Теория массового обслуживания

1. Основные понятия теории массового обслуживания
2. Входные характеристики системы массового обслуживания
3. Уравнения Колмогорова

Дополнительная литература:

Просветов, Г.И. Математические методы в логистике: задачи и решения/ Г.И.Просветов. Учебно-практическое пособие. – М.:Изд-во «Альфа-Пресс», 2008. – 304 с.

При исследовании операций часто приходится сталкиваться с системами, предназначенными для многоразового использования при решении однотипных задач.

Возникающие при этом процессы получили название процессов обслуживания, а системы — систем массового обслуживания (**СМО**).

Каждая СМО состоит из определенного числа обслуживающих единиц (приборов, устройств, пунктов, станций), которые называют **каналами обслуживания**.

Каналами могут быть линии связи, рабочие точки, вычислительные машины и др.

По числу каналов СМО подразделяют на одноканальные и многоканальные.

Заявки поступают в СМО обычно не регулярно, а случайно, образуя так называемый случайный поток заявок (требований).

Обслуживание заявок также продолжается какое-то случайное время.

Это приводит к тому, что СМО оказывается загруженной неравномерно: в какие-то периоды времени скапливается очень большое количество заявок (они либо становятся в очередь, либо покидают СМО необслуженными), в другие же периоды СМО работает с недогрузкой или простаивает.

Предметом теории массового обслуживания

является построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, характер потока заявок и т.п.) с показателями эффективности СМО, описывающими ее способность справляться с потоком заявок.

В качестве показателей эффективности СМО используются:

- среднее число заявок, обслуживаемых в единицу времени;
- среднее число заявок в очереди;
- среднее время ожидания обслуживания;
- вероятность отказа в обслуживании без ожидания;
- вероятность того, что число заявок в очереди превысит определенное значение
- и т.п.

СМО делят на два основных типа (класса):

- СМО с отказами
- СМО с ожиданием (очередью).

СМО с ожиданием подразделяются на разные виды в зависимости от того, как организована очередь:

- с ограниченной или неограниченной длиной очереди,
- с ограниченным временем ожидания
- т.п.

дисциплина обслуживания

определяет порядок выбора заявок из числа поступивших и порядок распределения их между свободными каналами.

По этому признаку обслуживание заявки может быть организовано по принципу *"первая пришла — первая обслужена"*, *"последняя пришла — первая обслужена"* (например, при извлечении для обслуживания изделий со склада, ибо последние из них оказываются часто более доступными) или *обслуживание с приоритетом* (когда в первую очередь обслуживаются наиболее важные заявки).

Приоритет может быть как абсолютным, когда более важная заявка "вытесняет" из-под обслуживания обычную заявку (например, в случае аварийной ситуации плановые работы ремонтных бригад прерываются до ликвидации аварии), так и относительным, когда более важная заявка получает лишь "лучшее" место в очереди.

ВОПРОС 2.

Входные характеристики системы массового обслуживания

Процесс работы СМО представляет собой *случайный процесс*.

Под **случайным (вероятностным или стохастическим)** процессом понимается процесс изменения во времени состояния какой-либо системы в соответствии с вероятностными закономерностями.

Процесс называется процессом с **дискретными состояниями**, если его возможные состояния можно заранее перечислить, а переход системы из состояния в состояние происходит мгновенно (скачком).

Процесс называется процессом с **непрерывным временем**, если моменты возможных переходов системы из состояния в состояние не фиксированы заранее, а случайны.

Процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем. Это означает, что состояние СМО меняется скачком в случайные моменты появления каких-то событий (например, прихода новой заявки, окончания обслуживания и т.п.).

Математический анализ работы СМО существенно упрощается, если процесс этой работы — марковский.

Случайный процесс называется марковским или случайным процессом без последствия, если для любого момента времени t_0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент t_0 и не зависят от того, когда и как система пришла в это состояние.

Пример марковского процесса:

система S — счетчик в такси.

