Лекция 14.09

Входящий поток требований

На прошлой лекции была рассмотрена общая схема системы массового обслуживания и перечислены входящие в нее элементы. Теперь нужно рассмотреть эти элементы более подробно. Начнем с входящего потока требований.

Входящий поток требований с математической точки зрения представляет собой разновидность потока событий (или случайного потока). Разберем это понятие более детально.

Потоком событий называют последовательность однородных событий, происходящих друг за другом в случайные моменты времени.

Геометрически поток событий можно представить как случайное множество точек числовой оси (см. рис. 1) образующие поток события называются элементарными событиями потока. Моменты их появления T_k представляют собой некоторые случайные величины.

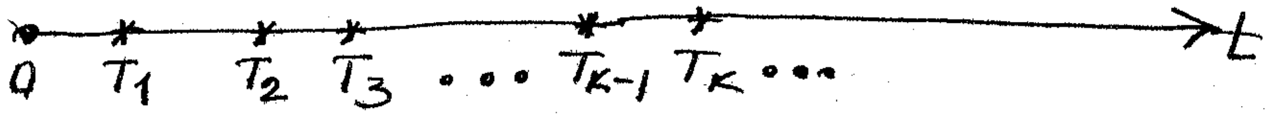

Рис. 1. К определению потока событий

Примеры случайных потоков могут быть чрезвычайно разнообразны. Важно только, чтобы моменты появления событий из потока были заранее неизвестны. Это может быть поток деталей, поступающих на обработку рабочему; поток корреспонденции, получаемой по электронной почте; поток читателей, заказывающих книги в библиотеке; поток шайб, заброшенных в хоккейном матче; поток стихийных бедствий (например, наводнений), происходящих в определенной местности и т.п.

С каждой СМО можно связать такой поток, элементарные события которого состоят в получении на вход системы последовательных требований. Такой поток называется <u>входящим потоком</u> данной СМО.

Математически входящий поток можно описать двумя способами: 1) Задать число появлений требований в любом фиксированном интервале. 2) Задать совместный закон распределения интервалов между требованиями. Вначале рассмотрим <u>первый</u> из указанных способов задания случайного потока.

Обозначим через $X(t_1,t_2)$ случайное число требований, попавших на интервал $[t_1, t_1+t_2]$. Для определения этой случайной величины нужно задать закон распределения X при любых $t_1 \geq 0$, $t_2 \geq 0$. Поскольку речь идет о дискретной целочисленной неотрицательной случайной величине, то придется задать бесконечный набор вероятностей

$$P_k(t_1,t_2) = \mathcal{P}\{X(t_1,t_2)=k\}, \quad (k=\overline{0,\infty}). \qquad (1)$$

Иначе говоря, нужно задать бесконечное множество функций вида (1) от двух аргументов при произвольных значениях t_1 и t_2.

Стационарные потоки и потоки без последействия

Описанная математическая модель получается весьма громоздкой, а практическое нахождение функций $P_k(t_1,t_2)$, например, из эксперимента, вызывает затруднение. Чтобы упростить модель, на поток обычно накладывают ряд физически ясных и практически приемлемых ограничений таким образом, чтобы поток можно было определить, задавая сравнительно небольшое конечное число функций и числовых параметров, которые легко поддаются экспериментальному определению.

Прежде всего выделяют так называемые стационарные потоки.

Поток событий называется стационарным, если все его вероятностные характеристики остаются неизменными при произвольном изменении начала отсчета времени.

В стационарном потоке все случайные величины $X(t_1,t_2)$, характеризуемые одним и тем же значением интервала наблюдения за потоком t_2, но разными значениями t_1, будут иметь один и тот же закон распределения. Иначе говоря, для стационарного потока

$$P_k(t_1,t_2)=P_k(t_2)=\mathrm{const}(t_1),\ (k=\overline{0,\infty}).\qquad(2)$$

Таким образом, в стационарном потоке придется задавать не функции двух аргументов (1), а функции одного аргумента (2).

Другим важным свойством случайных потоков является отсутствие в них последействия (то есть памяти). Рассмотрим два интервала числовой оси

$$\Omega_1=\{t:\ t_1\le t\le t_1+t_2\},\ \Omega_2=\{t:\ t_3\le t\le t_3+t_4\}\qquad(3)$$

и предположим, что они не пересекаются, то есть

$$\Omega_1\cap\Omega_2=\varnothing.\qquad(4)$$

Далее определим случайные величины

$$X_1=X(t_1,t_2),\quad X_2=X(t_3,t_4).\qquad(5)$$

Тогда говорят, что поток не имеет последействия, если случайные величины X_1 и X_2 не зависят друг от друга для любых непересекалощихся интервалов (3).

