СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В настоящее время большое внимание уделяется вопросам организации и управления различных отраслей народного хозяйства. Быстрое развитие новых видов техники, увеличения масштабов и стоимости проводимых мероприятий приводит к необходимости разработки научного анализа структуры и организации сложных процессов.

От науки требуются рекомендации по оптимальному управлению такими процессами.

Потребности практики вызвали необходимость разработки специальных методов, которые объединяются под общим названием “Исследование операций”, под которым понимается применение математических методов для обоснования оптимальных решений в различных областях человеческой деятельности. Это означает, что прошли те времена, когда различные управления находились методом проб и ошибок, так как такое управление приводит к большим потерям, связанные с ошибками.

В настоящее время исследование операций одна из быстро развивающихся отраслей науки, которая используется в промышленности, транспорте, торговле, сельском хозяйстве, финансовых организациях, военном деле и т.п.

Основные задачи, решаемые для гражданской авиации с использованием методов теории “Исследования операций”:

1. Разработка структуры управления воздушным движением (УВД) в масштабах страны, зоны и аэропорта.
2. Размещение радиоэлектронных средств радиолокации, навигации, посадки и связи.
3. Использование парка воздушных судов.
4. Разработка организационной структуры организации эксплуатации бортовых и наземных РЭС.

Глава 1. Основы теории исследования операций

1.1. Основные понятия теории исследования операций

Под операцией понимается любое мероприятие, подчиняющееся определенному замыслу и направленное на достижение конкретной цели [1, 3]. Операция всегда является управляемым мероприятием, т.е. выбор того или иного способа её организации зависит от исследователя. Всякий набор необходимых параметров называется решением. Решения, которые по тем или другим признакам предпочтительны перед другими, называются оптимальными. Основное назначение исследования операций состоит в предварительном обосновании оптимальных решений. Иногда в результате исследования удается указать одно единственное оптимальное решение. Чаще выделить область практически равноценных разумных решений, в пределах которых может быть сделан окончательный выбор. Процесс принятия реального решения выходит за рамки теории исследования операций и относится к компетенции ответственного лица (или группы лиц), которым предоставлено право окончательного выбора и на которого возложена ответственность за этот выбор.

Помимо основной задачи – обоснования оптимальных решений, к области исследования операций относятся следующие задачи:

- сравнительная оценка вариантов организации операций (решений);
- оценка влияния на результат операции параметров (заданных условий и определяемых параметров);
- исследование “узких мест”, т.е. таких элементов системы, нарушение действий которых существенно повлияет на успешность операции.

Исследование операций содержит в себе четыре основные проблемы:

- выбор показателей эффективности и критериев, по которым проводится оптимизация;
- построение модели операции;
- определение исходных данных, вводимых в модель;
- отыскание оптимального решения.

1.2. Показатели эффективности операций (критерии эффективности)

Пусть рассматривается некоторая операция. Естественным требованием к принимаемому решению является эффективность операции, понимая это как степень её готовности к выполнению своего предназначения. Чем лучше организована операция, тем она эффективнее. Чтобы судить об эффективности операции и сравнивать между собой по эффективности различно организованные операции, нужно иметь некоторый критерий оценки или показатель эффективности. Обозначим показатель эффективности через W. Важнейшими требованиями, которым должны отвечать показатели эффективности в исследовании операций, являются [2, 5]:

- представительность, то есть показатель должен позволять оценивать эффективность решения основной задачи операции;
- критичность, то есть чувствительным к изменениям исследуемых параметров;
- простота, так как введение в показатель второстепенных величин может усложнить исследования, не приводя ни к каким уточнениям выводов;
- объединение в себе по возможности всех основных элементов исследуемой операции, желательно, чтобы критерий был единым, так как при наличии двух критериев крайне затруднительно решение задачи;
- правильный учет стохастичности (случайности) процесса. Для данного случая показателя эффективности не детерминированное W, а её среднее значение (математическое ожидание) $M\ W$.

Приведем примеры удачного и неудачного выбора показателя. В начале второй мировой войны на торговых судах ставили зенитные орудия для защиты судов от авиации противника. В качестве показателя установки зениток использовался ущерб, наносимый самолетам противника. Статические данные показали, что только 4% самолетов, атаковавших эти корабли, были сбиты. Это показывает что орудия не окупали затрат на установку и расход боеприпасов. Было принято решение об их демонтаже. Безоружные торговые суда стали легкой добычей самолетов противника. Проведенный анализ по другому показателю – проценту потопленных судов из числа атакованных при отсутствии зенитного прикрытия составил 25%, а при наличии – 10%. Экономический анализ показал, что исходя из этих цифр, расходы на установку зенитных орудий с избытком окупают стоимость сохраненных судов.

В качестве другого примера неудачного решения является разработка самолета Ил-2 с двумя кабинами – бронированная кабина летчика с пулеметом для защиты передней полусферы и кабина стрелка-радиста с пулеметом для защиты задней полусферы. Используя показатель – ущерб истребителям противника и расход боеприпасов, было принято решение о демонтаже пулемета стрелка-радиста. Истребителями противника было сбито много самолетов Ил-2. После было принято решение по оснащению кабины стрелка-радиста пулеметами, и данный самолет очень хорошо воевал до конца войны.