Состояние системы в момент t характеризуется числом километров (десятых долей километров), пройденных автомобилем до данного момента. Пусть в момент t_0 счетчик показывает S_0. Вероятность того, что в момент $t>t_0$ счетчик покажет то или иное число километров $S1$, зависит от S_0, но не зависит от того, в какие моменты времени изменялись показания счетчика до момента t_0 .

Многие процессы можно приближенно считать марковскими.

В ряде случаев предысторией рассматриваемых процессов можно просто пренебречь и применять для их изучения марковские модели.

При анализе случайных процессов с дискретными состояниями удобно пользоваться геометрической схемой — так называемым **графом состояний**.

Обычно состояния системы изображаются *прямоугольниками* (*кружками*), а возможные переходы из состояния в состояние — *стрелками* (*ориентированными дугами*), соединяющими состояния.

Пример 1.

Построить граф состояний следующего случайного процесса:

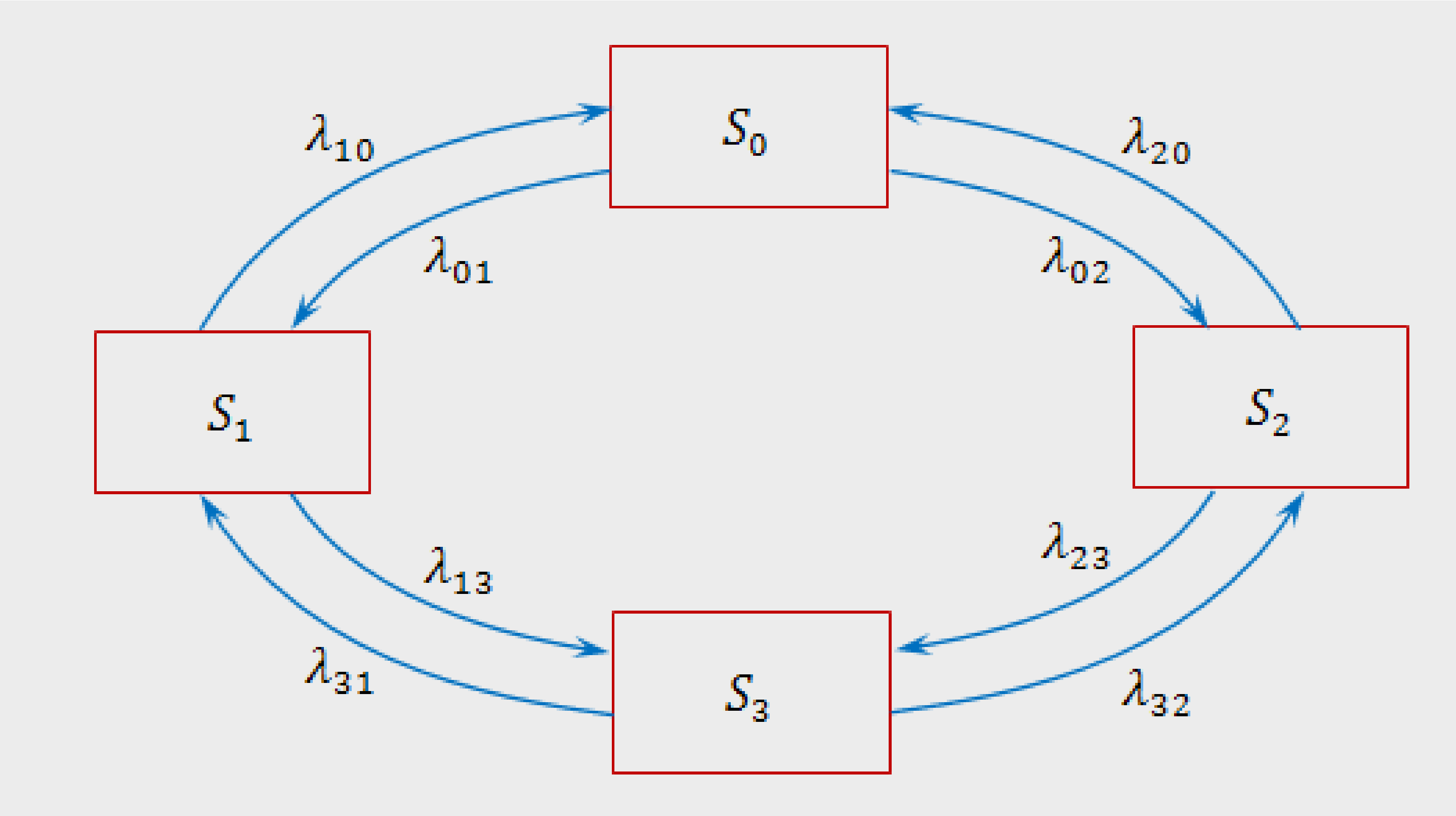
устройство S состоит из двух узлов, каждый из которых в случайный момент времени может выйти из строя, после чего мгновенно начинается ремонт узла, продолжающийся заранее неизвестное случайное время.