Отсутствие последействия означает, что поток *не имеет памяти*. Это можно пояснить следующим образом. Будем обозначать время через τ. Настоящий момент времени обозначим через t. Тогда значения $\tau > t$ будут обозначать будущие моменты времени, а значения $\tau < t$ будут относиться к прошлому (см. рис. 2).

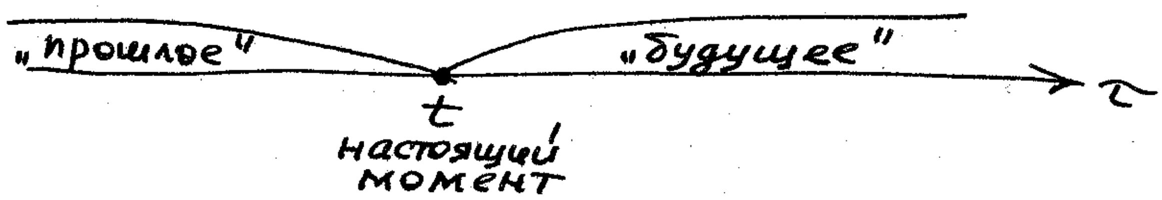

Рис. 2. К вопросу о памяти потока

Рассмотрим произвольный будущий интервал $[t_1, t_1+t_2] \subset (t, \infty)$ и произвольный уже прошедший интервал $[t_3, t_3+t_4] \subset [0, t)$. Если поток не имеет последействия, то поступление требований в двух указанных интервалах не зависит друг от друга, так как они не пересекаются (первый целиком лежит правее точки t, а второй целиком расположен левее этой точки). Следовательно, в потоке без последействия поступление заявок в будущем не зависит от того, как они поступали в прошлом, то есть поток не имеет памяти.

Здесь необходимо ввести одно важное понятие. Рассмотрим случайный процесс $U(t)$. *Случайным процессом* называют функцию аргумента t, значения которой при любом значении t являются *случайными величинами*. Обозначим через t настоящий момент времени, через τ – некоторый будущий момент, а через s – прошедший момент.

Обозначим ординату процесса в настоящий момент $U(t)$ через X, в будущий момент τ через $Y = U(\tau)$, а в прошедший момент s – $Z = U(s)$. Здесь моменты времени s, t, τ

удовлетворяют неравенству $s<t<\tau$, а все ординаты процесса X, Y, Z являются некоторыми случайными величинами, вообще говоря, связанными друг с другом.

Фиксируем настоящий момент времени t и ординату процесса $X=x$ в этот момент. Тогда, если задание t и x однозначно определяет закон распределения Y при любом $\tau>t$, то такой случайный процесс называется марковским процессом.

Если процесс $U(t)$ является марковским, то его ордината в будущий момент $\tau>t$ не зависит от ординаты $U(s)$ при любом $s<t$. Иначе говоря, будущее состояние марковского процесса полностью определяется его состоянием в настоящий момент и не зависит от того, каким образом это настоящее состояние было достигнуто.

Вернемся к рассмотрению случайных потоков. Введем случайный процесс $N(t)$, определяемый как общее число требований, полученных к моменту t

$$N(t)=X(0,t). \tag{6}$$

Если обозначить подобно рис. 1 через T_k момент прихода произвольного k-го требования в потоке, то функция $N(t)$ может быть графически представлена так, как это показано на рис. 3.

