\lambda(-a_{rr}-1+a_{r0})\ldots(-a_{rr}-1+a_{rr})$$

все скобки отрицательны, так что оно имеет знак $(-1)^{r+1}$. Это показывает, что $B_{r+1}(t)$ имеет корень между $-a_{rr}-1$ и $-\infty$.

Сопоставляя все полученное, мы видим, что многочлен $B_{r+1}(t)$ имеет по меньшей мере по одному корню в каждом из $r+2$ промежутков:

$$(0,-a_{r0}),(-a_{rk}-1,\ -a_{r,k+1})\,(0\leqslant k\leqslant r-1),\ (-a_{rr}-1,\ -\infty),$$

которые попарно не имеют общих точек. Так как $B_{r+1}(t)$ есть многочлен степени $r+2$, то этим его корни исчерпаны; а так как рассмотренные нами $r+2$ промежутков таковы, что расстояние между двумя соседними из них равно единице, то расстояние между двумя соседними корнями многочлена $B_{r+1}(t)$ всегда превосходит единицу. Таким образом, наше утверждение о расположении корней многочленов $B_r(t)$ полностью доказано.

Из этого следует, что разложение функции $\psi_r(t)$ на простые дроби имеет вид:

$$\psi_r(t)=\sum_{k=0}^{r}\frac{C_{rk}}{t+a_{rk}},$$

где числители C_{rk} легко могут быть выражены через числа a_{rk} хорошо известными методами.

§ 33. Заключение

Так как функция $\frac{C_{rk}}{t+a_{rk}}$ служит, как легко непосредственно убедиться, преобразованием Лапласа функции $C_{rk}e^{-a_{rk}x}$, то функция $\psi(t)$, вид которой мы только что нашли, есть преобразование Лапласа функции

$$\sum_{k=0}^{r} C_{rk} e^{-a_{rk}x}. \tag{33.1}$$

Но мы определили $\psi_r(t)$ как преобразование Лапласа искомой функции $\varphi_r(x)$. Можем ли мы отсюда заключить, что $\varphi_r(x)$ совпадает с функцией (33.1)?

Теория преобразований Лапласа (в детали которой мы здесь не можем входить) показывает, что среди функций, обладающих данным преобразованием Лапласа, может быть только одна ограниченная и неотрицательная при $0 \leqslant x < +\infty$; а так как функция $\varphi_r(x)$, очевидно, обладает обоими этими свойствами, то для ее совпадения с функцией (33.1) достаточно убедиться в положительности этой последней; для этого же в свою очередь, очевидно, достаточно показать, что $C_{rk} > 0$ $(0 \leqslant k \leqslant r)$. В своем цитированном нами исследовании Пальм дает явное выражение чисел C_{rk} через числа a_{rk} и путем анализа этих выражений действительно доказывает положительность всех коэффициентов C_{rk}. Таким образом, мы имеем для всех $r \geqslant 0$

$$\varphi_r(x) = \sum_{k=0}^{r} C_{rk} e^{-a_{rk}x} \quad (0 \leqslant x < +\infty),$$

и задача Пальма может считаться полностью решенной. Мы видим, что для любого r функция Пальма $\varphi_r(x)$, однозначно определяющая собой поток вызовов, теряемых на L_r (т. е. поступающих на L_{r+1}), представляет собой линейную комбинацию $r+1$ показательных функций. В частности, Пальм приводит следующее явное выражение для $\varphi_1(x)$:

$$\varphi_1(x) = \left(\frac{1}{2} + \frac{1}{4\sqrt{\lambda + \frac{1}{4}}}\right) e^{-\left(\lambda + \frac{1}{2} - \sqrt{\lambda + \frac{1}{4}}\right)x} + \left(\frac{1}{2} - \frac{1}{4\sqrt{\lambda + \frac{1}{4}}}\right) e^{-\left(\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}}\right)x}.$$

ЧАСТЬ III

СИСТЕМЫ С ОЖИДАНИЕМ

Глава 9

Случай показательного распределения длин разговоров

§ 34. Вероятности состояний

Мы обращаемся теперь к изучению полнодоступного пучка с ожиданием. Поступившему вызову приходится ждать разговора тогда и только тогда, если он застает все n линий пучка занятыми. Поэтому здесь наша прежняя «вероятность потери» (т. е. вероятность застать все линии занятыми) может быть названа «вероятностью ожидания». Эта величина понятным образом играет известную роль в оценке качества работы пучка. Однако для систем с ожиданием эта роль сравнительно невелика, так как, если даже значительному большинству вызовов приходится ожидать, обслуживание должно быть признано вполне удовлетворительным во всех тех случаях, когда промежутки ожидания оказываются в своем большинстве очень малыми. Решающую роль здесь играет не частота ожиданий («потерь»), а природа времени ожидания γ как случайной величины; частота же ожиданий («потерь») дает нам только один штрих этой картины — вероятность неравенства $\gamma > 0$. Понятно поэтому, что конечной целью в исследовании систем с ожиданием всегда служит отыскание *закона распределения времени ожидания* γ.

Для всех задач, которые мы будем рассматривать, безразлично, является ли данный пучок линий упорядоченным или нет. Входящий поток вызовов мы будем всегда предполагать простейшим с параметром λ. Вызовы обслуживаются в порядке их поступления. Длины разговоров всегда будут мыслиться независимыми как друг от друга, так и от течения потока вызовов. Что касается закона распределения этих длин, то именно он составляет собой основной момент различия в задачах теории систем с ожиданием. Обычно бывает так, что при различных распределениях длин разговоров к исследованию времени ожидания приходится подходить различными методами. В настоящей главе мы рассмотрим самый простой случай, когда длины l разговоров подчиняются показательному закону

$$\mathsf{P}\{l > t\} = e^{-\beta t}, \quad t > 0, \quad \beta > 0 \text{ — постоянная.}$$

Для этого случая полное решение задачи было дано еще Эрлангом [7].

Условимся обозначать через $\mathsf{P}_k(t)$ вероятность того, что в момент t система находится в «состоянии k», т. е. имеется всего k «наличных» (говорящих или ожидающих) вызовов. При $k \leqslant n$ занято k линий, ожидающих нет; при $k > n$ все n линий заняты и имеется $k - n$ ожидающих.

Если $k < n$, то мы находимся в тех же условиях, что и в случае систем с потерями, так как при $k < n$ нет ни потерь, ни ожиданий. Все соображения, приведенные нами в § 20, поэтому остаются в силе и, как там, приводят нас к системе уравнений

$$\mathsf{P}_0'(t) = -\lambda \mathsf{P}_0(t) + \beta \mathsf{P}_1(t);$$
$$\mathsf{P}_k'(t) = \lambda \mathsf{P}_{k-1}(t) - (\lambda + k\beta)\mathsf{P}_k(t) + (k+1)\beta \mathsf{P}_{k+1}(t) \quad (0 < k < n)$$

[ср. $(\mathscr{E})$ § 20, где мы полагали $\beta = 1$]. Но последнее уравнение системы $(\mathscr{E})$ должно быть теперь заменено другим, так как теперь возможен переход в состояние n из состояния $n+1$, которое не имело смысла в случае системы с потерями.