Рассмотренные примеры показывают, что правильный выбор показателя эффективности следует признать необходимым условием полезности исследования, выполняемого для обоснования выбора оптимального решения.

Приведем несколько примеров, в каждом из которых показатель эффективности выбран в соответствие с целевой направленностью операции.

1. Исследуется рентабельность авиационного предприятия, причем предполагается осуществить ряд мер с целью повышения эффективности его работы. В качестве показателя эффективности предлагается прибыль (или средняя прибыль), полученная за год.

2. Авиационная техническая база занимается ремонтом технических средств. Её доходность определяется числом этих средств, отремонтированных в течение дня. Показатель эффективности – среднее число отремонтированных средств.

3. План снабжения предприятия. Требуется разработать такой план снабжения, чтобы потребность в сырье и других комплектующих изделиях удовлетворялась. Показатель эффективности – расходы на перевозки.

4. Составление расписания движения воздушных судов. Показатель эффективности – регулярность движения воздушных судов.

5. Размещение средств УВД, навигации, посадки и связи в масштабах страны, зон УВД. Показатель эффективности – пропускная способность воздушных судов, безопасность воздушного движения.

6. Создание объединенного центра УВД. Показатели эффективности – пропускная способность, безопасность воздушного движения, стоимость создания объединенного центра.

7. Разработка средств УВД, навигации, посадки и связи. Показатели эффективности:

- средств УВД – дальность, вероятность правильного обнаружения, вероятность ложной тревоги.
- средств связи – дальность, разборчивость речи.
- средств навигации – дальность, точность.
- средств посадки – дальность, точность.

1.3. Математические модели исследования операций

Для применения количественных методов исследования операций необходимо построить математическую модель исследуемого объекта или процесса.

Общих способов построения математических моделей не существует. В каждом конкретном случае модель строится исходя из целевой направленности операции и задачи научного исследования, с учетом требуемой точности решения, а также точности, с какой могут быть известны исходные данные.

Требования модели противоречивы. С одной стороны, она должна быть достаточно полной, то есть должны быть учтены все важные факторы, от которых существенно зависит исход операции. С другой стороны модель должна быть достаточно простой для того, чтобы можно было установить обозримые зависимости между входящими в нее параметрами. Наиболее простой пример – электрический расчет усилителя. Показателем эффективности является мощность (напряжение) на выходе усилителя при известном напряжении на входе. Количественные связи – $K_У = \prod_{i=1}^{n} K_i$, где n – количество каскадов. Но при расчетах K_i делаются определенные допуски, что снижает точность расчета усилителя и он, как правило, работать не будет. Второй пример. Вероятность безотказной работы усилителя $P\ t\ = e^{\lambda t}$, где λ - интенсивность отказов. Среднее время безотказной работы $T_0 = \frac{1}{\lambda}$. Но если использовать табличные значения λ и не учитывать режим работы усилителя, то среднее время безотказной работы T_0 будет меньше в несколько раз. Математические модели, применяемые в настоящее время в задачах исследования операций, подразделяются на два класса: аналитические и статические [1, 3].

Для первых характерно установление аналитических (формульных) зависимостей между параметрами задачи, записанных в виде алгебраических, дифференциальных уравнений, уравнения с частными производными и так далее. С помощью аналитических моделей с удовлетворительной точностью описываются операции, где число взаимодействующих элементов мало.

Аналитические модели, как правило, более грубые, описывают явления приближенно и позволяют получить представление результатов скорее на качественном уровне. Но зато аналитическая форма наиболее полно отражает присущие явлению основные закономерности.

В операциях, где учитывается множество факторов, в том числе и случайные, на первый план выходит метод статического моделирования. Статические модели более точные, но трудно поддаются анализу.

Опыт показывает, что наилучшие результаты достигаются при совместном использовании аналитических и статических моделей. Рассмотрим некоторую операцию (управляемое мероприятие), на исход которой можно повлиять, выбирая те или иные значения управляющих параметров. Эффективность операции характеризуется показателем эффективности W, который заранее известен. Требуется максимизировать W при определенных ограничивающих условиях.

Допустим, что построенная математическая модель операции даст возможность вычислить значение показателя эффективности для той совокупности условий, в которых выполняется операция. Рассмотрим простейший случай, когда факторы, от которых зависит успешность выполнения операции, можно разбить на 2 группы:

- заданные, известные факторы (параметры) $a_1, a_2, \ldots, a_n$, которые представляют собой числовые характеристики неизменяемых условий проведения операций;
- факторы (элементы решения) $x_1, x_2, \ldots, x_n$, которые подлежат определению в процессе выбора решения.

Показатель эффективности является функцией, зависящей от обеих групп факторов, и записывается в виде

$$W = W\left(a_1, a_2, \ldots, a_n, x_1, x_2, \ldots, x_n\right).$$

Поэтому задачу исследования операций математически сформулировать следующим образом: при заданных условиях $a_1, a_2, \ldots, a_n$ найти элементы решения $x_1, x_2, \ldots, x_n$, которые максимизируют показатель эффективности W.