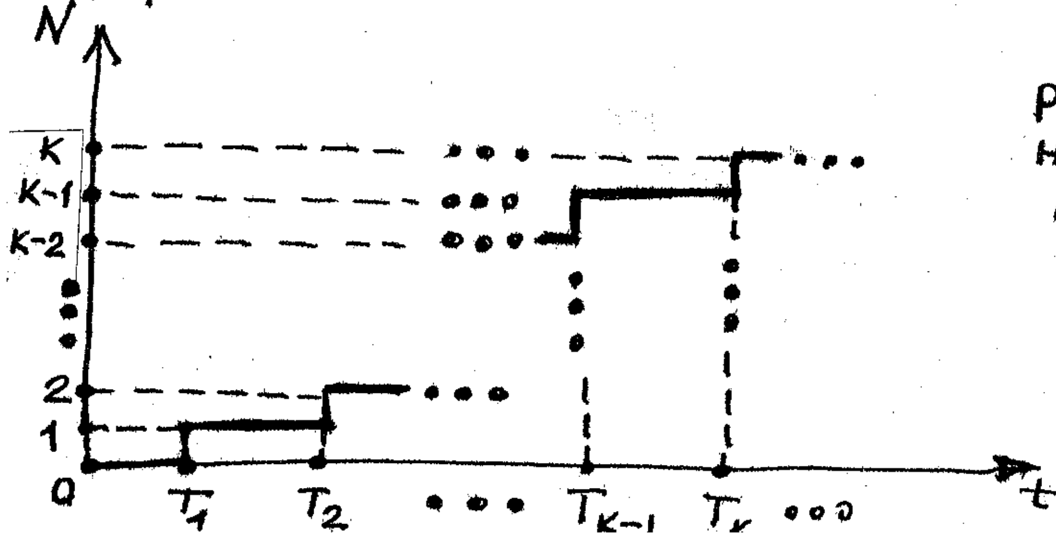

Рис. 3. К определению функции $N(t)=X(0,t)$

Из приведенных выше рассуждений следует, что для потока без последействия общее число поступивших требований $N(t)$ является **марковским процессом**. Данное свойство впервые было сформулировано А.А. Марковым, по имени которого оно впоследствии и было названо. Марковость процесса $N(t)$ и отсутствие последействия в исходном потоке — это полностью эквивалентные понятия.

Ординарные потоки требований. Понятие интенсивности потока

Понятие ординарности потока событий, с физической точки зрения имеет простой и ясный смысл: в ординарном потоке требования поступают по одиночке, не допускается их приход группами, одновременно по нескольку требований одновременно. Математически это требование нуждается в более строгом и корректном определении.

Строгая формулировка ординарности дается в терминах поступления заявок на бесконечно малых интервалах времени. Считается, что вероятность поступления на элементарном интервале длиной dt двух и более требований должна быть пренебрежимо малой по сравнению с вероятностью поступления на этом же бесконечно малом интервале одного требования.

Дадим строгое определение понятия ординарности. Для этого введем вероятность попадания на интервал $[t_1, t_1+t_2]$ не менее двух требований

$$P_{>1}(t_1,t_2) = \mathcal{P}\{X(t_1,t_2)>1\} = \sum_{k=2}^{\infty} P_k(t_1,t_2). \quad (7)$$

Далее рассмотрим бесконечно малый интервал $[t, t+dt]$ и предположим, что для него выполнялось условие

$$P_1(t,dt)=\lambda(t)dt+o(dt),\ P_{>1}(t,dt)=o(dt), \quad (8)$$

где $\lambda(t)$ – некоторая интегрируемая функция, а $o(dt)$ обозначает бесконечно малую величину более высокого порядка малости, чем dt. При соблюдении условий (8) поток событий будем называть ординарным.

Смысл условий (8) очевиден. Ясно, что при уменьшении длины интервала наблюдения вероятность получить одно требование должна уменьшаться. Согласно (8) при бесконечно малой длине интервала эта вероятность предполагается пропорциональной dt. В то же время вероятность получить два и более требований имеет более высокий порядок по dt, что и свидетельствует о пренебрежимой малости этой последней вероятности.

Выясним физический смысл коэффициента пропорциональности $\lambda(t)$ при dt в первом равенстве (8). Введем математическое ожидание числа требований $X(t_1,t_2)$, попавших на интервал $[t_1, t_1+t_2]$

$$\bar{x}(t_1,t_2)=M[X(t_1,t_2)]=\sum_{k=0}^{\infty} k P_k(t_1,t_2). \quad (9)$$

Допустим, что для рассматриваемого потока математическое ожидание (9) существует и чисто формально перепишем сумму (9) в виде

$$\bar{x}(t_1,t_2)=P_1(t_1,t_2)+\bar{x}_0(t_1,t_2)P_{>1}(t_1,t_2), \quad (10)$$

где введено обозначение

$$\bar{x}_0(t_1,t_2)=\sum_{k=2}^{\infty} k\frac{P_k(t_1,t_2)}{P_{>1}(t_1,t_2)}. \quad (11)$$