Проведем общее рассмотрение для любого $k \geqslant n$. Будем, как прежде, обозначать через $\mathsf{P}_{rs}(\tau)$ вероятность перехода системы из состояния r в состояние s за время τ и понимать знак $\approx$ как равенство с точностью до бесконечно малых вида $o(\tau)$ при $\tau \to 0$. Тогда мы легко находим, аналогично § 20, при $k \geqslant n$ и $\tau \to 0$

$$\mathsf{P}_{k-1,\,k}(\tau) \approx \lambda\tau, \quad \mathsf{P}_{k+1,\,k}(\tau) \approx n\beta\tau;$$
$$\mathsf{P}_{kk}(\tau) \approx 1 - \lambda\tau - n\beta\tau, \quad \mathsf{P}_{ik}(\tau) \approx 0 \quad (|i-k| > 1)$$

[отличие от случая $k < n$ состоит в том, что теперь мы имеем $\mathsf{P}_{k+1,\,k}(\tau) \approx n\beta\tau$, а не $\approx (k+1)\beta\tau$, как прежде, так как при $k \geqslant n$ в состоянии $k+1$ мы имеем n, а не $k+1$ занятых линий]. Эти оценки с помощью рассуждений, в точности аналогичных проведенным нами в § 20, приводят к соотношению

$$\mathsf{P}_k(t+\tau) \approx \mathsf{P}_{k-1}(t)\lambda\tau + \mathsf{P}_k(t)(1 - \lambda\tau - n\beta\tau) + \mathsf{P}_{k+1}(t)\, n\beta\tau,$$

откуда мы, снова в тесной аналогии с рассуждениями § 20, получаем с помощью предельного перехода

$$\mathsf{P}_k'(t) = \lambda \mathsf{P}_{k-1}(t) - (\lambda + n\beta)\mathsf{P}_k(t) + n\beta \mathsf{P}_{k+1}(t) \quad (k \geqslant n).$$

Таким образом, в целом система ($\mathcal{E}$) § 20 теперь заменяется (бесконечной) системой уравнений

$$\left.\begin{aligned} &\mathsf{P}_0'(t) = -\lambda\mathsf{P}_0(t) + \beta\mathsf{P}_1(t); \\ &\mathsf{P}_k'(t) = \lambda\mathsf{P}_{k-1}(t) - (\lambda + k\beta)\mathsf{P}_k(t) + (k+1)\beta\mathsf{P}_{k+1}(t) \quad (0 < k < n); \\ &\mathsf{P}_k'(t) = \lambda\mathsf{P}_{k-1}(t) - (\lambda + n\beta)\mathsf{P}_k(t) + n\beta\mathsf{P}_{k+1}(t) \qquad (k \geqslant n). \end{aligned}\right\} \quad (34.1)$$

Как и в случае систем с потерями, мы принимаем в качестве вероятностей состояний пределы, к которым стремятся вероятности $\mathsf{P}_k(t)$ при $t \to \infty$. В случае систем с потерями мы доказали во всей полноте существование этих пределов [теорема Маркова, § 19]. В случае систем с ожиданием [т. е. в случае уравнений (34.1)] такое доказательство также может быть проведено; однако здесь оно несравненно более сложно и требует существенно новых идей, так как метод Маркова тесно связан с предположением конечного числа возможных состояний системы. Мы не можем поместить этого доказательства здесь и вынуждены ограничиться ссылкой на его возможность*. Необходимо еще отметить, что в нашем новом случае мы, кроме существования пределов величин $\mathsf{P}_k(t)$ при $t \to \infty$, должны еще доказать возможность предельного перехода во всей системе (34.1) — вопрос, который в § 20 у нас не возникал, так как там мы имели дело с конечной системой уравнений.

Итак, мы допускаем, что при $t \to \infty$ существуют пределы

$$\lim_{t \to \infty} \mathsf{P}_k(t) = p_k \quad (k = 0, 1, \ldots)$$

и что соответствующий предельный переход возможен одновременно во всех уравнениях системы (34.1). Так как при этом левые части всех этих уравнений в пределе обращаются в нуль (это доказывается в точности так же, как в § 20), то мы приходим к системе линейных уравнений

$$\left.\begin{aligned} -\lambda p_0 + \beta p_1 &= 0; \\ \lambda p_{k-1} - (\lambda + k\beta) p_k + (k+1)\beta p_{k+1} &= 0 \quad (0 < k < n); \\ \lambda p_{k-1} - (\lambda + n\beta) p_k + n\beta p_{k+1} &= 0 \quad (k \geqslant n), \end{aligned}\right\} \quad (34.2)$$

из которой, в соединении с нормирующим условием

$$\sum_{k=0}^{\infty} p_k = 1$$

мы и должны определить числа p_k.

Полагая

$$\lambda p_{k-1} - k\beta p_k = z_k \quad (1 \leqslant k \leqslant n),$$

* Мы не знаем в литературе ни одного изложения этого доказательства, хотя сама задача рассматривалась приводимым здесь методом много раз (Эрланг [7], Фрай [4], Колмогоров [2], Феллер [3]). Эрланг вводит возможность предельного перехода как особый постулат; все остальные упомянутые авторы ограничиваются кратким указанием на возможность доказательства.

мы легко находим из системы (34.2)

$$z_1 = 0, z_k - z_{k+1} = 0 \quad (1 \leqslant k \leqslant n),$$

откуда

$$z_k = 0 \quad (0 \leqslant k \leqslant n),$$

или

$$p_k = \frac{\lambda}{k\beta} p_{k-1} \quad (0 < k \leqslant n).$$

Это дает, если для краткости положить еще $\lambda / \beta = y$,

$$p_k = \frac{y^k}{k!} p_0 \quad (0 \leqslant k \leqslant n);$$

в частности, при $k = n$ мы находим

$$p_n = \frac{y^n}{n!} p_0. \tag{34.3}$$

Чтобы найти значения p_k при $k > n$, мы обращаемся к последней группе уравнений (34.2). Перепишем их в виде:

$$n\beta (p_{k+1} - p_k) = \lambda (p_k - p_{k-1}) \quad (k \geqslant n),$$

и просуммируем по k от n до $n + r$:

$$n\beta (p_{n+r+1} - p_n) = \lambda (p_{n+r} - p_{n-1}),$$

откуда

$$n\beta p_{n+r+1} + z_n = \lambda p_{n+r},$$

или, так как $z_n = 0$, $\lambda / \beta = y$,

$$p_{n+r+1} = \frac{y}{n} p_{n+r} \quad (r \geqslant 0),$$

и следовательно, в силу (34.3)

$$p_{n+r+1} = \left(\frac{y}{n}\right)^{r+1} \frac{y^n}{n!} p_0 \quad (r \geqslant 0).$$

Таким образом, для любого $k \geqslant n$ мы имеем

$$p_k = \left(\frac{y}{n}\right)^{k-n} p_n = \frac{y^k}{n!\, n^{k-n}} p_0. \tag{34.4}$$

Соединяя этот результат с полученным прежде для $k \leqslant n$, мы находим

$$\left.\begin{aligned} p_k &= \frac{y^k}{k!} p_0 \qquad (0 \leqslant k \leqslant n); \\ p_k &= \frac{y^k}{n!\, n^{k-n}} p_0 \qquad (k \geqslant n). \end{aligned}\right\} \tag{34.5}$$

Нам остается найти p_0. Нормирующее условие

$$\sum_{k=0}^{\infty} p_k = 1$$

дает

$$\frac{1}{p_0}=\sum_{k=0}^{n-1}\frac{y^k}{k!}+\frac{n^n}{n!}\sum_{k=n}^{\infty}\left(\frac{y}{n}\right)^k=s_{n-1}(y)+\frac{y^n}{(n-1)!\,(n-y)},$$

где положено $s_m(y)=\sum_{k=0}^{m}(y^k/k!)$. Вспомним для сравнения, что в случае системы с потерями мы имели (§ 20)

$$\frac{1}{p_0}=s_n(y)=s_{n-1}(y)+\frac{y^n}{n!}.$$

Отметим еще, что вероятность найти все линии занятыми («вероятность ожидания») равна в силу (34.4)

$$\Pi=\sum_{k=n}^{\infty}p_n=\frac{n^n p_0}{n!}\sum_{k=n}^{\infty}\left(\frac{y}{n}\right)^k=\frac{y^n}{n!}\,\frac{p_0}{1-\dfrac{y}{n}}. \tag{34.6}$$

§ 35. Закон распределения времени ожидания

Теперь мы уже легко можем найти вероятность $\mathsf{P}\{\gamma>t\}$ того, что для поступившего в произвольно выбранный момент вызова время ожидания будет больше чем t. Обозначим через $\mathsf{P}_k\{\gamma>t\}$ условную вероятность того же неравенства в предположении, что произведенный вызов застал систему в состоянии k. По формуле полной вероятности мы имеем

$$\mathsf{P}\{\gamma>t\}=\sum_{k=0}^{\infty}p_k\,\mathsf{P}_k\{\gamma>t\},$$

или, так как, очевидно, $\mathsf{P}_k\{\gamma>t\}=0$ при $k<n$ и $t\geqslant 0$,

$$\mathsf{P}\{\gamma>t\}=\sum_{k=n}^{\infty}p_k\,\mathsf{P}_k\{\gamma>t\}. \tag{35.1}$$

Величины p_k нам известны; остается определить величины $\mathsf{P}_k\{\gamma>t\}$ при всех $k\geqslant n$.