Решение.

Возможные состояния системы:

- S_0 — оба узла исправны;
- S_1 — первый узел ремонтируется, второй исправен;
- S_2 — второй узел ремонтируется, первый исправен;
- S_3 — оба узла ремонтируются.

Граф системы



Стрелка, направленная, например, из S_0 в S_1, означает переход системы в момент отказа первого узла, из S_1 в S_0 — переход в момент окончания ремонта этого узла.

На графе отсутствуют стрелки из S_0 в S_3 и из S_1 в S_2. Это объясняется тем, что выходы узлов из строя предполагаются независимыми друг от друга и, например, вероятностью одновременного выхода из строя двух узлов (переход из S_0 в S_3) или одновременного окончания ремонтов двух узлов (переход из S_3 в S_0) можно пренебречь.

Под потоком событий понимается последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени (например, поток вызовов на телефонной станции, поток отказов ЭВМ, поток покупателей и т.п.).

Поток характеризуется интенсивностью λ — частотой появления событий или средним числом событий, поступающих в СМО в единицу времени

Поток событий называется регулярным, если события следуют одно за другим через определенные равные промежутки времени.

Например, поток изделий на конвейере сборочного цеха (с постоянной скоростью движения) является регулярным.

Поток событий называется **стационарным**, если его вероятностные характеристики не зависят от времени.

В частности, интенсивность стационарного потока есть величина постоянная: $\lambda(t)=\lambda$.

Например, поток автомобилей на городском проспекте не является стационарным в течение суток, но этот поток можно считать стационарным в течение суток в часы пик.

В последнем случае фактическое число проходящих автомобилей в единицу времени (например, в каждую минуту) может заметно отличаться друг от друга, но среднее их число будет постоянно и не будет зависеть от времени.

Вопрос 3.

Уравнения Колмогорова.

Пример 1.

