Тема 12 Модели и системы массового обслуживания

Случайные процессы.

Случайный процесс называется процессом с дискретным временем, если переходы системы из состояния в состояние возможны только в строго определенные, заранее фиксированные моменты времени. В промежутке времени между этими моментами система сохраняет свое состояние.

Случайный процесс называется процессом с непрерывным временем, если переход системы из состояния в состояние возможен в любой случайный момент времени.

Многие операции развиваются как случайные процессы, ход и исход которых зависит от ряда случайных факторов, сопровождающих эти операции. Для математического описания таких процессов может быть применен аппарат марковских случайных процессов.

Примерами случайных процессов могут быть:

- процесс функционирования ЭВМ;
- процесс обслуживания клиентов;
- процесс выполнения плана снабжения…

Случайный процесс называется марковским, если он обладает следующими свойствами:

- для каждого момента времени t_0 вероятность любого состояния системы в будущем зависит только от её состояния в настоящем и не зависит от того, когда и каким образом система пришла в это состояние (т.е., как развивался процесс в прошлом), по-другому - будущее не зависит от “предыстории” процесса.

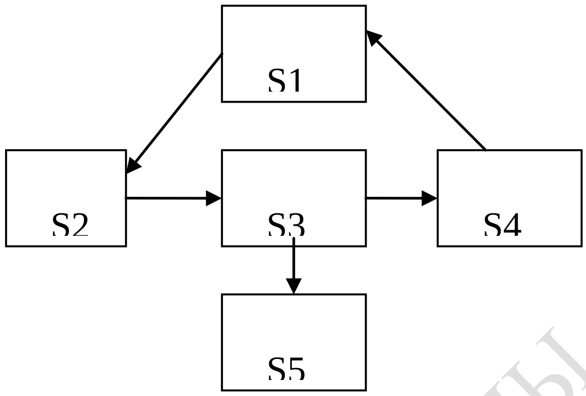
Марковские процессы можно разделить на классы по виду случайного процесса:

- с дискретными состояниями, если состояния можно перечислить, а переход из одного состояния в другое происходит скачком (мгновенно);
- с непрерывными состояниями, если характерен постепенный, плавный переход из состояния в состояние, например, процесс изменения напряжения в осветительной сети.

При анализе случайных процессов с дискретными состояниями удобно пользоваться геометрической схемой – так называемым графом состояний. Граф состояний изображает возможные состояния системы и её возможные переходы из состояния в состояние.

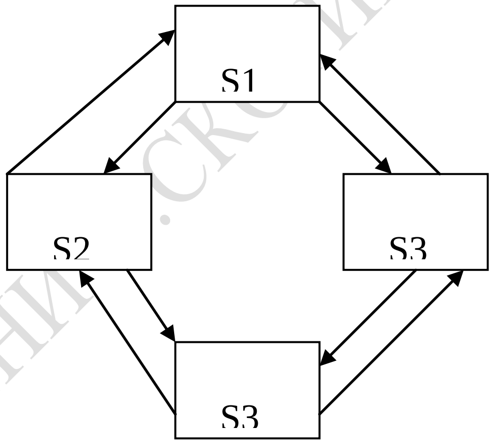
Пример 1: Система S – автомашина, которая переходит в одно из следующих состояний:

- S1 - исправна (работает),
- S2 - неисправна (ожидает ремонта),
- S3 - осмотр,
- S4 - ремонт,
- S5 - списана.

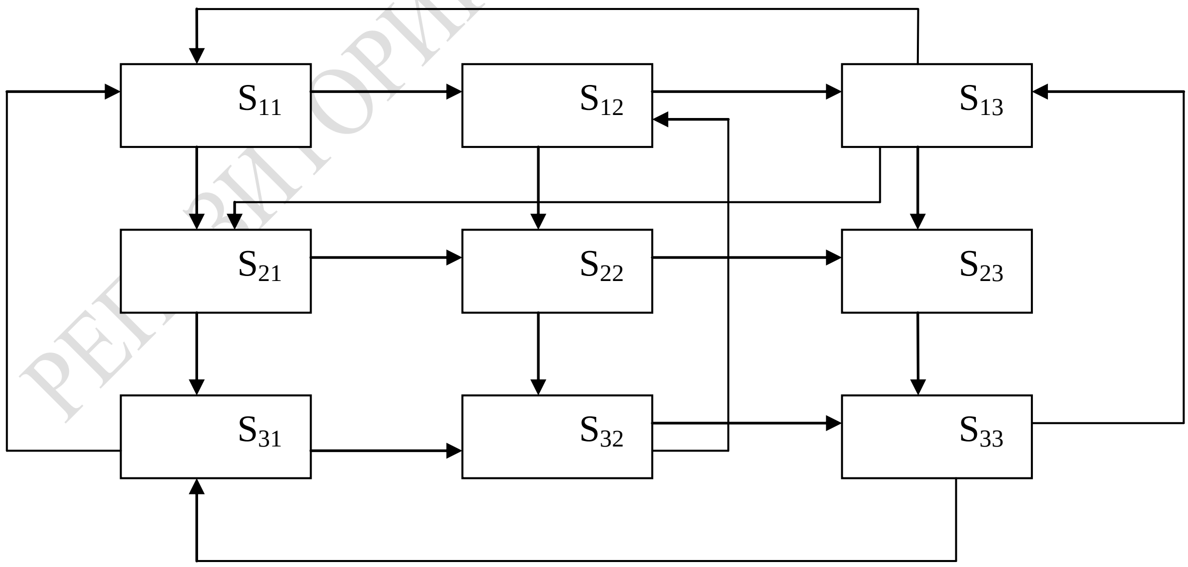


Пример 2. Система S представляет собой устройство, состоящее из 2 узлов, каждый из которых в какой-то момент времени может отказать. Отказавший узел немедленно начинают восстанавливать. Возможные состояния системы:

- S1 - оба узла работают,
- S2 - первый восстанавливается, второй работает,
- S3 - второй восстанавливается, первый работает,
- S4 - оба узла восстанавливаются.



Пример 3. В условиях предыдущего примера каждый узел перед восстановлением подвергается осмотру с целью локализации неисправности. Теперь нужно добавить ещё одно состояние для функции осмотра. Введя индексы i_1- работает, i_2 – осмотр, i_3 – восстановление, получим следующий граф состояний:



Рассмотрим марковский процесс с дискретными состояниями и дискретным временем. Состояния системы обозначим $S_1, S_2, \dots, S_n$, а моменты переходов $t_1, t_2, \dots, t_k$. Так что шаги (этапы) процесса можно рассматривать как функцию целочисленного аргумента – номера шага. Тогда случайный процесс системы может быть таким:

S_1 $S_2 \longrightarrow S_1 \longrightarrow S_2 \longrightarrow S_3 \longrightarrow S_4 \longrightarrow S_1 \longrightarrow \ldots \longrightarrow$

Если $S_i^{(k)}$ включают в себя все состояния системы, то сами события образуют полную группу и несовместны. Процесс, представимый как последовательность событий, называется <u>марковской цепью</u>, если для каждого шага вероятность перехода из S_i в S_j не зависит от того, когда и как система пришла в S_i. Такая цепь может быть описана с помощью вероятностей состояния $P(S_n^{(k)})$, $\sum_i P_i^{(k)} = 1$ $P_i^{(k)}$ = 1 для любого k, т.к. это вероятности несовместных событий, образующих полную группу.

Марковскую цепь можно представить себе так, как будто точка, изображающая систему S, случайным образам перемещается по графу состояний.

Для любого шага k существуют вероятности перехода из любого состояния в любое другое, такие вероятности называются переходными вероятностями марковской цепи.

Если переходные вероятности не зависят от номера шага k, то марковская цепь однородная, в противном случае – неоднородная (нестационарная).