Положим $k-n=\nu$ ($\nu=0, 1, 2, \ldots$). Наша задача — найти вероятность неравенства $\gamma>t$ при условии, что в момент вызова все линии были заняты и сверх того имелось ν ожидающих. Очевидно, что при этом наш вызов получает разговор после $(\nu+1)$-го освобождения линии. Искомая вероятность есть поэтому вероятность того, что за время t после появления нашего вызова произойдет не более чем ν освобождений линии. Пусть $q_r(t)$ $(0\leqslant r\leqslant\nu)$ есть вероятность того, что за это время произойдет ровно r освобождений; тогда в силу $k-n=\nu$

$$\mathsf{P}_k\{\gamma>t\}=\sum_{r=0}^{k-n}q_r(t)\qquad(k\geqslant n).$$

Но поток освобождений за время ожидания нашего вызова представляет собой в силу показательного закона распределения длин разговоров простейший поток с параметром $n\beta$, так как вероятность того, что не произойдет ни одного освобождения за время t после такого момента, когда все линии заняты, равна $(e^{-\beta t})^n = e^{-\beta n t}$. Величина $q_r(t)$ есть вероятность того, что за время t наступит r событий этого потока; в силу формул главы 1 поэтому

$$q_r(t) = e^{-n\beta t} \frac{(n\beta t)^r}{r!} \quad (0 \leqslant r \leqslant \nu),$$

и мы находим

$$\mathsf{P}_k\{\gamma > t\} = \sum_{r=0}^{k-n} e^{-n\beta t} \frac{(n\beta t)^r}{r!} \quad (k \geqslant n).$$

Возвращаясь к формуле (35.1) и используя соотношение (34.4), мы находим

$$\mathsf{P}\{\gamma > t\} = \sum_{k=n}^{\infty} p_k \sum_{r=0}^{k-n} e^{-n\beta t} \frac{(n\beta t)^r}{r!} = e^{-n\beta t} \sum_{k=n}^{\infty} \left(\frac{y}{n}\right)^{k-n} p_n \sum_{r=0}^{k-n} \frac{(n\beta t)^r}{r!} =$$

$$= p_n e^{-n\beta t} \sum_{r=0}^{\infty} \frac{(n\beta t)^r}{r!} \sum_{k=n+r}^{\infty} \left(\frac{y}{n}\right)^{k-n} =$$

$$= p_n e^{-n\beta t} \sum_{r=0}^{\infty} \frac{(n\beta t y)^r}{n^r\, r!} \sum_{k=n+r}^{\infty} \left(\frac{y}{n}\right)^{k-n-r} =$$

$$= \frac{p_n e^{-n\beta t}}{1 - \frac{y}{n}} \sum_{r=0}^{\infty} \frac{(\lambda t)^r}{r!} = \frac{p_n}{1 - \frac{y}{n}} e^{-(n\beta - \lambda) t},$$

или, так как в силу (34.3) и (34.6) $p_n = \frac{y^n}{n!} p_0 = \Pi\left(1 - \frac{y}{n}\right)$

$$\mathsf{P}\{\gamma > t\} = \Pi e^{-(n\beta - \lambda) t} \quad (t \geqslant 0).$$

Этим наша задача решена. Мы видим, что в принятых нами условиях время ожидания подчиняется показательному закону распределения с параметром $n\beta - \lambda$. Вместе с тем мы получаем

$$\mathsf{P}\{\gamma > 0\} = \Pi,$$

как оно и должно быть: через Π мы как раз обозначили в конце § 34 вероятность застать все линии занятыми («вероятность ожидания»).

Глава 10

Однолинейные системы в случае стандартной длины разговора

§ 36. Разностно-дифференциальное уравнение задачи

За пределами показательного распределения длин разговоров исследование систем с ожиданием сопряжено с большими трудностями. Простые и законченные результаты здесь удается получить лишь в некоторых частных предположениях. Особенно важным в практическом отношении здесь является случай систем с одной линией (короче: однолинейных систем), для которых задача исследования времени ожидания может быть продвинута весьма далеко при статистических предпосылках широкой общности. Этим случаем мы и будем теперь заниматься до конца книги.

В настоящей главе мы изложим созданную Эрлангом интересную как оригинальностью метода, так и законченностью результатов теорию однолинейных систем в предположении, что все разговоры имеют в точности одну и ту же длину τ. Во всех других отношениях мы сохраняем статистические предпосылки предшествующей главы.

Рассмотрим какие-либо два последовательных вызова. Пусть γ_0 — время ожидания первого, а γ — второго вызова. Обозначим через z расстояние между этими двумя вызовами. Очевидно, что если $z \geqslant \gamma_0 + \tau$, то при появлении второго вызова (единственная) линия свободна и $\gamma = 0$; если же $z < \gamma_0 + \tau$, то при появлении второго вызова линия занята и ему приходится ждать в течение промежутка времени $\gamma = \gamma_0 + \tau - z$. Таким образом, мы находим, что при данном $z = u$

$$\gamma = \begin{cases} 0 & (\gamma_0 + \tau - u \leqslant 0); \\ \gamma_0 + \tau - u & (\gamma_0 + \tau - u \geqslant 0), \end{cases}$$

или, что то же,

$$\gamma = \max(0, \quad \gamma_0 + \tau - u).$$

Пусть теперь t — любое положительное число. Убедимся, что при данном $z = u$ неравенства $\gamma < t$ и $\gamma_0 < t + u - \tau$ равносильны между собой. В самом деле, если $\gamma < t$, то либо $\gamma = 0$, и тогда $\gamma_0 \leqslant u - \tau < u + t - \tau$, либо $\gamma = \gamma_0 + \tau - u$, и тогда $\gamma_0 + \tau - u < t$, и значит, $\gamma_0 < t + u - \tau$; обратно, если $\gamma_0 < t + u - \tau$, то либо $\gamma_0 < u - \tau$, $\gamma_0 + \tau - u < 0$, $\gamma = 0 < t$, либо $\gamma_0 > u - \tau$, $\gamma = \gamma_0 + \tau - u < t$. Таким образом, если P_u означает условную вероятность, вычисленную в предположении, что $z = u$, то мы имеем при любом $u > 0$

$$\mathsf{P}_u\{\gamma < t\} = \mathsf{P}_u\{\gamma_0 < t + u - \tau\} \quad (t > 0);$$

но γ_0 (время ожидания первого вызова), очевидно, не зависит (как случайная величина) от того, когда последует второй вызов, т. е. какое значение получит величина z; поэтому условная вероятность $\mathsf{P}_u\{\gamma_0<t+u-\tau\}$ неравенства $\gamma_0<\tau+u-\tau$ равна безусловной вероятности $\mathsf{P}\{\gamma_0<t+u-\tau\}$ того же неравенства, и мы получаем

$$\mathsf{P}_u\{\gamma<t\}=\mathsf{P}\{\gamma_0<t+u-\tau\}\quad(t>0). \tag{36.1}$$

Обозначим через $f(t)$ закон распределения величины γ, полагая

$$f(t)=\mathsf{P}\{\gamma<t\}.$$

Так как поток вызовов мы предполагаем простейшим с параметром λ, то вероятность неравенств $u<z<u+du$ (с точностью до бесконечно малых высших порядков) равна $\lambda e^{-\lambda u}\,du$, и мы по формуле полной вероятности находим

$$f(t)=\int_0^\infty \lambda e^{-\lambda u}\,\mathsf{P}_u\{\gamma<t\}\,du,$$

или в силу (36.1) при $t\geqslant 0$

$$f(t)=\int_0^\infty \lambda e^{-\lambda u}\,\mathsf{P}\{\gamma_0<t+u-\tau\}\,du\qquad(t\geqslant 0).$$

Но закон распределения величины γ_0 дается той же функцией $f(t)$, что и для величины γ; поэтому мы находим

$$f(t)=\int_0^\infty \lambda e^{-\lambda u}\,f(t+u-\tau)\,du\qquad(t\geqslant 0).$$

Это уравнение и будет служить нам основой для определения искомой функции $f(t)$. Прежде всего, предполагая эту функцию дифференцируемой, мы находим

$$f'(t)=\int_0^\infty \lambda e^{-\lambda u}\,f'(t+u-\tau)\,du\qquad(t\geqslant 0),$$