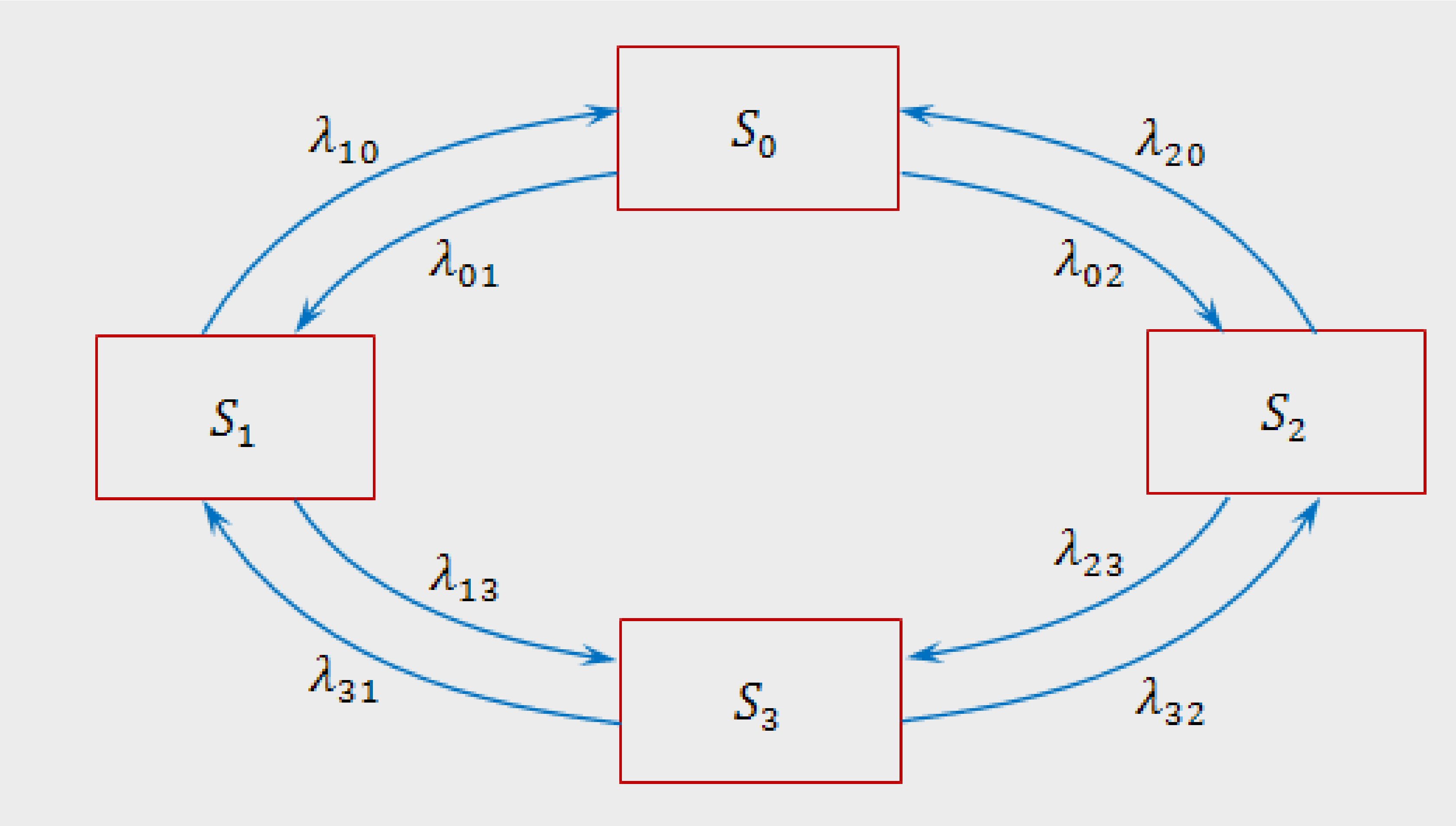
Построить граф состояний следующего случайного процесса:

устройство S состоит из двух узлов, каждый из которых в случайный момент времени может выйти из строя, после чего мгновенно начинается ремонт узла, продолжающийся заранее неизвестное случайное время.

Возможные состояния системы:

- S_0 — оба узла исправны;
- S_1 — первый узел ремонтируется, второй исправен;
- S_2 — второй узел ремонтируется, первый исправен;
- S_3 — оба узла ремонтируются.

Граф системы



Вероятностью i-го состояния называется вероятность $p_i(t)$ того, что в момент Т система будет находиться в состоянии S_i. Очевидно, что для любого момента t сумма вероятностей всех состояний равна единице:

$$\sum_{i=0}^{3} p_i(t) =$$

$$=p_0(t)+ p_1(t)+ p_2(t)+ p_3(t)=1$$

Правило составления уравнений Колмогорова.

