Потоки событий

Потоком событий называется последовательность событий, происходящих одно за другим в какие-то случайные моменты времени. Различают потоки однородных и неоднородных событий. *Поток событий* называется *однородным,* если он характеризуется только моментами поступления этих событий (вызывающими моментами) и задается последовательностью $\{t_n\} = \{0 \le t_1 \le t_2 \le ... \le t_n \le \ldots\}$, где t_n — момент наступления n-го события — неотрицательное вещественное число. Однородный поток событий также может быть задан в виде последовательности промежутков времени между n-м и $(n—1)$-м событиями $\{\tau_n\}$, которая однозначно связана с последовательностью вызывающих моментов $\{ t_n \}$, где $\tau_n = t_n - t_{n-1}$.

Потоком неоднородных событий называется последовательность $\{t_n, f_n\}$, где t_n—вызывающие моменты; f_n —набор признаков события. Например, применительно к процессу обслуживания для неоднородного потока заявок может быть задана принадлежность к тому или иному источнику заявок, наличие приоритета, возможность обслуживания тем или иным типом канала и т. п.

Рассмотрим поток, в котором события разделены интервалами времени τ_1, τ_2, …, которые вообще являются случайными величинами. Пусть интервалы τ_1, τ_2, ... независимы между собой. Тогда поток событий называется потоком с *ограниченным последействием*.

Пусть на оси времени 0t случайным образом возникают точки – моменты появления каких-то однородных событий.

Поток событий называется *ординарным,* если вероятность того, что на малый интервал времени Δt примыкающий к моменту времени t, попадает больше одного событий $P_{>1}(t, \Delta t)$, пренебрежительно мала по сравнению с вероятностью того, что на этот же интервал времени Δt попадает ровно одно событие $P_1(t, \Delta t)$, т. е. $P_1(t, \Delta t) >> P_{>1}(t, \Delta t)$.

Стационарным потоком событий называется поток, для которого вероятность появления того или иного числа событий на интервале времени τ зависит лишь от длины этого участка и не зависит от того, где на оси времени Ot взят этот участок.

Рассмотрим на оси времени Ot ординарный поток событий и найдем среднее число событий, наступающих на интервале времени Δt, примыкающем к моменту времени t. Получим

$$0 \cdot P_0(t, \Delta t) + 1 \cdot P_1(t, \Delta t) = P_1(t, \Delta t).$$

Тогда среднее число событий, наступающих на участке времени Δt в единицу времени, составит $[P_{>1}(t,\Delta t)]/\Delta t$. Рассмотрим предел этого выражения при $\Delta t \to 0$. Если этот предел существует, то он называется *интенсивностью (плотностью) ординарного потока событий*

$$\lim_{\Delta t \to 0} [P_1(t, \Delta t)/\Delta t] = \lambda(t).$$

Интенсивность потока может быть любой неотрицательной функцией времени, имеющей размерность, обратную размерности времени. Для стационарного потока его интенсивность не зависит от времени и представляет собой постоянное значение, равное среднему числу событий, наступающих в единицу времени $\lambda(t) = \lambda =$ const.

Поток событий называется *потоком без последействия*, если для любых неперекрывающихся интервалов времени число событий, попадающих на один из них не зависит от числа событий, попадающих на другие интервалы.

Поток, обладающий тремя перечисленными выше свойствами, а именно стационарностью, ординарностью и отсутствием последействия, называется ***простейшим потоком***.

Разделим участок τ на n частей так, чтобы $\Delta t=\tau/n$. Вероятность появления единицы на элементарном участке Δt определяется так:

$$P(1, \Delta t)=\lambda\Delta t=\lambda\tau/n.$$

Представим n участков как n независимых испытаний. Такое представление справедливо в силу свойства отсутствия последействия для простейшего потока. В каждом испытании событие может произойти с вероятностью $P(1, \Delta t) =\lambda\tau/n$. Число наступления события в течении времени τ есть случайная величина, распределенная по биноминальному закону с вероятностью

$$P(1,\Delta t) = C_n^k\left(\frac{\lambda\tau}{n}\right)^k\left(1-\frac{\lambda\tau}{n}\right)^{n-k}$$

При неограниченном увеличении числа интервалов Δt, т.е. при $n\to\infty$ и $\Delta t\to 0$ биноминальное распределение сходится к распределению Пуассона и вероятность определяется так:

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

Рассмотрим случайную величину T – интервал времени между соседними событиями.