и интеграция по частям дает

$$f'(t)=\lambda e^{-\lambda u}\,f(t+u-\tau)\Big|_0^\infty+\lambda^2\int_0^\infty e^{-\lambda u}\,f(t+u-\tau)\,du=-\lambda f(t-\tau)+\lambda f(t),$$

или

$$f'(t)=\lambda\,[f(t)-f(t-\tau)]\quad(t\geqslant 0).$$

Мы получаем, таким образом, для определения функции $f(t)$ разностно-дифференциальное уравнение простого вида. Это уравнение может быть еще упрощено преобразованием искомой функции

$$f(t) = e^{\lambda t} g(t),$$

что, как легко видеть, дает для новой неизвестной функции уравнение

$$g'(t) = -\lambda e^{-\lambda\tau} g(t-\tau) \quad (t > 0). \tag{36.2}$$

§ 37. Закон распределения времени ожидания

Рассмотрим сначала отрезок времени $0 < t \leqslant \tau$; в силу $t > 0$ при этом имеют место все соотношения, выведенные в § 36. Но при $t \leqslant \tau$ мы имеем $f(t-\tau) = \mathsf{P}\{\gamma < t - \tau\} = 0$, и следовательно, соотношение (36.2) дает

$$g'(t) = 0, \quad g(t) = c = \text{const} \quad (0 < t \leqslant \tau),$$

откуда

$$f(t) = ce^{\lambda t} \quad (0 < t \leqslant \tau).$$

Чтобы определить постоянную c, заставим в последнем равенстве t стремиться к нулю; в пределе мы получим $c = f(+0)$, но $f(-0) = 0$, а потому $c = f(+0) - f(-0) = \mathsf{P}\{\gamma = 0\}$; это есть вероятность того, что поступившему в произвольно выбранный момент вызову не придется ожидать (линия окажется свободной); поэтому $c = 1 - \alpha$, где α означает вероятность в произвольно выбранный момент застать линию занятой. Иначе говоря, α есть математическое ожидание суммарной длительности всех разговоров, ведущихся в течение единицы времени. Но математическое ожидание числа разговоров в единицу времени равно λ, а длительность каждого разговора равна τ; следовательно,

$$\alpha = \lambda\tau, \quad c = 1 - \alpha = 1 - \lambda\tau.$$

Таким образом, окончательно

$$f(t) = (1-\alpha) e^{\lambda t} \quad (\alpha = \lambda\tau, \ 0 < t \leqslant \tau),$$

и закон распределения времени ожидания найден нами для отрезка $0 < t \leqslant \tau$.

Мы теперь докажем, *что для любого неотрицательного целого числа n в отрезке $n\tau < t \leqslant (n+1)\tau$ функция $g(t)$ определяется формулой*

$$g(t) = (1-\alpha) \sum_{k=0}^{n} e^{-\lambda k\tau} \frac{(k\tau - t)^k}{k!} \lambda^k. \tag{37.1}$$

Так как функция $f(t)$ просто выражается через функцию $g(t)$, то этим мы получим явное выражение для закона распределения времени ожидания.

При $n = 0$ формула (37.1) дает

$$g(t) = 1 - \alpha \quad (0 < t \leqslant \tau),$$

что нами доказано выше. Поэтому мы можем доказать формулу (37.1) с помощью индукции. Мы допустим, что она верна для какого-либо числа $n \geqslant 0$, и покажем, что в таком случае она останется верной и для числа $n+1$.

Итак, пусть соотношение (37.1) верно при $n\tau < t \leqslant (n+1)\tau$; если $(n+1)\tau < t \leqslant (n+2)\tau$, то в силу (36.2) мы будем иметь

$$g'(t) = -\lambda e^{-\lambda\tau} g(t-\tau) = -\lambda e^{-\lambda\tau}(1-\alpha)\sum_{k=0}^{n} e^{-\lambda k\tau}\frac{(k\tau - t + \tau)^k}{k!}\lambda^k;$$

интегрируя это равенство по t от $(n+1)\tau$ до числа, которое мы снова обозначим через t, мы найдем

$$g(t) - g[(n+1)\tau] = -\lambda e^{-\lambda\tau}(1-\alpha)\sum_{k=0}^{n}\frac{e^{-\lambda k\tau}\lambda^k}{k!}\int_{(n+1)\tau}^{t}[(k+1)\tau - u]^k du.$$

Здесь

$$\int_{(n+1)\tau}^{t}[(k+1)\tau - u]^k du = -\frac{1}{k+1}\{[(k+1)\tau - t]^{k+1} - [(k-n)\tau]^{k+1}\},$$

а потому

$$g(t) - g[(n+1)\tau] =$$

$$= \lambda e^{-\lambda\tau}(1-\alpha)\sum_{k=0}^{n}\frac{e^{-\lambda k\tau}\lambda^k}{(k+1)!}\{[(k+1)\tau - t]^{k+1} - [(k+1)\tau - (n+1)\tau]^{k+1}\} =$$

$$= (1-\alpha)\sum_{k=0}^{n}\frac{e^{-\lambda(k+1)\tau}\lambda^{k+1}}{(k+1)!}\{[(k+1)\tau - t]^{k+1} - [(k+1)\tau - (n+1)\tau]^{k+1}\} =$$

$$= (1-\alpha)\sum_{r=1}^{n+1}\frac{e^{-\lambda r\tau}\lambda^r}{r!}\{[r\tau - t]^r - [r\tau - (n+1)\tau]^r\} =$$

$$= (1-\alpha)\sum_{r=0}^{n+1}\frac{\lambda^r e^{-\lambda r\tau}}{r!}(r\tau - t)^r - (1-\alpha)\sum_{r=0}^{n+1}\frac{\lambda^r e^{-\lambda r\tau}}{r!}[r\tau - (n+1)\tau]^r, \quad (37.2)$$

где суммирование по r можно вести от нуля, так как члены с $r=0$ в обоих суммах, очевидно, взаимно уничтожаются.

Формула (37.1) по нашему допущению верна при $n\tau < t \leqslant (n+1)\tau$. Полагая в ней $t = (n+1)\tau$, мы находим

$$g[(n+1)\tau] = (1-\alpha)\sum_{r=0}^{n+1}\frac{\lambda^r e^{-\lambda r\tau}}{r!}[r\tau - (n+1)\tau]^r, \quad (37.3)$$

где суммирование можно вести до $n+1$ потому, что член суммы с $r = n+1$, очевидно, равен нулю. Наконец, складывая между собой равенства (37.2) и (37.3), мы находим

$$g(t) = (1-\alpha)\sum_{r=0}^{n+1} \frac{\lambda^r e^{-\lambda r\tau}}{r!}(r\tau - t)^r$$

при любом t в промежутке $(n+1)\tau < t \leqslant (n+2)\tau$, что и требовалось доказать.

Таким образом, мы получаем для любого $t > 0$

$$\mathsf{P}\{\gamma < t\} = f(t) = e^{\lambda t}(1-\lambda\tau)\sum_{k=0}^{n} e^{-\lambda k\tau}\frac{(k\tau - t)^k}{k!}\lambda^k,$$

где целое неотрицательное число n определяется неравенствами

$$n\tau < t \leqslant (n+1)\tau.$$

Глава 11

Общая теория однолинейных систем

§ 38. Постановка задачи и обозначения

В двух предшествующих главах мы, следуя классическим работам Эрланга, нашли закон распределения времени ожидания в двух наиболее простых предположениях относительно закона распределения длин разговоров: в случае показательного распределения (см. гл. 9) и (для однолинейных систем) в случае стандартной длительности разговора (см. гл. 10). Однако можно считать, что на практике мы встречаемся с тем или другим из этих двух простейших распределений лишь в очень редких случаях. В большинстве же приложений мы можем в лучшем случае рассчитывать лишь на некоторое приближение реального распределения длин разговоров к той или другой из наших двух предпосылок; и даже для такого расчета в очень многих случаях мы не имеем в сущности никаких оснований. Поэтому представляется весьма желательным построить метод, позволяющий определить закон распределения времени ожидания (или хотя бы важнейшие его статистические характеристики) при возможно широких предпосылках относительно распределения длин разговоров.

В своей общей постановке эта задача приводит к расчетам, трудно обозримым по своей сложности. Поэтому во всем дальнейшем мы сосредоточим наше внимание на практически весьма важном случае *системы с одной линией*. Зато в отношении закона распределения длин разговоров мы ограничимся естественным требованием существования конечного математического ожидания, оставляя этот закон во всем остальном совершенно произвольным. Мы увидим, что при этих предпосылках задача отыскания закона распределения времени ожидания принципиально решается до конца сравнительно несложными приемами.