Для однородной цепи переходные вероятности могут быть записаны как условные Pij = P ($S_j^{(k)} / S_i^{(k-1)}$) и $\sum_j P_{ij} = 1$,

так как события несовместны и образуют полную группу.

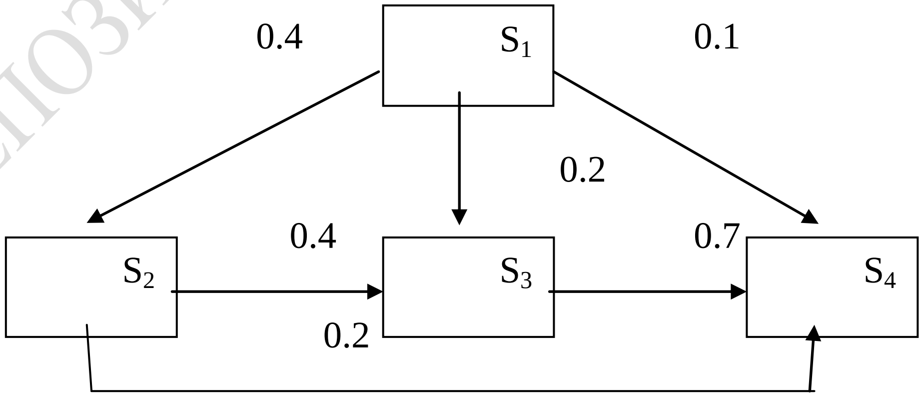
Если на графе состояний у стрелок перехода проставить соответствующие переходные вероятности, то такой граф называется <u>размеченным графом состояний</u>.

Имея в расположении размеченный граф и зная начальное состояние можно найти вероятности состояний $P_i^{(k)}$ после любого шага k.

По формуле полной вероятности

$$P_i(k) = \sum_j P_j(k-1)\, P_{ji}\,,\; j = 1,\ldots,n, \quad \sum_i P_i(k) = 1$$

Пусть дана однородная марковская цепь, представленная графом:



На графе не проставлены стрелки перехода из состояния в само себя, т.к. эта вероятность определяется как дополнение до единицы, т.е. $\sum_j P_{ij} = 1$.

Из рисунка имеем матрицу переходных вероятностей:

$$P_{ij} = \begin{array}{c|cccc} i \downarrow \ \ j \rightarrow & 1 & 2 & 3 & 4 \\ \hline 1 & 0.3 & 0.4 & 0.2 & 0.1 \\ 2 & & 0.4 & 0.4 & 0.2 \\ 3 & & & 0.3 & 0.7 \\ 4 & & & & 1 \end{array}$$

Пусть система находится в начальном состоянии S_1, т.е. $P_1(0) = 1$.Тогда

$$P_1(1) = \sum_j p_j(0)P_{j1} = P_1(0)*0,3 = 0,3$$

$$P_2(1) = \sum_j p_j(0)P_{j2} = P_1(0)*0,4 = 0,4$$

$$P_3(1) = \sum_j p_j(0)P_{j3} = P_1(0)*0,2 = 0,2$$

$$P_4(1) = \sum_j p_j(0)P_{j4} = P_1(0)*0,1 = 0,1$$

Для второго шага:

$$P_1(2) = \sum_j p_j(1)P_{j1} = P_1(1)*P_{11} = 0,3*0,3 = 0,09$$

$$P_2(2) = \sum_j p_j(1)P_{j2} = P_1(1)*P_{12} + P_2(1)*P_{22} = 0,3*0,4 + 0,4*0,4 = 0,28$$

$$P_3(2) = \sum_j p_j(1)P_{j3} = P_1(1)*P_{13} + P_2(1)*P_{23} + P_3(1)*P_{33} = 0,3*0,2 + 0,4*0,4 + 0,2*0,3 = 0,28$$

$$P_4(2) = \sum_j p_j(1)P_{j4} = P_1(1)*P_{14} + P_2(1)*P_{24} + P_3(1)*P_{34} + P_4(1)*P_{44} = 0,3*0,1 + 0,4*0,2 +$$
$$+ 0,2*0,2 + 0,1*1 = 0,35$$

$$\sum P_i(k) = 1$$

$P_i(k) =$ строка $P_j(k-1)*$ столбец P_{ji}

Если представить приведенную систему, как модель последовательной стрельбы по цели и обозначить:

S1 – цель невредима,

S2 – цель незначительно повреждена,

S3 – цель получила существенные повреждения,

S4 – цель поражена,

и предположить, что число выстрелов 4, то, продолжая аналогично, получим

$$P_i(3) = [0,027 \quad 0,148 \quad 0,214 \quad 0,611],$$

$P_i(4) = [0{,}0081 \quad 0{,}07 \quad 0{,}1288 \quad 0{,}7931]$.

Так что вероятности исходов обстрела цели берутся из матрицы $P_i(4)$.

Для неоднородной марковской цепи матрица переходных вероятностей зависит от номера шага (k), так, что должны быть определены $P_{i,j}^{(k)}$ для всех шагов. В остальном схема вычислений не меняется. Имеем:

$$P_{i,j}^{(k)} = P(S_j^{(k)}) / S_i^{(k-1)} \quad \text{и} \quad P_i(k) = \sum_j P_j^{(k-1)} P_{j,i}^{(k)},\ i = 1, n.$$

Пусть для предыдущего графа состояний заданы следующие матрицы: $P_{i,j}^{(1)} = P_{i,j}$

$$P_{i,j}^{(2)} = \begin{bmatrix} 0.1 & 0.4 & 0.3 & 0.2 \\ & 0.2 & 0.5 & 0.3 \\ & & 0.2 & 0.8 \\ & & & 1 \end{bmatrix} \qquad P_{i,j}^{(3)} = \begin{bmatrix} 0.05 & 0.3 & 0.4 & 0.25 \\ & 0.1 & 0.6 & 0.3 \\ & & 0.1 & 0.9 \\ & & & 1 \end{bmatrix}$$

Тогда из примера выше:

$P_i(1) = [0{,}3 \quad 0{,}4 \quad 0{,}2 \quad 0{,}1]$,

$P_i(2) = [0{,}03 \quad 0{,}2 \quad 0{,}33 \quad 0{,}44]$,

$P_i(3) = [0{,}02 \quad 0{,}029 \quad 0{,}165 \quad 0{,}786]$.

Схема марковского процесса с дискретными состояниями и непрерывным временем называется – непрерывной Марковской цепью (цепью Маркова).

Обозначив через $P_i(t)$ вероятность того, что система S будет находиться в состоянии S_i в момент времени t для любого момента времени t будем иметь: $\sum_i p_i(t) = 1$, так как $\{S_i\}$ независимы и образуют полную группу.

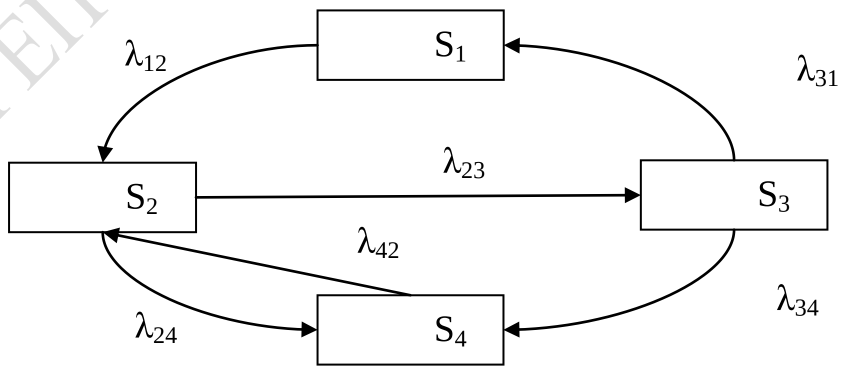
Вместо переходных вероятностей введем в рассмотрение плотность вероятностей перехода λ_{ij}:

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t},$$

так что $p_{ij}(\Delta t) \approx \lambda_{ij}\Delta t$ при малом Δt. Если λ_{ij} не зависит от t, то марковский процесс однороден, иначе неоднороден.