В левой части каждого из них стоит производная вероятности i-го состояния. В правой части — сумма произведений вероятностей всех состояний (из которых идут стрелки в данное состояние) на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность данного (i-го состояния).

система дифференциальных уравнений Колмогорова для вероятностей состояний:

$$\begin{cases} p_0' = \lambda_{10}p_1 + \lambda_{20}p_2 - (\lambda_{01} + \lambda_{02})p_0, \\ p_1' = \lambda_{01}p_0 + \lambda_{31}p_3 - (\lambda_{10} + \lambda_{13})p_1, \\ p_2' = \lambda_{02}p_0 + \lambda_{32}p_3 - (\lambda_{20} + \lambda_{23})p_2, \\ p_3' = \lambda_{13}p_1 + \lambda_{23}p_2 - (\lambda_{31} + \lambda_{32})p_3. \end{cases}$$

Уравнения Колмогорова дают возможность найти все вероятности состояний как функции времени.

Особый интерес представляют вероятности системы $p_i(t)$ в предельном стационарном режиме, т.е. при $t \to \infty$, которые называются предельными (или финальными) вероятностями состояний.

В теории случайных процессов доказывается, что если число состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое состояние, то предельные вероятности существуют.

Предельная вероятность состояния Si имеет четкий смысл: она показывает среднее относительное время пребывания системы в этом состоянии. Например, если предельная вероятность состояния So , т.е. p_0=0,5 , то это означает, что в среднем половину времени система находится в состоянии So.

Так как предельные вероятности постоянны, то, заменяя в уравнениях Колмогорова их производные нулевыми значениями, получим систему линейных алгебраических уравнений, описывающих стационарный режим.

Для системы с графом состояний S такая система уравнений имеет вид:

$$\begin{cases} (\lambda_{01} + \lambda_{02})p_0 = \lambda_{10}p_1 + \lambda_{20}p_2, \\ (\lambda_{10} + \lambda_{13})p_1 = \lambda_{01}p_0 + \lambda_{31}p_3, \\ (\lambda_{20} + \lambda_{23})p_2 = \lambda_{02}p_0 + \lambda_{32}p_3, \\ (\lambda_{31} + \lambda_{32})p_3 = \lambda_{13}p_1 + \lambda_{23}p_2. \end{cases}$$

Систему можно составить непосредственно по размеченному графу состояний, если руководствоваться правилом, согласно которому слева в уравнениях стоит предельная вероятность данного состояния p_i, умноженная на суммарную интенсивность всех потоков, ведущих из данного состояния, а справа — сумма произведений интенсивностей всех потоков, входящих в i-е состояние, на вероятности тех состояний, из которых эти потоки исходят.

Пример 2.

Найти предельные вероятности для системы S из примера 1, граф состояний которой приведен на рис. 1, при

$$\lambda_{01} = 1, \quad \lambda_{02} = 2, \quad \lambda_{10} = 2, \quad \lambda_{13} = 2, \quad \lambda_{20} = 3, \quad \lambda_{23} = 1, \quad \lambda_{31} = 3, \quad \lambda_{32} = 2.$$

Система алгебраических уравнений, описывающих стационарный режим для данной системы, имеет вид:

$$\begin{cases} 3p_0 = 2p_1 + 3p_2, \\ 4p_1 = p_0 + 3p_3, \\ 4p_2 = 2p_0 + 2p_3, \\ p_0 + p_1 + p_2 + p_3 = 1. \end{cases}$$

Решив систему получим:

$P_0=0,4$

$P_1=0,2$

$P_2=0,27$

$P_3=0,13$

Т.е. в предельном, стационарном режиме система S в среднем

- 40% времени будет находиться в состоянии P_0 (оба узла исправны),
- 20% — в состоянии P_1 (первый узел ремонтируется, второй работает),
- 27% — в состоянии P_2 (второй узел ремонтируется, первый работает)
- 13% времени — в состоянии P_3 (оба узла ремонтируются).