В течение всей настоящей главы мы будем иметь дело с пучком из одной линии, на которую поступает простейший поток вызовов с параметром λ. Вызовы, заставшие линию занятой, ожидают ее освобождения и занимают ее в порядке их поступления. Длины разговоров не зависят ни друг от друга, ни от числа ожидающих. Вероятность того, что длина произвольно выбранного разговора окажется больше чем t, мы будем обозначать через $F(t)$, так что средняя длительность разговора будет

$$-\int_0^\infty t\,dF(t) = \int_0^\infty F(t)\,dt = s.$$

Кроме того, мы вводим следующие обозначения, которыми будем пользоваться на протяжении всей главы:

α — вероятность того, что в произвольно выбранный момент времени линия окажется занятой; иначе говоря, математическое ожидание суммарного времени занятости линии за 1 час (часом мы будем условно называть принятую единицу времени);

π_k $(k=0, 1, 2, \ldots)$ — вероятность того, что в начале произвольно выбранного разговора мы будем иметь k ожидающих;

$v_k(t) = e^{-\lambda t}(\lambda t)^k/k!$ — вероятность того, что в течение промежутка времени длины t на линию поступит k вызовов;

γ (случайная величина) — время ожидания для вызова, поступившего в произвольно выбранный момент времени.

Дальнейшие обозначения будут объяснены по мере их введения.

§ 39. Вспомогательные предложения

Лемма 1. *Пусть $0 < a < b$. Вероятность застать линию в произвольно выбранный момент времени занятой разговором, длина которого заключена между a и b, равна*

$$-\lambda\int_a^b u\,dF(u).$$

Доказательство. Обозначим через $\mathsf{P}_T(a, b)$ вероятность застать линию занятой разговором длины, заключенной между a и b*, в момент времени, произвольно выбранный в промежутке $(0, T)$. Лемма 1 утверждает тогда, что

$$\lim_{T\to\infty} \mathsf{P}_T(a, b) = -\lambda\int_a^b u\,dF(u).$$

* В дальнейшем мы такой разговор для краткости будем называть «разговором длины (a, b)».

Обозначим далее через $L_T^*(a, b)$ суммарную длительность всех разговоров и частей разговоров длины (a, b), ведущихся в промежутке $(0, T)$. $L_T^*(a, b)$ есть случайная величина; если она принимает какое-либо значение l, то соответствующая условная вероятность застать разговор длины (a, b), очевидно, равна l/T. Поэтому в силу формулы полной вероятности

$$\mathsf{P}_T(a, b) = \sum_l \mathsf{P}\{L_T^*(a, b) = l\} \frac{l}{T} = \frac{1}{T} \mathsf{M} L_T^*(a, b),$$

где M — символ математического ожидания.

Для доказательства леммы 1 достаточно поэтому установить, что при $T \to \infty$

$$\lim \frac{1}{T} \mathsf{M} L_T^*(a, b) = -\lambda \int_a^b u\, dF(u). \tag{39.1}$$

Пусть $L_T(a, b)$ означает суммарную длительность разговоров длины (a, b), *начинающихся* в отрезке $(0, T)$. Так как, очевидно,

$$|L_T(a, b) - L_T^*(a, b)| < b,$$

а следовательно, и

$$|\mathsf{M} L_T(a, b) - \mathsf{M} L_T^*(a, b)| < b,$$

то (39.1) равносильно соотношению

$$\frac{1}{T} \mathsf{M} L_T(a, b) \to -\lambda \int_a^b u\, dF(u) \quad (T \to \infty);$$

а так как $\mathsf{M} L_T(a, b)$, очевидно, пропорционально T, то последнее соотношение означает просто, что при любом $T > 0$

$$\frac{1}{T} \mathsf{M} L_T(a, b) = -\lambda \int_a^b u\, dF(u).$$

В частности, это соотношение будет установлено (и, значит, лемма 1 доказана), если мы убедимся, что

$$\lim_{T \to \infty} \frac{1}{T} \mathsf{M} L_T(a, b) = -\lambda \int_a^b u\, dF(u). \tag{39.2}$$

Пусть u — любое положительное число и $T < u$. В промежутке $(0, T)$ может тогда начаться не более одного разговора длины $(u, u + du)$, так что случайная величина $L_T(u, u + du)$ может, кроме значения 0, принять только значение вида $u + \theta\, du$, где $0 < \theta < 1$; такое значение она примет, очевидно, если (1°) в отрезке $(0, T)$ начнется по меньшей мере один разговор и (2°) последний из начавшихся в $(0, T)$ разговоров будет длины $(u, u + du)$. Вероятность (2°) есть $F(u) - F(u + du)$, а вероятность (1°) при $T \to 0$ имеет вид $\lambda T + o(T)$, где

λ — среднее число начал разговоров в единицу времени, совпадающее, очевидно, с параметром входящего потока вызовов. Таким образом, случайная величина $L_T(u, u+du)$ в рассматриваемом нами случае либо равна нулю, либо принимает значение вида $u+\theta\,du$, причем последнее — с вероятностью $[\lambda T + o(T)]\ [F(u) - F(u+du)]$. Таким образом,

$$\mathsf{M}L_T(u, u+du) = (u+\theta\,du)\,[\lambda T + o(T)]\,[F(u) - F(u+du)],$$

и следовательно, суммируя по u от a до b,

$$\mathsf{M}L_T(a, b) = [-\lambda T + o(T)]\int\limits_a^b u\,dF(u).$$

Отсюда следует (39.2), а значит, и лемма 1.

Лемма 1, таким образом, доказана. При $a=0$, $b=+\infty$ рассматриваемая в ней вероятность, очевидно, есть вероятность α того, что в произвольно выбранный момент времени линия окажется занятой. Мы приходим таким образом к важной формуле

$$\alpha = -\lambda\int\limits_0^\infty t\,dF(t) = \lambda s$$

(которую впрочем можно было бы усмотреть и непосредственным рассуждением).

Условимся называть ***разговором типа*** k ($k = 0, 1, 2, \ldots$) разговор, в начале которого имеется k ожидающих, так что введенную нами в § 38 величину π_k можно определить как долю разговоров типа k среди всех ведущихся разговоров (вероятность того, что наудачу выбранный разговор окажется разговором типа k).

Лемма 2. *Если в некоторый момент линия оказалась занятой, то условная вероятность того, что она занята разговором типа k, равна π_k.*

Доказательство. По определению величины π_k среди λ разговоров, ведущихся в среднем в единицу времени, мы будем иметь в среднем $\lambda\pi_k$ разговоров типа k. Так как длительность разговора, как случайная величина, не зависит от числа ожидающих в его начале (т. е. от его типа), то средняя длительность разговора типа k равна s, а следовательно, суммарная длительность разговоров типа k, ведущихся в единицу времени, в среднем равна $\lambda\pi_k s = \alpha\pi_k$; но условная вероятность, о которой идет речь в лемме 2, есть отношение этой средней суммарной длительности ко времени занятости линии, т. е. к α; таким образом, эта условная вероятность равна π_k, и лемма 2 доказана.

Выберем теперь произвольный разговор длины $t > 0$ и найдем вероятность $u_k(z)$ $(0 < z \leqslant t)$ того, что через промежуток времени z после начала этого разговора число ожидающих будет равно k

($k=0,\ 1,\ 2,\ldots$). Так как длина и тип разговора взаимно независимы, то вероятность того, что в начале выбранного разговора будет r ожидающих, равна π_r; если же это случится, то для того, чтобы по истечении времени z число ожидающих достигло k, необходимо и достаточно, чтобы было $r\leqslant k$ и чтобы за этот промежуток времени z поступило $k-r$ новых вызовов, вероятность чего есть

$$v_{k-r}(z)=e^{-\lambda z}\frac{(\lambda z)^{k-r}}{(k-r)!}.$$

Мы приходим таким образом к следующему предложению.