В размеченном графе состояний вместо переходных вероятностей проставим λ_{ij} – плотность переходных вероятностей. Пусть размеченный граф имеет вид:

Для $p_1(t)$ – вероятность того, что система в момент t будет в состоянии S_1:



а) – в момент t система была в S_1, а за время Δt из него не вышла;

б) – в момент t система была в S_3, а за время Δt перешла в S_1

Вероятность а) $p_1(t)*(1-\lambda_{12}\Delta t)$, вероятность б) $p_3(t)*\lambda_{31}\Delta t$.

По правилу сложения вероятностей

$p_1(t+\Delta t) = p_1(t)*(1-\lambda_{12}\Delta t)+ p_3()*\lambda_{31}\Delta t$

$$\frac{p_1(t+\Delta t)-p_1(t)}{\Delta t} = -\lambda_{12}p_1(t)+\lambda_{31}p_3(t) = \frac{dp_1(t)}{dt}.$$

Для состояния S_2 найдём $P_2(t+\Delta t)$:

1) система была в S_2 и за Δt не перешла в S_3 или S_4,

2) система была в S_1 и за Δt перешла в S_2,

3) система была в S_4 и за Δt перешла в S_2,

1) $P_2(t)\cdot(1-\lambda_{23}\Delta t-\lambda_{24}\Delta t)$,

2) $P_1(t)\cdot\lambda_{12}\Delta t$,

3) $P_4(t)\cdot\lambda_{42}\Delta t$.

$$P_2(t+\Delta t) = P_2(t)\cdot(1-\lambda_{23}\Delta t-\lambda_{24}\Delta t) + P_1(t)\cdot\lambda_{12}\Delta t + P_4(t)\cdot\lambda_{42}\Delta t$$

$$\frac{dP_2}{dt} = -P_2(t)\cdot(\lambda_{23}+\lambda_{24}) + P_1(t)\lambda_{12} + P_4(t)\lambda_{42}.$$

Аналогично:

$$\frac{dP_3}{dt} = -\lambda_{31}P_3(t)-\lambda_{34}P_3(t)+\lambda_{23}P_2(t),$$

$$\frac{dP_4}{dt} = -\lambda_{41}P_4(t)+\lambda_{24}P_2(t)+\lambda_{34}P_3(t).$$

Получили четыре дифференциальных уравнения, образующих полную систему, решение которой определяется начальными условиями при t=0: номер состояния S_i и $(0\ldots1_i,0\ldots)=P_i$, кроме того, $\sum P_i = 1$ указывает на дополнительное соотношение.

Уравнения для вероятности состояний в дифференциальной форме называются **уравнениями Колмогорова**

$$\frac{dPi}{dt} = \sum \lambda_{ij}P_i + \sum_i \lambda_{ji}P_i.$$

В левой части каждого уравнения стоит производная вероятности события, а правая часть содержит столько членов, сколько стрелок связаны с данным состоянием. Если стрелка направлена из состояния, соответствующий член имеет знак минус, если в состояние, то знак плюс. Каждый член равен произведению плотности вероятностей перехода, соответствующего данной стрелке, умноженной на вероятность того состояния, из которого исходит стрелка.

Уравнения Колмогорова имеют достаточно простое решение для случая $\lambda_{i,j}$ = const , т.е. интенсивности потоков не меняются со временем, такие потоки называются **простейшими или стационарными пуассоновскими**.

Если число состояний конечно и из каждого состояния можно перейти в другое, то при $t\to\infty$ функции $p_i(t) \to p_i$. Вероятности p_i называются **финальными или предельными вероятностями** состояний и не зависят от начального состояния системы, т.е. в системе S устанавливается предельный стационарный режим, переход в каждое из состояний осуществляется с некоторой постоянной вероятностью. Она представляет собой среднее относительное время пребывания системы в данном состоянии.

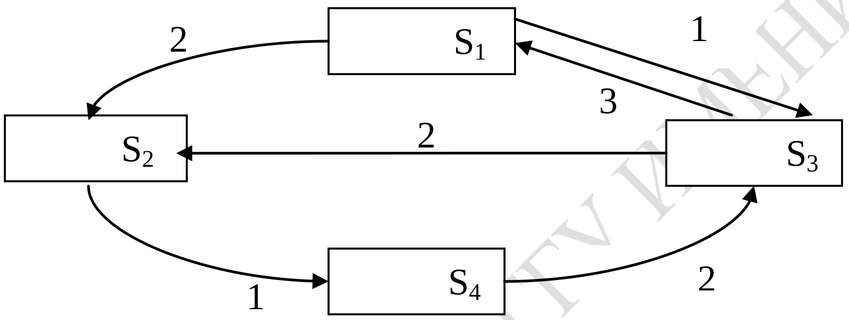
Отсюда следует, что в системе Колмогорова все производные равны нулю и система превращается в систему обычных алгебраических уравнений. В этой системе одно из уравнений зависит от других, так как система связана соотношением $\sum p_i = 1$. Поэтому, добавив к системе условие нормировки и исключая произвольно одно из уравнений системы, получим систему из n алгебраических уравнений относительно n-переменных.

Для записи финальных уравнений полезно мнемоническое правило:

«что втекает, то и вытекает».

То есть, для каждого состояния сумма членов, соответствующих входящим стрелкам, равна сумме членов, соответствующих выходящим, а каждый член равен интенсивности потока событий, переводящего систему по данной стрелке, умноженной на вероятность состояния, из которого выходит стрелка.

Пусть задан размеченный граф системы:



$$\left.\begin{aligned}(2+3)P_1 &= 1P_3\\ 1P_2 &= 2P_1 + 2P_3\\ (1+2)P_3 &= 3P_1 + 2P_4\\ 2P_4 &= 1P_2\end{aligned}\right\}$$

Имеем:

$$\left.\begin{aligned}P_3 &= 5P_1\\ P_2 &= 2P_4\\ \sum P_i &= 1\end{aligned}\right\} \qquad \left.\begin{aligned}2P_4 = 2P_1 + 5\cdot 2P_1 &= 12P_1\\ P_1 + 12P_1 + 5P_1 + P_4 &= 1\end{aligned}\right\}$$

Т.к. $P_4 = 6P_1$, то $24P_1 = 1$, отсюда

$$P_1 = {}^1/_{24} = {}^1/_{24} \quad P_4 = {}^1/_4 = {}^6/_{24}$$

$$P_2 = {}^1/_2 = {}^{12}/_{24} \quad P_3 = {}^5/_{24} = {}^5/_{24}, \quad \sum P_i = 1$$

В стационарном режиме система будет в S_1 ${}^1/_{24}$ части времени, в состоянии S_2 ${}^1/_2$ часть времени, в S_3 - ${}^5/_{24}$ и в S_4 - ${}^1/_4$ часть времени.