$$F(t) = P(T < t) = P_{\geq 1}(t) = 1 - P_0(t) = 1 - e^{-\lambda t}\,.$$

Таким образом интервал времени t между соседними событиями в простейшем потоке есть случайная величина, распределенная по показательному закону с параметром λ:

$$f(t)=\begin{cases} \lambda e^{-\lambda t},\ t>0 \\ 0,\ t<0 \end{cases}$$

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

Поток Пальма часто получается в виде выходных потоков систем массового обслуживания. Если на какую-нибудь систему поступает поток заявок, то он разделяется на два потока: поток обслуженных и поток необслуженных заявок. Поток необслуженных заявок может поступать н некоторую другую систему массового обслуживания, поэтому представляет – интерес рассмотреть его свойства. Основной в теории выходных потоков является теорема Пальма:

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

Примером потока с ограниченным последействием являются *потоки Эрланга.* Они образуются «просеиванием» из простейшего потока.

Потоком Эрланга k-го порядка ($Э_k$) называется поток, получаемый из простейшего, если сохранить каждую (k+1)-ю точку, а остальные выбросить. Очевидно, что простейший поток можно рассматривать как поток Эрланга нулевого порядка. Случайная величина интервала времени между соседними событиями в потоке Эрланга может быть представлена как сумма k случайных величин интервалов времени между соседними событиями в исходном простейшем потоке: $T = \sum_{k=1}^{k+1} T_i$. Плотность распределения этой случайной величины (закон распределения Эрланга k-го порядка) имеет вид:

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

Числовые характеристики распределения Эрланга определяются через свойства числовых характеристик случайных величин:

$$m_k = (k+1)m_0; \qquad D_k=(k+1)D_0.$$

Подставив в эти выражения значения математического ожидания и дисперсии случайной величины T_0 интервалов времени между соседними событиями исходного простейшего потока, получаем:

$$m_k = \frac{(k+1)}{\lambda}; \qquad D_k = \frac{k+1}{\lambda^2}.$$

Из этих выражений следует, что с увеличением порядка потока Эрланга увеличиваются математическое ожидание промежутка времени между соседними событиями, а интенсивность потока падает. Пронормеруем величину T_k таким образом, чтобы математическое ожидание нормированной случайной величины не изменялось при увеличении порядка потока Эрланга: $T^* = T_k/(k+1)$. Такой поток называется нормированным потоком Эрланга k-го порядка. Закон распределения промежутка T^* между событиями для этого потока будет:

$$f_k^*(t) = \frac{\Lambda_k (\Lambda_k t)^k}{k!} e^{-\Lambda_k t}, \qquad \text{где } \Lambda_k = (k+1)\Lambda.$$

Математическое ожидание случайной величины T^* не зависит от k и равно: $m_k^* = m_0 = \frac{1}{\lambda}$, где λ – интенсивность исходного простейшего потока.

Дисперсия случайной величины T^* равна

$$D_k^* = \frac{D_k}{(k+1)^2} = \frac{1}{\lambda^2 (k+1)}$$

и неограниченно убывает с увеличением k.

Таким образом, при неограниченном увеличении k нормированный поток Эрланга приближается к регулярному потоку с постоянными интервалами, равными 1/λ.

Кроме характеристик входного потока заявок, режим работы СМО зависит также от производительности самой системы: числа каналов n и быстродействия каждого канала.

Одной из важнейших характеристик, связанных с производительностью системы является *время обслуживания одной заявки $T_{об}$*, которая может быть как случайной так и регулярной величиной.

Для практики особый интерес представляет случай, когда закон распределения времени обслуживания является экспоненциальным с параметром μ:

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

где μ – величина, обратная среднему времени обслуживания заявки.

Особая роль, которую играет в теории массового обслуживания показательный закон распределения времени обслуживания связана со следующим свойством: *если в какой-то момент времени t_0 происходит обслуживание заявки, то закон распределения* <u>*оставшегося*</u> *времени обслуживания не зависит от того, сколько времени обслуживание* <u>*уже продолжалось*</u>.

Показательный закон не является универсальным законом распределения времени обслуживания. Часто время обслуживания лучше описывается другими законами, например, нормальным или Эрланга. Однако, пропускная способность и другие характеристики СМО сравнительно мало зависят от вида закона распределения времени обслуживания, а зависят, главным образом, от его среднего значения. Поэтому, в теории массового обслуживания чаще всего пользуются допущением, что время обслуживания распределено по показательному закону.