Лемма 3. *Вероятность $u_k(z)$ того, что через промежуток времени $z\leqslant t$ после начала произвольно выбранного разговора длины t мы будем иметь k ожидающих, равна*

$$u_k(z)=\sum_{r=0}^{k}\pi_r v_{k-r}(z)\quad (k=0,\ 1,\ 2,\ldots).$$

Условимся теперь обозначать через τ промежуток времени между произвольно выбранным моментом и окончанием того разговора, который мы в этот момент застали (остаточная длительность случайно застигнутого разговора). Имеет место

Лемма 4. *Пусть $k\geqslant 0$, $x>0$ и $dx>0$ мало. Вероятность застать в произвольно выбранный момент k ожидающих и разговор с $x<\tau<x+dx$ равна*

$$-\lambda\,dx\int_{x}^{\infty}u_k(t-x)\,dF(t)+o(dx).$$

Доказательство. Трудность обоснования леммы 4 заключается в том, что закон распределения случайной величины τ зависит от того, сколько ожидающих мы застаем в выбранный нами момент времени. Если мы обозначим через

a — событие, состоящее в том, что в выбранный момент имеется k ожидающих;

b — событие, состоящее в том, что в выбранный момент ведется разговор с $x<\tau<x+dx$,

то эти события взаимно зависимы; наша задача — определить вероятность $\mathsf{P}(ab)$ их совместного наступления. С этой целью рассмотрим сначала условную вероятность $\mathsf{P}_t(ab)$ той же пары событий при условии, что застигнутый разговор имеет длину t. Мы можем записать в легко понятных обозначениях

$$\mathsf{P}_t(ab)=\mathsf{P}_t(b)\,\mathsf{P}_{tb}(a). \tag{39.3}$$

Так как выбранный нами момент, если он застал разговор длины t, с одинаковой вероятностью мог попасть в любой момент этого разговора, то

$$\mathsf{P}_t(b)=\begin{cases} dx/t & (t>x); \\ 0 & (t\leqslant x). \end{cases} \qquad (39.4)$$

С другой стороны событие b равносильно событию

$$t-x-dx<t-\tau<t-x,$$

где $t-\tau$ есть возраст застигнутого разговора. Таким образом, $\mathsf{P}_{tb}(a)$ есть условная вероятность застать в произвольно выбранный момент k ожидающих, если известно, что в этот момент мы застали разговор длины t и возраста, заключенного между $t-x-dx$ и $t-x$. Согласно определению функции $u_k(x)$ (см. лемму 3) эта вероятность при $dx\to 0$ асимптотически равна $u_k(t-x)$*.

Таким образом, в силу (39.3) и (39.4) мы имеем

$$\mathsf{P}_t(ab)=\begin{cases} u_k(t-x)\,dx/t+o(dx) & (t>x); \\ 0 & (t\leqslant x). \end{cases}$$

Такова условная вероятность совместного наступления событий a и b *при условии, что застигнут разговор длины* t. Нам остается теперь освободиться от этого условия. Согласно лемме 1 закон распределения длины застигнутого разговора есть

$$G(t)=-\lambda\int_t^\infty u\,dF(u),$$

и мы имеем по формуле полной вероятности

$$\mathsf{P}(a,\,b)=-\int_0^\infty \mathsf{P}_t(a,\,b)\,dG(t)=-\int_x^\infty u_k(t-x)\frac{dx}{t}\,dG(t)+o(dx)=$$

$$=-\lambda\int_x^\infty u_k(t-x)\frac{dx}{t}\,t\,dF(t)=-\lambda\,dx\int_x^\infty u_k(t-x)\,dF(t)+o(dx),$$

что и требовалось доказать.

Лемма 5.

$$-\int_0^\infty u_{k+1}(x)\,dF(x)=\pi_k \quad (k>0);$$

$$-\int_0^\infty u_1(x)\,dF(x)=\pi_0\left\{1+\int_0^\infty e^{-\lambda x}\,dF(x)\right\}.$$

Доказательство. По смыслу функции $u_{k+1}(x)$ [см. формулировку леммы 3] интеграл

$$-\int_0^\infty u_{k+1}(x)\,dF(x)$$

---

* Непрерывность функции $u_k(z)$ вытекает из леммы 3.

означает вероятность того, что в конце произвольно выбранного разговора мы будем иметь $k+1$ ожидающих. В случае $k>0$ это равносильно тому, что следующий разговор начнется при k ожидающих, вероятность чего равна π_k. Этим доказано первое из двух утверждаемых равенств.

В случае $k=0$ вероятность того, что выбранный разговор окончится при одном ожидающем, будет *меньше* чем π_0; дело в том, что разговор без ожидающих может начаться не только при окончании предшествующего разговора с одним ожидающим, но и другим способом — появлением вызова в момент, когда линия свободна. Поэтому

$$-\int_0^\infty u_1(x)\,dF(x)=\pi_0-\rho,$$

где $\rho>0$. Для определения ρ мы сложим между собой все установленные нами соотношения, замечая при этом, что

$$\sum_{k=0}^\infty \pi_k=1 \quad \text{и} \quad \sum_{k=0}^\infty u_k(x)=1$$

при любом x; поэтому мы получаем

$$-\int_0^\infty [1-u_0(x)]\,dF(x)=1-\rho;$$

а так как в силу леммы 3 $u_0(x)=\pi_0 v_0(x)=\pi_0 e^{-\lambda x}$, то отсюда

$$\rho=-\pi_0\int_0^\infty e^{-\lambda x}\,dF(x), \tag{39.5}$$

и лемма 5 доказана полностью.

Лемма 6. *При* $|a|\leqslant 1$, $z>0$

$$\sum_{k=0}^\infty a^k u_k(z)=e^{\lambda z(a-1)}\sum_{k=0}^\infty \pi_k a^k.$$

Доказательство. В силу леммы 3

$$\sum_{k=0}^\infty a^k u_k(z)=\sum_{k=0}^\infty a^k\sum_{r=0}^k \pi_r e^{-\lambda z}\frac{(\lambda z)^{k-r}}{(k-r)!}=e^{-\lambda z}\sum_{r=0}^\infty \pi_r\sum_{k=r}^\infty a^k\frac{(\lambda z)^{k-r}}{(k-r)!}=$$
$$=e^{-\lambda z}\sum_{r=0}^\infty \pi_r\, a^r\sum_{k=r}^\infty\frac{(a\lambda z)^{k-r}}{(k-r)!}=e^{\lambda z(a-1)}\sum_{k=0}^\infty \pi_k a^k,$$

что и требовалось доказать.

Обозначим через $\psi(\xi)$ характеристическую функцию длительности разговора, т. е. положим

$$\psi(\xi)=-\int_0^\infty e^{i\xi t}\,dF(t);$$

положим, далее, для любого вещественного ξ

$$\chi(\xi)=\sum_{k=0}^{\infty}\pi_k\,[\psi(\xi)]^k;$$

наконец, пусть

$$\eta=\eta(\xi)=\frac{\lambda}{i}\,[\psi(\xi)-1].$$

Лемма 7.

$$\chi(\xi)=(1-\alpha)\,\frac{1-\psi(\xi)}{\psi(\eta)-\psi(\xi)}\,.$$

Доказательство. В силу леммы 5 мы имеем [определяя ρ формулой (39.5)]

$$\chi(\xi)=\sum_{k=0}^{\infty}[\psi(\xi)]^k\pi_k=-\sum_{k=0}^{\infty}[\psi(\xi)]^k\int_0^{\infty}u_{k+1}(x)\,dF(x)+\rho=$$

$$=\rho-\frac{1}{\psi(\xi)}\int_0^{\infty}\left\{\sum_{k=1}^{\infty}[\psi(\xi)]^k u_k(x)\right\}dF(x),$$