Вычисления:

$$\left.\begin{array}{c} -5P_1 + P_3 = 0 \\ -P_2 + 2P_1 + 2P_3 = 0 \\ -3P_3 + 3P_1 + 2P_4 = 0 \\ -2P_4 + P_2 = 0 \end{array}\right| \qquad P_3 = 5P_1$$

$$\left.\begin{array}{c} -P_2 + 2P_1 + 10P_1 = 0 \\ -15P_1 + 3P_1 + 2P_4 = 0 \\ -2P_4 + P_2 = 0 \end{array}\right| \qquad P_2 = 2P_4$$

$$\left.\begin{array}{c} -2P_4 + 12P_1 = 0 \\ -12P_1 + 2P_4 = 0 \end{array}\right| \qquad P_4 = 6P_1$$

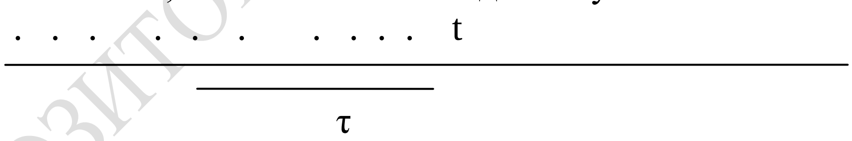
$$-12P_1 + 12P_1 = 0$$

Поток событий

Потоком событий называется последовательность однородных событий, следующих одно за другим в случайные моменты времени (поток вызовов на телефонной станции, поток заявок обслуживания в Интернете, поток посетителей в магазине, поток неисправности (сбоев) вычислительных машин, поток выстрелов, направляемых на цель …)

Поток событий в системе называется **регулярным**, если события следуют одно за другим через строго определённые промежутки времени (тактовая частота, поток синхроимпульсов …). Реальные потоки событий в исследуемых операциях характеризуются случайными моментами наступления событий.

1. Поток событий в системе называется **стационарным**, если вероятность попадания заданного числа событий на участок времени длиной τ не зависит от его положения, а зависит лишь от длины участка.



2. Поток событий называется потоком **без последствия**, если для любых непересекающихся участков времени число событий, попадающих на один из них, не зависят от того, сколько событий попало на другой.

3. Поток событий называется **ординарным**, если вероятность попадания на элементарный участок двух и более событий пренебрежимо мала по сравнению с вероятностью попадания одного события.

В действительности стационарность существует на ограниченном участке времени, а распространение на бесконечность лишь удобный прием. Например, вызовы телефона в пределах часа можно считать стационарными, но час днём и на час ночью – большая разница.

Отсутствие последствия: поток (грузовых) поездов – если они не могут следовать один за другим чаще чем через τ_0, то между событиями в потоке имеется зависимость и отсутствие последствия нарушается. Если интервал $\tau >> \tau_0$, то нарушение несущественно.

Ординарность: события в потоке приходят поодиночке, а не парами, тройками и т. д. Поток в парикмахерскую ординарен, а в загс, куда приходят парами не соответствует, однако если пару считать новой единицей, то можно рассматривать ординарным потоком «пар».

Если единицей является пакет, то наряду с пакетом приходится рассматривать случайную величину X – число событий в пакете, т.е. модель потока усложняется.

Поток, обладающий всеми тремя свойствами, называется простейшим или пуассоновским стационарным потоком, его математическое описание оказывается наиболее простым.

Простейший поток играет особую роль: при суперпозиции большого числа потоков с последствием (они должны быть стационарными и ординарными) образуется простейший поток (требуется, чтобы они были сравнимы по интенсивности).

В нестационарном пуассоновском потоке среднее число событий в единицу времени зависит от времени λ=λ(t), а для простейшего λ=const. Число событий потока, попадающих на любой участок, распределено по закону Пуассона.

Вероятность попадания равно m событий на участок длиной

$$P_m = \frac{(\lambda\tau)^m}{m!} \cdot e^{-\lambda\tau}$$

Для нестационарного потока $\lambda\tau = \int\limits_{t_0}^{t_0+\tau} \lambda(t)dt$. λ называется интенсивностью потока.

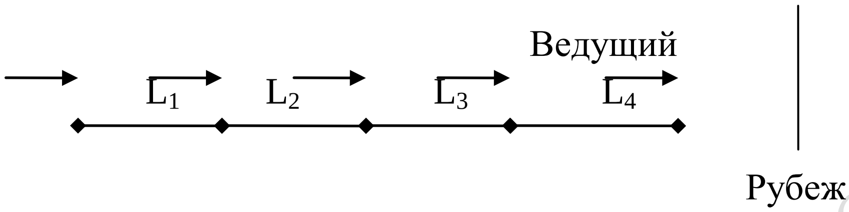
Если обозначить через T интервал времени (T<t) между двумя событиями, то её плотность распределения $f(t) = \lambda e^{-\lambda t}$, t > 0, $m_t = \frac{1}{\lambda}$, $D_t = \frac{1}{\lambda^2}$, $\delta_t = \frac{1}{\lambda}$. Вероятность появления события на Δt есть $\lambda \cdot \Delta t = P_1(\Delta t)$, т.к. считаем поток ординарным . $P_1(\Delta t) = \lambda(t) \cdot \Delta t$ для нестационарного потока.

Потоки Пальма и Эрланга

Если промежутки времени между последовательными событиями представляют собой независимые, одинаково распределённые случайные величины T_1, T_2, … , T_i, то такой поток событий называется *потоком Пальма***.**

Простейший поток есть частный случай потока Пальма, когда величины T_i распределены по одному и тому же показательному закону.

Пример. Группа самолётов идёт в боевом порядке «колонна» с одинаковой для всех скоростью V. Каждый из них должен выдерживать строй, т.е. держаться на определённом расстоянии от впереди идущего. Моменты пересечения этого рубежа, определяемого дальномером с ошибками, образуют поток Пальма, так как $T_1 = \frac{L_1}{V}, T_2 = \frac{L_2}{V},...$ независимы. Однако, если расстояние измерять не от соседа, а от ведущего колонну, то поток не будет потоком Пальма.



Важными для практики <u>образцами</u> потоков Пальма являются потоки Эрланга. Эти потоки образуются в результате «просеивания» простейших потоков.

Например, если из простейшего взять только каждое второе событие, то новый поток образует поток Эрланга второго порядка, k-ю точку – поток k-го порядка. Простейший поток представляет собой поток Эрланга 1-го порядка: $Э_1$.

Интервал времени $T = T_1 + ... + T_k = \sum_{i=1}^{k} T_i$ представляет собой сумму k независимых случайных величин – расстояний между событиями в простейшем потоке с распределением:

$$f_k(t) = \frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t}, t > 0,$$

$$m_t^{(k)} = \frac{k}{\lambda}, D_t^{(k)} = \frac{k}{\lambda^2}, \sigma_t^{(k)} = \frac{\sqrt{k}}{\lambda},$$

где λ – интенсивность простейшего потока, который подвергся прореживанию.

Если закон Эрланга записать через свои собственные характеристики, определив $\Lambda_к$ -- интенсивность потока $Э_к$, то $\Lambda_к = \lambda / k$, т.к. берется k-я часть из λ.

Выражая $\lambda = к\Lambda_к$ и подставив в $f_k(t)$, получим

$$f_k(t) = \frac{(k\Lambda_k)^k}{(k-1)!} t^{k-1} e^{-k\Lambda_k}, \; t > 0;$$

$$m_t^{(k)} = \frac{1}{\Lambda_k},\ D_t^{(k)} = \frac{1}{k\Lambda_k^2},\ \sigma_t^{(k)} = \frac{1}{\sqrt{k}\Lambda_k}.$$

Из формул видно, что при $\Lambda_k = const = \Lambda = \dfrac{1}{T}$ и $k \to \infty$ $\sigma_t^{(k)} \to 0$ поток Эрланга приближается к регулярному потоку с интервалом между событиями *T*, т.е. имеет жёсткую функциональную связь – последействие максимально, при $k = 1$ -- отсутствие последействия.

Таким образом, порядок потока k может служить в какой-то степени «мерой последействия». Реальный поток можно заменить потоком Эрланга, рассчитав порядок *k* из величин дисперсии.