Рассмотрим величины

$$q_k(t_1,t_2)=\frac{P_k(t_1,t_2)}{P_{>1}(t_1,t_2)}, \quad (k\geq 2). \quad (12)$$

Ясно, что все величины (12) неотрицательны и удовлетворяют следующему условию нормировки:

$$\sum_{k=2} q_k(t_1,t_2) \equiv 1. \qquad (13)$$

Следовательно, q_k можно трактовать как некоторые вероятности. Очевидно $q_k(t_1,t_2)$ представляет собой вероятность получить ровно k требований в интервале $[t_1, t_1+t_2]$ но условную при условии, что $X(t_1,t_2) \geq 2$. Это означает, что $\overline{x}_0(t_1,t_2)$, определяемое (11), равно среднему числу полученных требований при том же условии.

Запишем равенство (10) для бесконечно малого интервала $[t, t+dt]$. Имеем

$$\overline{x}(t,dt) = p_1(t,dt) + \overline{x}_0(t,dt)\, p_{>1}(t,dt). \qquad (14)$$

Для большинства реальных потоков величина $\overline{x_0}(t,dt)$ остается ограниченной при $dt \to 0$. Допустим, что поток является ординарным и удовлетворяет условию (8), тогда

$$\overline{x}(t,dt) = \lambda(t)dt + o(dt). \qquad (15)$$

Теперь вновь обратимся к общему числу требований, полученных к моменту t, определяемому согласно (6) и подсчитаем среднее число таких требований

$$\overline{n}(t) = M[N(t)] = \overline{x}(0,t). \qquad (16)$$

Нетрудно понять, что

$$d\overline{n}(t) = \overline{n}(t+dt) - \overline{n}(t) = \overline{x}(t,dt). \qquad (17)$$

Следовательно, на основании (15) получим

$$d\overline{n}(t) = \lambda(t)dt + o(dt). \qquad (18)$$

Поделив обе части (18) на dt и переходя к пределу при $dt \to 0$, получаем дифференциальное уравнение

$$\frac{d\overline{n}(t)}{dt} = \lambda(t). \qquad (19)$$

Среднее число событий, происходящих в случайном потоке в единицу времени, называется <u>интенсивностью</u> этого потока. Отсюда следует, что функция $\lambda(t)$ в определении ординарного потока (8) есть нечто иное, как его интенсивность

Допустим, что наблюдение за потоком началось при $t=0$, причем в этот момент не было получено ни одного требования, то есть $N(0)=0$. Ясно, что тогда и среднее число требований

$$\bar{n}(0)=0 \qquad (20)$$

также равнялось нулю. Решая уравнение (19) при начальном условии (20), получаем

$$\bar{n}(t)=\int^{t}\lambda(s)ds=\Lambda(t), \qquad (21)$$

где функция

$$\Lambda(t)=\int_0^t\lambda(s)ds \qquad (22)$$

называется <u>ведущей функцией потока</u> (этот термин был введен А.Я. Хинчиным). Ведущая функция представляет собой среднее число требований потока, попавших на интервал $[0,t]$.

Рассмотрим частный случай, когда поток является <u>стационарным</u>. Для стационарного потока, очевидно, $\lambda=const(t)$ и тогда

$$\bar{n}(t)=\Lambda(t)=\lambda t, \qquad (23)$$

то есть среднее число требований будет расти по <u>линейному закону</u>.

Интенсивность стационарного потока сравнительно просто можно оценить по опытным данным. Рассмотрим достаточно большой по длительности интервал наблюдения за потоком $[0,T]$ и фиксируем все полученные на опыте требования. Обозначим их общее число в интервале $[0,T]$ через $N(T)$. Тогда в качестве оценки интенсивности λ можно принять

$$\hat{\lambda}=\frac{N(T)}{T}. \qquad (24)$$

Конечно, эта оценка будет случайной величиной, так как зависит от опытных данных.