откуда в силу леммы 6

$$\chi(\xi)=\rho-\frac{1}{\psi(\xi)}\int_0^{\infty}\left\{e^{\lambda x[\psi(\xi)-1]}\sum_{k=0}^{\infty}\pi_k\,[\psi(\xi)]^k-u_0(x)\right\}dF(x)=$$

$$=\rho-\frac{1}{\psi(\xi)}\int_0^{\infty}\left\{e^{\lambda x[\psi(\xi)-1]}\chi(\xi)-\pi_0e^{-\lambda x}\right\}dF(x)=$$

$$=\rho-\frac{\chi(\xi)}{\psi(\xi)}\int_0^{\infty}e^{\lambda x[\psi(\xi)-1]}dF(x)+\frac{1}{\psi(\xi)}\int_0^{\infty}\pi_0e^{-\lambda x}dF(x)=$$

$$=\rho-\frac{\chi(\xi)}{\psi(\xi)}\int_0^{\infty}e^{i\eta x}dF(x)-\frac{\rho}{\psi(\xi)}=$$

$$=\rho\,\frac{\psi(\xi)-1}{\psi(\xi)}+\frac{\xi(\xi)\,\psi(\eta)}{\psi(\xi)}\,;$$

$$\chi(\xi)\,\frac{\psi(\xi)-\psi(\eta)}{\psi(\xi)}=\rho\,\frac{\psi(\xi)-1}{\psi(\xi)}\,;$$

$$\chi(\xi)=\rho\,\frac{1-\psi(\xi)}{\psi(\eta)-\psi(\xi)}\,.$$

Для доказательства леммы 7 остается показать, что $\rho=1-\alpha$. Но ρ не зависит от ξ, а так как $\chi(0)=1$, то из последней формулы

$$\rho=\lim_{\xi\to 0}\frac{\psi(\eta)-\psi(\xi)}{1-\psi(\xi)}\,,$$

или, по правилу Лопиталя,

$$\rho = \lim_{\xi \to 0} \frac{\psi'(\eta)\dfrac{d\eta}{d\xi} - \psi'(\xi)}{-\psi'(\xi)} = 1 - \lim_{\xi \to 0} \frac{\psi'(\eta)}{\psi'(\xi)} \frac{d\eta}{d\xi};$$

но $\psi'(0) = is$, $\dfrac{d\eta}{d\xi} = \dfrac{\lambda}{i}\psi'(\xi) \to \lambda s = \alpha$ при $\xi \to 0$, и мы действительно находим $\rho = 1 - \alpha$.

§ 40. Характеристическая функция времени ожидания

Целью нашего исследования является отыскание закона распределения случайной величины γ — времени ожидания для вызова, производимого в произвольно выбранный момент времени. Так как всякий закон распределения однозначно определяется соответствующей характеристической функцией, а отыскание важнейших свободных характеристик случайной величины по ее характеристической функции обычно бывает проще, чем по ее закону распределения, то мы естественно будем считать нашу цель достигнутой, если нам удастся найти характеристическую функцию $\varphi(\xi)$ величины γ, т. е. математическое ожидание величины $e^{i\gamma\xi}$, как функцию вещественного параметра ξ. К этому и будут направлены наши усилия.

Пусть в дальнейшем M есть символ математического ожидания, так что $\varphi(\xi) = \mathsf{M}e^{i\gamma\xi}$. Мы будем различать следующие возможности. Во-первых, в произвольно выбранный момент мы можем застать линию свободной; вероятность этого равна $1 - \alpha$, и в этом случае наверняка $\gamma = 0$, $e^{i\gamma\xi} = 1$. Во-вторых, мы можем застать линию занятой при k ожидающих ($k = 0, 1, 2, \ldots$); вероятность этого мы обозначим через $\mathsf{P}(k)$, а математические ожидания, вычисленные при этом условии, будем обозначать через M_k. Очевидно, мы имеем

$$\varphi(\xi) = \mathsf{M}e^{i\gamma\xi} = 1 - \alpha + \sum_{k=0}^{\infty} \mathsf{P}(k)\, \mathsf{M}_k e^{i\gamma\xi}.$$

Если наш вызов застал линию занятой при k ожидающих, то время ожидания γ для него слагается из двух частей: 1) остаточная длительность τ того разговора, который ведется в момент появления нашего вызова, и 2) суммарная длительность T разговоров тех k ожидающих, которых наш вызов застал при своем появлении. Мы имеем поэтому

$$\gamma = \tau + T, \quad \mathsf{M}_k e^{i\gamma\xi} = \mathsf{M}_k (e^{i\tau\xi} e^{iT\xi}).$$

Но очевидно, что *при данном k* величины τ и T взаимно независимы. Поэтому мы имеем

$$\varphi(\xi) = 1 - \alpha + \sum_{k=0}^{\infty} \mathsf{P}(k)\, \mathsf{M}_k e^{i\tau\xi} \mathsf{M}_k e^{iT\xi}.$$

Величина $\mathsf{M}_k e^{iT\xi}$ вычисляется непосредственно. В самом деле, величина T есть сумма k взаимно независимых случайных величин, распределенных по закону $F(x)$ с характеристической функцией $\psi(\xi)$. Поэтому

$$\mathsf{M}_k e^{iT\xi} = [\psi(\xi)]^k \quad (k = 0, 1, 2, \ldots)$$

и мы находим

$$\varphi(\xi) = 1 - \alpha + \sum_{k=0}^{\infty} [\psi(\xi)]^k \mathsf{P}(k) \mathsf{M}_k e^{i\tau\xi}.$$

Обозначим через $Q_k(x)$ закон распределения величины τ (вероятность неравенства $\tau > x$) при k ожидающих. Тогда

$$\mathsf{M}_k e^{i\tau\xi} = -\int_0^{\infty} e^{i\xi x} dQ_k(x);$$

$$\mathsf{P}(k) \mathsf{M}_k e^{i\tau\xi} = -\int_0^{\infty} e^{i\xi x} \mathsf{P}(k) dQ_k(x).$$

Но величина

$$\mathsf{P}(k)[Q_k(x) - Q_k(x + dx)]$$

есть, очевидно, вероятность того, что наш вызов застанет линию занятой при k ожидающих и что при этом будет $x < \tau < x + dx$. Эта вероятность в силу леммы 4 с точностью до малых высших порядков при $dx \to 0$ равна

$$-\lambda dx \int_x^{\infty} u_k(t - x)\, dF(t).$$

Поэтому мы получаем

$$\mathsf{P}(k) \mathsf{M}_k e^{i\xi\tau} = -\lambda \int_0^{\infty} e^{i\xi x} dx \int_x^{\infty} u_k(t - x)\, dF(t),$$

и следовательно,

$$\varphi(\xi) = 1 - \alpha - \lambda \int_0^{\infty} e^{i\xi x} dx \int_x^{\infty} \left\{\sum_{k=0}^{\infty} [\psi(\xi)]^k u_k(t - x)\right\} dF(t),$$

откуда в силу леммы 6

$$\varphi(\xi) = 1 - \alpha - \lambda \int_0^{\infty} e^{i\xi x} dx \int_x^{\infty} e^{\lambda(t-x)[\psi(\xi)-1]} \sum_{k=0}^{\infty} \pi_k [\psi(\xi)]^k dF(t) =$$

$$= 1 - \alpha - \lambda\chi(\xi) \int_0^{\infty} e^{i\xi x} dx \int_x^{\infty} e^{\lambda(t-x)[\psi(\xi)-1]} dF(t),$$

где положено, как в § 39, $\sum\limits_{k=0}^{\infty} \pi_k [\psi(\varepsilon)]^k = \chi(\xi)$. Если мы еще в согласии с § 39 положим

$$\frac{\lambda}{i}[\psi(\xi)-1]=\eta=\eta(\xi),$$

то получим

$$\begin{aligned}\varphi(\xi)&=1-\alpha-\lambda\chi(\xi)\int_0^\infty e^{i(\xi-\eta)x}dx\int_x^\infty e^{i\eta t}dF(t)=\\&=1-\alpha-\lambda\chi(\xi)\int_0^\infty e^{i\eta t}dF(t)\int_0^t e^{i(\xi-\eta)x}dx=\\&=1-\alpha-\lambda\chi(\xi)\int_0^\infty e^{i\eta t}dF(t)\frac{e^{i(\xi-\eta)t}-1}{i(\xi-\eta)}=\\&=1-\alpha-\frac{\lambda\chi(\xi)}{i(\xi-\eta)}\int_0^\infty[e^{i\xi t}-e^{i\eta t}]\,dF(t)=\\&=1-\alpha+\frac{\lambda}{i}\chi(\xi)\frac{\psi(\xi)-\psi(\eta)}{\xi-\eta}.\end{aligned}$$

Наконец, выражая $\chi(\xi)$ согласно лемме 7, получаем

$$\begin{aligned}\varphi(\xi)&=1-\alpha+(1-\alpha)\frac{\lambda}{i}\frac{\psi(\xi)-1}{\xi-\eta}=\\&=(1-\alpha)\left\{1+\frac{\lambda}{i}\frac{\psi(\xi)-1}{\xi-\frac{\lambda}{i}[\psi(\xi)-1]}\right\}=\\&=(1-\alpha)\frac{\frac{i\xi}{\lambda}}{\frac{i\xi}{\lambda}-[\psi(\xi)-1]}=\\&=\frac{1-\alpha}{1-\alpha\frac{\psi(\xi)-1}{is\xi}}.\end{aligned}$$

Полученная таким образом формула

$$\varphi(\xi)=\frac{1-\alpha}{1-\alpha\frac{\psi(\xi)-1}{is\,\xi}} \tag{40.1}$$

может считаться полным решением поставленной задачи, так как характеристическая функция $\varphi(\xi)$ искомого закона распределения времени ожидания нами выражена через данные постоянные s и $\alpha=\lambda s$ и через характеристическую функцию $\psi(\xi)$ данного закона распределения $F(t)$ длин разговоров.