Пример. Пусть получен поток с характеристиками $m_t = 2\min$, $\sigma_t = 0.9\min$. Подобрать поток $Э_к$.

$$\lambda = \frac{1}{m_t} = \frac{1}{2} = 0.5 \text{ события/min.}$$

$$k = \left(\frac{1}{\sigma_t^{(k)}\lambda_k}\right)^2 = \left(\frac{1}{0.9\cdot 1/2}\right)^2 \approx 4.9 \Rightarrow 5$$

Требуемый поток $$t_5(t) = \frac{(5\cdot 0.5)^5}{4!} t^4 e^{-t*5/2} \approx 4.1 t^4 e^{-2.5t}$$

При помощи потоков Эрланга можно сводить немарковские процессы к марковским.

Циклические процессы

Ранее было установлено, что предельные вероятности для стационарного процесса существуют, если есть переход из любого состояния в любое, и следовательно, для любого события k в цепи событий найдется подходящая последовательность перехода из $S^{(k)}_j$ в $S^{(n+k)}_j$. Расчет такой системы был приведен ранее и в общем случае не может быть записано его решение.

Однако существуют схемы систем, в которых расчет может быть выполнен и для общего случая.

Начнем с простого примера. Техническое устройство состоит из трёх одинаковых узлов, каждый может выходить из строя и после восстановления включается в систему. Перенумеруем состояния системы по числу неисправных узлов:

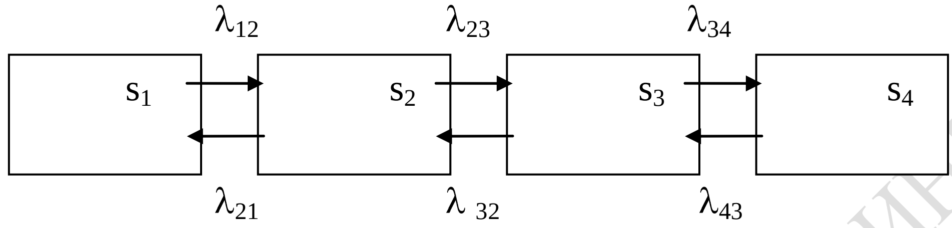
s_1 – все узлы исправны,

s_2 – один узел неисправен,

s_3 – два узла неисправны,

s_4 – все три узла неисправны – восстанавливаются

Граф состояний системы будет следующим:



Такая схема называется «схемой гибели и размножения»

Решение её ищется последовательной подстановкой.

Используя мнемоническое правило «что втекает, то и вытекает», получаем:

$p_1\lambda_{12} = p_2\lambda_{21}$

$p_1\lambda_{12}+p_3\lambda_{32} = p_2(\lambda_{21}+\lambda_{23}) \Rightarrow p_3\lambda_{32} = p_2\lambda_{23}$

$p_2\lambda_{23}+p_4\lambda_{43} = p_3(\lambda_{32}+\lambda_{34}) \Rightarrow p_4\lambda_{43} = p_3\lambda_{34}$

$p_3\lambda_{34} = p_4\lambda_{43}$

Отбросив последнее уравнение и добавив $p_1+p_2+p_3+p_4 = 1$, разрешаем уравнения через p_1:

$p_2=p_1\lambda_{12}/\lambda_{12}$

$p_3=p_2\lambda_{23}/\lambda_{32}=p_1(\lambda_{12}/\lambda_{21})(\lambda_{23}/\lambda_{32})$

$p_4=p_1(\lambda_{12}/\lambda_{21})(\lambda_{23}/\lambda_{32})(\lambda_{34}/\lambda_{43})$

Подставив в условие нормировки найденные вероятности, найдем Р1 и, следовательно, другие вероятности.

Очевидно, что схема может быть продолжена вправо до любого числа состояний n, выражение для

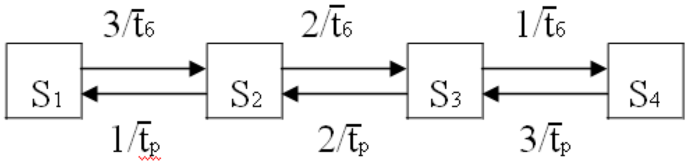
$$P_n = \frac{\lambda_{n-1,n}\ldots\lambda_{12}}{\lambda_{n,n-1}\ldots\lambda_{21}}P_1 = \frac{П(стрелки\_вправо)}{П(стрелки\_влево)}P_1\ ,$$

$$P_1 = \frac{1}{1+\dfrac{\lambda_{12}}{\lambda_{21}}+\dfrac{\lambda_{23}\lambda_{12}}{\lambda_{32}\lambda_{21}}+\ldots+\dfrac{\lambda_{n-1,n}\ldots\lambda_{12}}{\lambda_{n,n-1}\ldots\lambda_{21}}}.$$

Такая непрерывная марковская цепь называется «процессом гибели и размножения».

Пример. Прибор состоит из трех узлов, время безотказной работы узла $\bar{t}_б$, время восстановления $\bar{t}_р$ (ремонт), потоки простейшие. Найти среднюю производительность прибора, если при трех работающих узлах она равна

100%, при двух 50%, а при одном и менее не работает. Обозначая состояние системы, как и ранее, через S_i получим



Из общего решения (см. предыдущий пример) имеем:

$\lambda12=3/\bar{t}_б$, $\lambda23=2/\bar{t}_б$, $\lambda34=1/\bar{t}_б$

$\lambda21=1/\bar{t}_р$, $\lambda32=2/\bar{t}_р$, $\lambda43=3/\bar{t}_р$

и общее решение определяется формулой для P_n и P_1.

Обозначим ($\bar{t}_р/\bar{t}_б$)=γ, тогда

$$P_1=\frac{1}{1+3\gamma+3\gamma^2+\gamma^3}$$

$P_2=3\gamma P_1$, $P_3=3\gamma^2P_1$, $P_4=\gamma^3P_1$, П=($100P_1+50P_2$)%

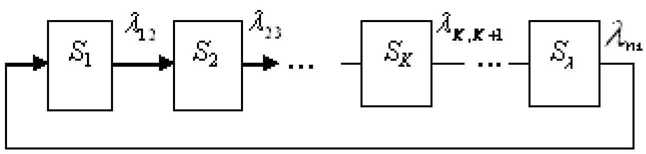
Пусть $\bar{t}_б$=10ч , $\bar{t}_р$=5ч, тогда γ=1/2 и

$$P_1=\frac{1}{1+\frac{3}{2}+\frac{3}{4}+\frac{1}{8}}=\frac{8}{27}\ ,\ \text{P2=3γ, P1=}\frac{12}{27},$$

$$P_3=3\gamma^2P_1=\frac{6}{27},\quad P_4=\gamma^3P_1=\frac{1}{27},$$

П=$\frac{14}{27}100=51{,}9\%$ номинала.

Марковский случайный процесс <u>называется циклическим</u>, если состояния связаны между собой в кольцо с односторонними переходами.