Исследуем свойства оценки (24). Вычисляя математическое ожидание обеих частей равенства (24). Используя (23), будем иметь

?) $$M[\tilde{\lambda}] = \frac{\overline{n}(T)}{T} = \frac{\lambda T}{T} = \lambda. \qquad (25)$$

Следовательно, оценка (24) является несмещенной оценкой для λ. Подсчитаем дисперсию оценки $\tilde{\lambda}$. Имеем,

$$D[\tilde{\lambda}] = \frac{D[N(T)]}{T^2}. \qquad (26)$$

Для многих реальных потоков событий, как, например, для потоков Пальма, о которых речь пойдет чуть позднее, дисперсия $N(T)$ с ростом T растет по линейному закону, значит

$$\lim_{T\to\infty} D[\tilde{\lambda}] = 0 \qquad (27)$$

В математической статистике показывается, что при соблюдении условия (27) оценка является состоятельной, то есть сходится по вероятности к оцениваемому параметру λ. Таким образом, при достаточно большом интервале наблюдения T оценка $\tilde{\lambda}$ дает хорошую аппроксимацию для интенсивности потока λ.

В самом начале лекции мы говорили, что случайные потоки можно описывать двумя способами. Пока мы разобрали лишь первый из них, когда фиксируется интервал $[t_1, t_1+t_2]$ и исследуется число требований $X(t_1,t_2)$, попавших на этот интервал. Существует и другой способ описания случайных потоков, который часто оказывается даже более удобным. Разберем его более подробно.

Потоки Пальма

обратимся к рисунку 1, приведенному ранее и воспользуемся обозначением T_k для момента прихода k-го требования. Очевидно, разность

$$\tau_k = T_k - T_{k-1}, \quad (k=\overline{1,\infty}) \qquad (28)$$

имеет смысл интервала между k-м и предшествующим ему требованием. При k = 1 в формуле (28) формально принимается $T_0 = 0$.

Альтернативный способ задания случайных потоков состоит в том, что задается совместный закон распределения системы случайных величин τ_k. Этот способ описания потока особенно удобен, если поток является стационарным или не имеет последействия.

Пусть, например, поток ординарен и стационарен. Первое требование в нем будет получено через случайное время τ_1. Сдвинем начало отсчета времени на величину τ_1 и будем наблюдать за формированием второго требования. Оно придет через некоторое время τ_2. В силу стационарности потока условия поступления второго требования будут такими же, как для первого требования. Следовательно, закон распределения τ_2 должен совпадать с законом распределения τ_1. Эти рассуждения можно повторить для любого номера интервала между требованиями τ_k. Поэтому в стационарном потоке все интервалы τ_k будут распределены по одному и тому же закону.

Теперь обратимся к потоку без последействия. Допустим, что получено уже k требований. Выделим интервал, непосредственно следующий за моментом прихода k-го требования t_k. Благодаря отсутствию последействия появление новых требований после момента T_k не будет зависеть от моментов прихода всех ранее полученных требований. Значит, очередной $(k+1)$-й интервал не будет зависеть от всех предшествующих интервалов $\tau_1, \ldots, \tau_k$. Если повторить эти рассуждения для всех k, то получим, что в потоке без последействия вся система случайных интервалов $\{\tau_k\}_{k=1}^{\infty}$ независима в совокупности.

Независимость и одинаковая распределенность интервалов в потоке являются чрезвычайно удобными и весьма желательными свойствами случайных потоков. В связи с этим шведский математик К. Пальм ввел важный класс потоков, в которых, во-первых, все интервалы независимы и, во-вторых, они одинаково распределены. Такие потоки называют потоками Пальма (иногда используют также термин рекуррентные потоки).

Допустим, что имеется поток Пальма в котором все интервалы распределены по некоторому заданному закону $a(\tau)$. Важнейшими характеристиками такого потока являются: средний интервал между последовательными требованиями

$$\bar{\tau} = M[\tau_k] = \int_0^{\infty} \tau a(\tau) d\tau, \qquad (29)$$

а также дисперсия интервала между требованиями

$$\sigma_\tau^2 = D[\tau_k] = M[(\tau_k - \bar{\tau})^2] = $$
$$= \int_0^{\infty} (\tau - \bar{\tau})^2 a(\tau) d\tau. \qquad (30)$$

Отметим, что обе величины (29) и (30)

в потоке Пальма не зависит от номера интервала k.

Важнейшей характеристикой потока Пальма является коэффициент вариации интервала между требованиями

$$\gamma_\tau = \frac{\sigma_\tau}{\bar{\tau}}, \qquad (31)$$

который служит безразмерной мерой рассеяния требований в потоке. Он, конечно, тоже одинаков для всех интервалов τ_k.

Модель потока Пальма является одной из основных в ТМО, и мы ее изучим более подробно в последующих лекциях