В качестве первой иллюстрации формулы (40.1) рассмотрим, какое выражение с ее помощью получает среднее значение $\bar{\gamma}$ времени ожидания. Мы имеем в виду $\varphi(0)=1$

$$i\bar{\gamma}=\varphi'(0)=\frac{\varphi'(0)}{\varphi(0)};$$

а так как в силу (40.1)

$$\varphi(\xi) = \frac{(1-\alpha)\, is\xi}{is\xi - \alpha[\psi(\xi) - 1]},$$

то

$$\frac{\varphi'(\xi)}{\varphi(\xi)} = \frac{d \ln \varphi(\xi)}{d\xi} = \frac{1}{\xi} - \frac{is - \alpha\psi'(\xi)}{is\,\xi - \alpha\,[\psi(\xi) - 1]} =$$

$$= \frac{\alpha\,\{\xi\psi'(\xi) - [\psi(\xi) - 1]\}}{\xi\,\{is\,\xi - \alpha\,[\psi(\xi) - 1]\}}.$$

Предел этого выражения при $\xi \to 0$ легко находится по правилу Лопиталя и равен

$$\frac{\alpha\psi''(0)}{2(1-\alpha)\,is}.$$

Величина

$$\psi''(0) = \int_0^\infty t^2\, dF(t) = -s_2$$

есть взятое с обратным знаком математическое ожидание квадрата длины разговора. Мы находим таким образом

$$i\bar{\gamma} = -\frac{\alpha}{2i(1-\alpha)}\frac{s_2}{s},$$

или

$$\bar{\gamma} = \frac{\alpha}{2(1-\alpha)}\frac{s_2}{s}.$$

Эта простая формула между прочим показывает, что при данной загрузке линии α и при данной средней длине разговора s время ожидания будет в среднем тем меньше, чем меньше дисперсия длин разговоров, т. е. чем более стандартизованы эти длины.

В качестве второй иллюстрации рассмотрим простейший случай показательного распределения

$$F(t) = e^{-\beta t}$$

длин разговоров. Мы легко находим в этом случае

$$s = \frac{1}{\beta}, \quad \psi(\xi) = \frac{\beta}{\beta - i\xi} = \frac{1}{1 - i\xi s},$$

откуда

$$\frac{\psi(\xi) - 1}{is\,\xi} = \frac{1}{1 - i\xi s} = \psi(\xi),$$

и, значит, в силу (40.1)

$$\varphi(\xi) = (1-\alpha)\frac{1 - i\xi s}{1 - \alpha - i\xi s} = 1 - \alpha + \alpha\frac{1-\alpha}{1 - \alpha - i\xi s}. \tag{40.2}$$

Но

$$\frac{1-\alpha}{1-\alpha-i\xi s} = \frac{\dfrac{1-\alpha}{s}}{\dfrac{1-\alpha}{s} - i\xi}$$

есть характеристическая функция показательного закона распределения с параметром $(1-\alpha)/s$. Поэтому (40.2) показывает, что закон распределения времени ожидания есть

$$\mathsf{P}\{\gamma \geqslant t\} = (1-\alpha) E(t) + \alpha e^{-\frac{1-\alpha}{s} t} \quad (t \geqslant 0),$$

где

$$E(t) = \begin{cases} 1 \ (t \leqslant 0); \\ 0 \ (t > 0). \end{cases}$$

Это согласуется с формулой

$$\mathsf{P}\{\gamma > t\} = \Pi e^{-(n\beta-\lambda)t},$$

полученной нами в конце главы 10 для случая n линий. В самом деле, при $n=1$ мы имеем $\Pi = \alpha = \lambda s$, $n\beta - \lambda = \frac{1}{s} - \lambda = \frac{1-\alpha}{s}$. Член же $(1-\alpha) E(t)$ при $t > 0$ равен нулю, а при $t = 0$ обращается в $1 - \alpha = \mathsf{P}\{\gamma = 0\}$.

ЛИТЕРАТУРНЫЕ УКАЗАНИЯ

(числа в прямых скобках относятся к списку литературы в конце настоящих указаний на стр. 119).

§ 1. Эти элементарные сведения помещаются почти во всех современных курсах теории вероятностей; см., в частности, Гнеденко [1], Феллер [3], Фрай [4], Хинчин [5], Erlang [7].

§ 2. Повидимому, публикуется впервые.

§ 3, 4. См. § 1.

§ 5. См. Фрай [4].

§ 6. В этой форме публикуется впервые.

§ 7. Публикуется впервые.

§ 8. См. Redheffer [9], где в другой форме решается та же задача.

§ 9, 10. Относительно функции $\varphi_0(t)$ см. Palm [8]. Относительно функций $\varphi_k(t)$ $(k > 0)$ публикуется впервые.

§ 11. Публикуется впервые.

§ 12. Публикуется впервые.

§ 13. Понятие потока с ограниченным последействием принадлежит Пальму [8]. В остальном содержание параграфа публикуется впервые.

§ 14. См. Palm [8].

§ 15, 16. Публикуется впервые.

§ 17, 18, 19, 20. См. Феллер [3], Фрай [4], Erlang [7].

§ 21. Этаэргодическая теорема является частным случаем известных общих теорем эргодической теории. В такой редакции, повидимому, публикуется впервые.

§ 22, 23. Erlang [7].

§ 24. Публикуется впервые. Частные случаи см. Palm [8].

§ 25. Публикуется впервые.

§ 26. См. Palm [8].

§ 27. Публикуется впервые,

§ 28, 29, 30, 31, 32, 33. См. Palm [8].

§ 34, 35. См. Колмогоров [2], Феллер [3], Фрай [4], Erlang [7].

§ 36, 37. См. Erlang [7].

§ 38, 39, 40. Вся глава 11 представляет собой существенную переработку статьи автора [6].

СПИСОК ЛИТЕРАТУРЫ

1. Гнеденко Б. В., Курс теории вероятностей, изд. 2-е, Гостехиздат, 1954.
2. Колмогоров А. Н., Sur le problème d'attente, Мат. сборн., 1931, 38, № 1—2, 101—106.
3. Феллер В., Введение в теорию вероятностей и ее приложения, Изд. иностр. лит., 1952.
4. Фрай Т., Теория вероятностей для инженеров, Гостехиздат, 1934.
5. Хинчин А. Я., Асимптотические законы теории вероятностей, ОНТИ, 1936.
6. Хинчин А. Я., Математическая теория стационарной очереди, Мат. сборн. 1932, 39, № 4, 73—84.
7. Erlang A. K., The Life and Works of, The Copenhagen Telephone Company, 1948.
8. Palm C., Intensitätsschwankungen im Fernsprechverkehr, Ericsson Technics, 1943, 44, 1—189.
9. Redheffer R. M., A note on the Poisson law, Math. Magazine, 1953, 26, № 4, 185—188.

ОГЛАВЛЕНИЕ

Часть III

СИСТЕМЫ С ОЖИДАНИЕМ

Утверждено к печати
Математическим институтом
Академии Наук СССР

Редактор издательства *А. З. Рывкин*
Технический редактор *Т. А. Землякова*
Корректор *В. Т. Макаров*
Сдано в набор 11/III 1955 г. Подп. к печ. 19/V 1955.
РИСО № 10-11В. Т-04430. Издат. № 1005.
Тип. заказ № 1163. Формат бум. $70\times108^1/_{16}$.
Печ. л. 9,06 Уч.-издат. 8,2. Тираж 2500
Цена 5 р. 75 к.

Издательство Академии Наук СССР, Москва, Подсосенский пер., 21

2-я тип. Издательства Академии Наук СССР.
Москва, Шубинский пер., д. 10.