Для предельных вероятностей имеем:

$$\left.\begin{array}{c} P_2\lambda_{23}=P_1\lambda_{12} \\ \dots\dots\dots\dots\dots\dots \\ P_K\lambda_{K,k+1}=P_{k-1}\lambda_{k-1,k} \\ \dots\dots\dots\dots\dots\dots \\ P_n\lambda_{n-1}=P_{n-1}\lambda_{n-1,n} \\ P_1\lambda_{12}=P_n\lambda_{n,1} \end{array}\right\}$$

Имеем n уравнений и условие нормировки, поэтому можно одно из исходных однородных уравнений отбросить, пусть последнее. Разрешаем, как обычно через P_1:

$$P_2 = \frac{\lambda_{12}}{\lambda_{23}} P_1$$

$$P_3 = \frac{\lambda_{23}}{\lambda_{34}} P_2 = \frac{\lambda_{23}}{\lambda_{34}} \frac{\lambda_{12}}{\lambda_{23}} P_1 = \frac{\lambda_{12}}{\lambda_{34}} P_1$$

....

$$P_K = \frac{\lambda_{12}}{\lambda_{k,k+1}} P_1$$

$$P_n = \frac{\lambda_{12}}{\lambda_{n,1}} P_1$$

Подставим в условие нормировки:

$$1 + \lambda_{12} \left(\frac{1}{\lambda_{23}} + \frac{1}{\lambda_{34}} + ... + \frac{1}{\lambda_{n1}} \right) = \frac{1}{P_1}.$$

Для простейшего потока $\bar{t}_i = \frac{1}{\lambda_{i,i+1}}$ и поэтому моно заменить решение в другой форме через средние времена пребывания системы в каждом из состояний.

$$p_1 = \frac{\bar{t}_1}{\sum\limits_i \bar{t}_i}; \ldots \; p_k = \frac{\bar{t}_k}{\sum\limits_i \bar{t}_i},$$

т.е. предельные вероятности состояний в циклической схеме относятся как среднее времени пребывания системы в каждом из состояний.

Пример. Электронная ЦВМ может находиться в одном из состояний:

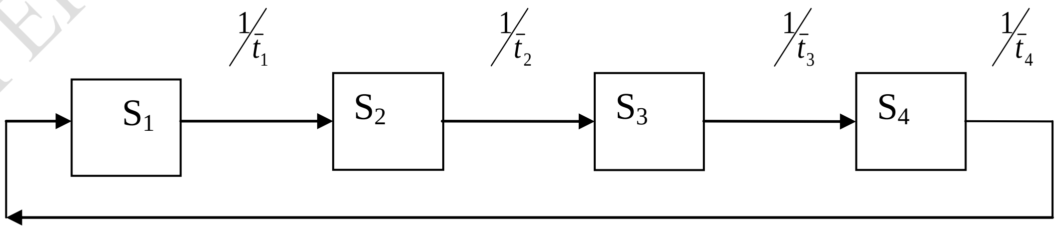
S1 – исправна,

S2 – поиск неисправности,

S3 – ремонт,

S4 – запуск.

Среднее время безотказной работы 0,5 суток, поиск неисправности 0,5 часа, ремонт – 6 часов, запуск –1 час



$$\bar{t}_1 = \frac{1}{2} \qquad \bar{t}_2 = \frac{1}{48} \qquad \bar{t}_3 = \frac{1}{4} \qquad \bar{t}_4 = \frac{1}{24} \quad (\textit{суток})$$

$$P_1 = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{48}+\frac{1}{4}+\frac{1}{24}} = \frac{1}{2}*\frac{48}{39} = \frac{24}{39}$$

$$P_2 = \frac{1}{48}*\frac{48}{39} = \frac{1}{39} \quad P_3 = \frac{1}{4}*\frac{48}{39} = \frac{12}{39} \quad P_4 = \frac{2}{39} \quad \sum P_i = 1$$

Сведение немарковских процессов к марковским методом псевдосостояний.

Реальные процессы всегда обладают последействием, кроме того и закон распределения отличен от показательного. Решение для немарковских процессов обычно приводит к уравнениям в частных производных и требует гораздо более сложного математического аппарата.

В некоторых случаях возможно приближённое приведение немарковских процессов к марковским. Обычно для этого требуется, чтобы число состояний системы было не велико, а отличающиеся от простейших потоки событий представляли собой (хотя бы приближённо) потоки Эрланга.

Тогда вводя в схему возможных состояний системы некоторые фиктивные состояния – псевдосостояния, удаётся свести немарковский процесс к марковскому и, следовательно, получить для предельных вероятностей алгебраические уравнения.

Обычный способ – замена состояния с выходным потоком Эрланга k-го порядка k-состояниями (последовательными) с простейшим потоком на входе, так как их последовательное соединение даёт разрежение потока в k раз и, следовательно, поток Эрланга k-го порядка.

Поясним идею метода псевдосостояний на примерах.

Пример1. Дана система S, которая может выходить из строя под влиянием простейшего потока интенсивности λ. Отказ восстанавливается, причём время восстановления подчиненно закону Эрланга 3-ого порядка:

$$f_3(t)= \mu(\mu t)^2 e^{-\mu t} /2 \qquad ,(t>0)$$

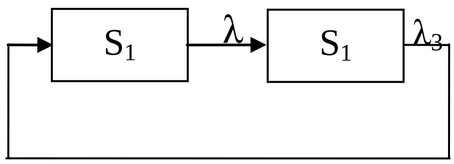
Решение: Время восстановления представляет собой сумму трёх случайных величин T_1, T_2, T_3 с распределением

$f_1(t)= \mu e^{-\mu t}$, (t>0)

Исходных состояний системы всего два:

S_1 – устройство исправно,

S_2 – устройство восстанавливается.



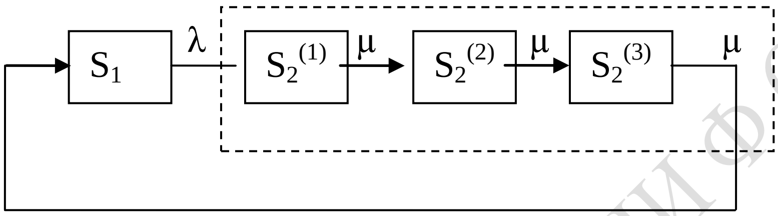
Введем вместо S_2 три последовательных состояния:

$S_2^{(1)}$ – ремонт начинается,

$S_2^{(2)}$ – ремонт продолжается,

$S_2^{(3)}$ – ремонт заканчивается.

Время пребывания в каждой из фаз S_2 будем считать распределенным по показательному закону $f_1(t)$. Процесс в такой системе



будет марковским и, следовательно, применимы формулы для циклического процесса.

Обозначим $\bar{t}_1 = 1/\lambda$, $\bar{t}_2 = 1/\mu$ из $P_k = \dfrac{\bar{t}_k}{\sum_i \bar{t}_i}$, получим:

$$P_1 = \frac{\bar{t}_1}{\bar{t}_1 + \bar{t}_2 + \bar{t}_2 + \bar{t}_2} = \frac{\bar{t}_1}{\bar{t}_1 + 3\bar{t}_2}$$

$$P_2^{(1)} = P_2^{(2)} = P_2^{(3)} = \frac{\bar{t}_2}{\bar{t}_1 + 3\bar{t}_2}, \qquad \sum_i P_i = 1$$

$$P_2 = P_2^{(1)} + P_2^{(2)} + P_2^{(3)} = \frac{3\bar{t}_2}{\bar{t}_1 + 3\bar{t}_2}$$

Произведя обратную замену через исходные данные получим:

$P_1 = \mu/(\mu+3\lambda)$, $P_2 = 3\lambda/(\mu+3\lambda)$.

Пример 2. Устройство состоит из двух одинаковых узлов, каждый из которых может выходить из строя под влиянием простейшего потока λ. Отказавший узел ремонтируется. Время ремонта T распределено по закону Эрланга второго порядка $f_2(t) = \mu(\mu t)e^{-\mu t}$, t>0. Требуется найти предельные вероятности.

Решение: Истинных состояний системы 3:

S_0 – оба узла работают,

S_1 – один работает, второй на ремонте,

S_2 – оба узла ремонтируются.

Поскольку порядок закона Эрланга два, представим ремонт в виде двух фаз:

1) ремонт начинается ,

2) ремонт заканчивается.

Длительность каждой фазы теперь находится по экспоненциальному закону:

$S_1^{(1)}$ - один работает, другой начинает ремонт;

$S_1^{(2)}$ - один работает, другой кончает ремонт;

$S_2^{(1,1)}$ - оба узла начинают ремонт;

$S_2^{(1,2)}$ - один начинает ремонт, а другой кончает;

$S_2^{(2,2)}$ - оба узла кончают ремонт.

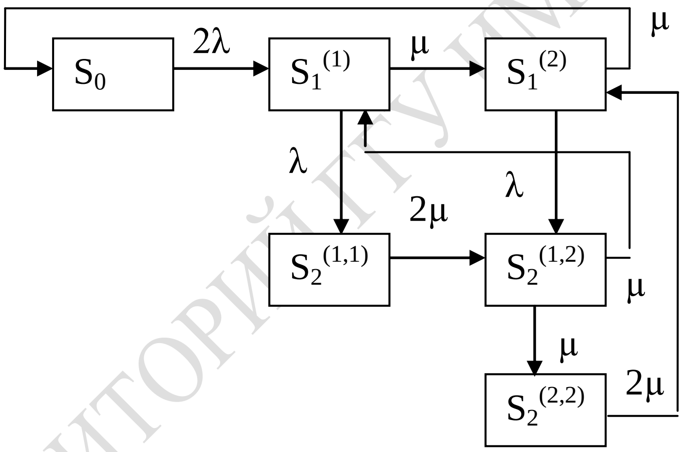
$2\lambda p_0+\mu p_2^{(1,2)}=p_1^{(1)}(\lambda+\mu)2$

$\mu p_1^{(1)}+2\mu p_2^{(2,2)}=p_1^{(2)}(\lambda+\mu)$

$\lambda p_1^{(1)}=p_2^{(1,1)}2\mu$

$2\mu p_2^{(1,1)}+\lambda p_1^{(2)}=p_2^{(1,2)}(\lambda+\mu)$

$\mu p_2^{(1,2)}=p_2^{(2,2)}2\mu$



$\mu p_1^{(2)}=p_0 2\lambda$

Решив уравнения через $\sum p_i=1$, находим вероятность псевдосостояний, а затем и самих состояний. $p_0=\dfrac{\mu^2}{\mu^2+4\lambda\mu+4\lambda^2}$, $p_1=p_1^{(1)}+p_1^{(2)}=\dfrac{4\lambda\mu}{\mu^2+4\lambda\mu+4\lambda^2}$

$$p_2=p_2^{(1,1)}+p_2^{(1,2)}+p_2^{(2,2)}=\frac{4\lambda^2}{\mu^2+4\lambda\mu+4\lambda^2}.$$

Заметим, что метод псевдосостояний достаточно громоздок и требует повышенного внимания.

Системы массового обслуживания.

Система массового обслуживания прежде всего характеризуется необходимостью пребывать в состоянии ожидания:

- Очередь покупателей возле касс,
- Колонна автомобилей у светофора ,
- Очередь больных на приём к врачу,
- Группа самолётов ожидающих разрешения на взлёт,
- Очередь на ремонт в цехе промышленного предприятия,
- Груда рукописных материалов для печатиу машинистки,
- Набор программ для выполнения на ЦВМ.

Встречающийся на каждом шагу «феномен» ожидания является прямым следствием вероятностного характера возникновения потребности в обслуживании и разброса показателей обслуживающих систем.

Ни время возникновения потребности в обслуживании, ни продолжительность обслуживания заранее неизвестны, иначе можно было бы исключить неприятную необходимость «ждать».

Цель изучения обслуживающей системы в том, чтобы взять под контроль некоторые количественные характеристики системы массового обслуживания: среднее время пребывания клиента в очереди, время простоя системы и т.д. Эти характеристики обычно противоречивы, так как рассматриваются две противоположных точки зрения: потребителя и системы.

Основными элементами системы массового обслуживания являются:

- заявка на обслуживание (требования),
- механизм обслуживания (обслуживающая система, узел обслуживания, обслуживающий прибор).

Взаимодействие между заявкой и механизмом представляет интерес лишь с точки зрения регистрации продолжительности интервала времени, который требуется для полного обслуживания заявки с момента её поступления в обслуживающую систему.

В теории массового обслуживания оперируют понятиями: распределение моментов поступления требований и распределение времени обслуживания требований.

Эти распределения вероятностей могут моделировать ситуации, когда требования поступают и обслуживаются индивидуально (одноканальная конфигурация системы) и группами (одновременное поступление и обслуживание нескольких требований) (многоканальная конфигурация системы или системы с параллельно-групповым обслуживанием).

Весьма существенны и другие характеристики:

дисциплина очереди и обслуживания.

Наиболее справедливо правило «первым пришёл – первым обслуживаешься» (ПЕРППО).

Однако на практике могут быть и другие дисциплины:

- «последним пришёл – первым обслуживаешься (ПОППО),
- случайный отбор заявок (СОЗ),
- очереди по уровню приоритета.

Структура обслуживающей системы также оказывает влияние на процесс обслуживания.

– несколько параллельных приборов (узлов) обслуживания;

– несколько последовательных разнотипных узлов обслуживания (тандем очередей);

– сетевая структура – имеется и параллельное и последовательное обслуживание.

Следующая характеристика – допустимая вместимость блока ожидания или длина очереди. В ситуации превышения длины очереди происходит отказ требования от обслуживания.

Факторы, порождаемые природой источника требований:

– источник конечной ёмкости (мощности), когда поступление очередного требования снижает частоту последующих требований. (Например, в цехе из М станков суммарное количество потенциальных заявок на ремонт равно М Другой пример: машинописное бюро обслуживает конечное число клиентов, но каждый из них может генерировать бесконечное число заявок, т. к. результат выхода не влияет на вход).

– источник бесконечной мощности не меняет своих характеристик.

Модели массового обслуживания, в которых в качестве пользователей или обслуживающих элементов выступают люди иногда называются бихевиоральными, учитывающими поведение человеческих индивидуумов. Например, при параллельном обслуживании может быть переход из одной очереди к другой с надеждой сократить время ожидания. В других случаях клиент просто отказывается от присоединения к очереди ожидающих, т. к. не переносит длительного бездействия (оно у разных людей может быть разным). Модель массового обслуживания не может учитывать сугубо индивидуальное поведение клиентов, а включает как редкое исключение. С другой стороны, если большинство клиентов в процессе ожидания обслуживания проявляют склонность к некоторому «нестандартному» поведению, например, излишней разговорчивости, то эта «привычка» клиентов влияет на операционную структуру системы массового обслуживания и должна быть учтена как фактор, позволяющий увеличить продолжительность обслуживания в расчете на одного клиента.

Таким образом, функциональные возможности модели массового обслуживания определяются следующими факторами:

1) распределением моментов поступления заявок на обслуживание (единичных или групповых);
2) распределением продолжительностей обслуживания;
3) конфигурацией обслуживающей системы (последовательное, параллельное, последовательно-параллельное обслуживание);
4) дисциплиной очереди (ПЕРППО, ПОСППО, СОЗ) и приоритетными характеристиками обслуживающей системы;
5) вместимостью блока ожидания (ограниченная или неограниченная);
6) ёмкостью (или мощностью) источника требований (конечная или бесконечно большая);
7) бихевиоральными характеристиками системы (возможность клиентов переходить из одной очереди в другую, ненулевая вероятность отказов от ожидания либо сразу, либо после непродолжительной задержки в очереди).

Можно построить столько моделей массового обслуживания, сколько существует различных комбинаций перечисленных операционных характеристик, ассоциированных с понятием «массовое обслуживание».

Входной поток

Входной поток наиболее просто описывается для Марковских процессов с учетом того, что поступившие требования непременно присоединяются к очереди и не покидают её, пока их не обслужат. Такие стохастические процессы называют процессами чистого рождения (Пример: роддом: данные о случайном процессе – записи точного времени рождения каждого ребенка в течение определенного периода, например суток).

Для бесконечно малого интервала h и n>0 имеем:

$$p_h(t+h) = p\begin{pmatrix} \text{n поступлен. в течение t единиц времени и ни одного на h} \\ \text{или n - 1 поступлен. в течение t единиц времени и одно на h} \end{pmatrix} =$$

$$= p_n(t)\,p_0(h) + p_{n-1}(t)\,p_1(h), n = 1,2,...$$

$$p_0(t+h) = p_0(t)\,p_0(h), n = 0.$$

Пусть среднее число поступлений в единицу времени есть λ, тогда:

$$p_0(h) \approx 1-\lambda h, \quad p_1(h) \approx \lambda h,$$

$$p_n(t+h) \approx p_n(t)(1-\lambda h) + p_{n-1}(t)\lambda h, \ n > 0$$

$$p_0(t+h) \approx p_0(t)(1-\lambda h), \ n = 0$$

или

$$[p_n(t+h) - p_n(t)]/h \approx -\lambda p_n(t) + \lambda p_{n-1}(t)$$

$$[p_0(t+h) - p_0(t)]/h \approx -\lambda p_0(t)$$

и при h→0

$$\left.\begin{aligned} p_n'(t) &= -\lambda p_n(t) + \lambda p_{n-1}(t) \\ p_0'(t) &= -\lambda p_0 t \end{aligned}\right\}$$ система разностно-дифференциальных уравнений.

Решение есть $p_n(t) = [(\lambda t)^n e^{-\lambda t}]/n!$ - распределение Пуассона, его среднее и дисперсия раны λt. Входной поток всегда является функцией времени, а λ – его интенсивность.

Выходной поток

Характеризует процесс на выходе системы массового обслуживания. Предполагается, что система понижает функционирование при наличии в ней N клиентов, которые после завершения их обслуживания выбывают из системы с интенсивностью μ, такие процессы называют процессами чистой гибели.

Например, изъятие из склада хранимых там запасов с интенсивностью μ единиц в единицу времени есть процесс чистой гибели.

Пусть $q_n(t)$ есть вероятности того, что в течении t единиц времени со склада будет выдано n единиц запасов. Для достаточно малого h имеем:

$$q_0(h) \approx 1 - \mu h$$

$$q_1(h) \approx \mu h$$

и для $q_n(t)$ имеем следующие соотношения:

$$q_N(t+h) \approx q_N(t)1 + q_{N-1}(t)\mu h,\ n = N$$

$$q_n(t+h) \approx q_n(t)(1-\mu h) + q_{n-1}(t)\mu h,\ 1 \le n < N$$

$$q_0(t+h) \approx q_0(t)(1-\mu h),\ n = 0$$

После преобразований получим систему дифференциальных уравнений:

$$q'_N(t) = \mu q_{N-1}(t),\ n = N$$

$$q'_n(t) = -\mu q_n(t) + \mu q_{n-1}(t),\ 1 \le n < N$$

$$q'_0(t) = -\mu q_0(t),\ n = 0,$$

которая имеет следующее решение:

$$q_n(t) = \frac{(\mu t)^n e^{-\mu t}}{n!},\ 1 \le n < N$$

$$q_N(t) = 1 - \sum_{n=0}^{N-1} q_n(t),\ n = N$$

Для вероятности того, что в системе остались n заявок на обслуживание имеем:

$P_n(t) = q_{N-n}(t)$ или:

$$P_n(t) = \frac{(\mu t)^{N-n} e^{-\mu t}}{(N-n)!},\ 1 \le n < N$$

$$P_0(t) = 1 - \sum_{n=1}^{N} P_n(t),\ n = N.$$

Как видим и здесь описание потоков Пуассоновское.

Оценка распределений (верификация)

Как видим, пуассоновское распределение играет важную роль при описании потоков в системе. Поэтому важно уметь оценивать истинный поток на принадлежность пуассоновскому. Грубая оценка проводится на основе сравнений среднего равного дисперсии, т.к. из всех известных дискретных распределений вероятностей таким свойством обладает лишь распределение Пуассона. Более точная оценка делается с использованием χ^2- критерия, основанный на анализе квадратов отклонений наблюдаемых значений от расчётных по выбранной формуле (распределения: пуассоновское, экспоненциальное, нормальное…)

Пусть на перекрёстке фиксируются моменты прибытия транспортных средств так что можем рассчитать среднее число прибытий за каждый час и для каждого числа прибытий в час определена соответствующая частота f_n.

$$P_n = [(11.65)^n e^{-11.65} \;] / n!$$

n	 n	e n	χ i	n	p_n
				0	0
				1	0.000 1
				2	0.000 6
				3	0.002 3
				4	0.006 7
				5	0.015 6
				6	0.030 3
			1	7	0.050 4
		11,3	,636	8	0.073 4
		5,9 9	0	9	0.095
0		6,9 7	0 ,557	10	0.110 6
1		7,3 8	0 ,356	11	0.117 2
		7,1	1	12	0.113

2	0	7	,117		8
3	1	6,4 3	3 ,248	13	0.102
4		5,3 4	1 ,325	14	0.084 8
5				15	0.065 9
6		12,	2	16	0.047 9
7		42	,365	17	0.083 4

$$\bar{n} = \frac{\sum n f_n}{\sum f_n} = \frac{734}{63} = 11.65 \text{ авто/час}$$

Ожидаемое улучшение частоты $e_n = (\sum f_n) p_n = 63 p_n$. Объединяем в группы так что бы e_n были примерно ровны, например можно потребовать что бы $f_n \geq 5$. $\chi^2 = \sum \frac{(f_n - e_n)^2}{e_n} = 10.6$.

Вычисленное χ^2=10.6 нужно сравнить с критическим для χ^2 – распределения.

Для этого задается уровень значимости α (аналог 1- $P_{доверительное}$) и число степеней свободы ν, которое в нашем случае :

ν = (число составных интервалов)-(число оцениваемых параметров)-1;

(число составных интервалов)=8, (число оцениваемых параметров)=1, ν=6.

Для ν=6 и α=0.05 из таблицы, χ_6^2 (0.05)=12.592. Выдвигаемая гипотеза принимается, если $\chi^2 \leq \chi_\nu^2(\alpha)$. Поскольку вычисленное χ^2=10.6 < χ_6^2 (0.05), то выборка соответствует пуассоновскому распределению со средним

$\bar{n}$=11.65 прибытий/час.

Классификация систем при наличии входного и выходного потоков использует принятую унификацию следующей структуры:

(a/b/c): (d/e/f),

где символы ассоциированы с конкретными наиболее существенными элементами модели:

a/b/c – обозначения Д.Г. Кендалла (1953 г.),
d/e – обозначения А.М. Ли (1966 г.),
f – обозначение Х. Таха (1982 г.);

a – распределение моментов поступления заявок на обслуживание,
b – распределение времени обслуживания,
c – число параллельно функционирующих узлов обслуживания (1,…, ∞),
d – дисциплина очереди (ПЕРППО, ПОСППО, СОЗ),

e – максимальное число допустимых в систему требований (число требований в очереди + число требований принятых на обслуживание),

f – емкость источника генерирующего заявки на обслуживание.

Для a/ƅ используют следующие обозначения:

M – пуассоновское (марковское),

D – фиксированный (детерминированный) интервал,

E_k – распределение Эрланга или Гамма-распределение,

GI – распределение произвольного вида моментов поступления,

G – распределение произвольного вида моментов выбытия.

Например (M/D/10): (GD/N/∞) –

пуассоновский входной поток,
фиксированное время обслуживания,
10 параллельных узлов обслуживания,
GD – не регламентированная дисциплина очереди,
система вмещает не более N требований,
ёмкость источника не ограничена.