Для нахождения экстремума функции необходимо найти производные по всем аргументам, прировнять их к нулю и решить полученную систему уравнений. Однако этот простой метод в задачах исследования операций имеет ограниченное применение, чему имеется несколько причин.

1. В задачах исследования операций обычно имеется большое количество элементов решения. Поэтому совместное решение системы уравнений, полученных дифференцированием, оказывается сложным.

2. На практике часто встречаются случаи, когда из-за некоторых особенностей исследуемых функций или из-за дискретного изменения аргумента производные не существуют.

Поэтому общих математических методов экстремумов функций произвольного вида при наличии любых ограничений не существует.

Глава 2. Линейное программирование
2.1. Постановка задач линейного программирования

Во многих областях производства возникают задачи оптимизации решения, для которых характерны следующие особенности [2, 3, 6]:

- показатель эффективности W представляет собой линейную функцию от элементов решения $x_1, x_2, \ldots, x_n$;
- ограничения, налагаемые на элементы решения, имеют вид линейных равенств.

К таким задачам относятся:

- задача планирования производства;
- задача о загрузке оборудования для технического обслуживания и ремонта воздушных судов и наземной техники;
- задача о перевозках (транспортная задача);
- задача снабжения запасными частями и сырьем и др.

2.2. Основная задача линейного программирования (ОЗЛП)

Любую задачу линейного программирования можно свести к стандартной форме, так называемой “основной задаче линейного программирования”, которая формулируется так: найти неотрицательные значения переменных $x_1, x_2, \ldots, x_n$, которые удовлетворяли бы условиям [2]:

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = B_1, \\ &a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = B_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = B_{mn}, \end{aligned} \qquad (1)$$

Кроме того минимизировали (максимизировали) линейную функцию

$$W = C_1X_1 + C_2X_2 + C_3X_3, \qquad (2)$$

Условимся называть допустимым решением ОЗЛП любую совокупность переменных

$x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$, удовлетворяющую уравнениям (1).

Оптимальным решением будем называть то из допустимых решений, при котором функция (2) обращается в максимум (минимум).

2.3. Методы нахождения допустимых решений

Пусть имеется система уравнений (1). Если система уравнений – ограничений ОЗЛП совместна, то матрица системы и её расширенная матрица имеют один и тот же ранг, и система уравнений совместна. Если ранги не равны, система несовместна.

Ранг системы не может быть больше числа уравнений m, т.е. $r \le m$. Ранг системы не может быть больше общего числа переменных n, т.е. $r \le n$.

Если $r = n$, то определитель, составленный из коэффициентов a_{ij} не равен нулю. Следовательно, в этом случае система (1) имеет единственное решение. Чтобы найти величину x_i достаточно в определителе Δ заменить i-й столбец – столбцом свободных членов и разделить на Δ, т.е.

$$\Delta = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{ni} \end{pmatrix} \rightarrow \Delta_i \begin{pmatrix} a_{11} & a_{12} \ldots b_{1m} & a_{1n} \\ a_{21} & a_{22} \ldots b_{2m} & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} \ldots b_{m} & a_{nn} \end{pmatrix}, \tag{3}$$

$$X_i = \frac{\Delta_i}{\Delta}.$$

На практике случай, когда количество уравнений равно количеству неизвестных − редкий случай. Рассмотрим более общий случай, когда число независимых уравнений меньше самих переменных. Для рассматриваемого случая существует бесконечное множество решений. При этом $n - r$ переменным мы можем придавать произвольные значения (т.е. свободные переменные), а остальные r переменных выразить через них (назовем их базисными).

Рассмотрим систему двух уравнений с четырьмя неизвестными:

$$\begin{aligned} &2x_1 - x_2 + x_3 - x_4 = 1, \\ &-x_1 + x_2 - 2x_3 + x_4 = 2. \end{aligned} \tag{4}$$

В данной системе количество уравнений m равно рангу $r = m = 2$. Количество неизвестных $n = 4$. В качестве базисных выбираем x_1 и x_2, а свободными членами x_3 и x_4.

Выразим базисные через свободные:

$$\begin{aligned} &2x_1 - x_2 = 1 - x_3 + x_4, \\ &-x_1 + x_2 = 2 + 2x_3 - x_4. \end{aligned} \tag{5}$$

Сложив, получим: $x_1 = 3 + x_3$.

Умножаем второе уравнение на 2 и складываем с первым. Получим $x_2 = 5 + 3x_3 - x_4$.

Свободным переменным можно придавать любые значения. Все равно получим совокупность x_1, x_2, x_3, x_4, удовлетворяющую системе уравнений.

Пусть $x_3 = x_4 = 0$, получим $x_1 = 3$ и $x_2 = 5$.

Если $x_3 = x_4 = 1$, то $x_1 = 4$, а $x_2 = 7$.

Таким образом, решений будет множество. Теперь возникает следующая задача − найти среди допустимых решений оптимальное, т.е. такое решение, для которого линейная функция $W = c_1x_1 + c_2x_2 + ... + c_nx_n \to \max(\min)$.

Но прежде, чем перейти к решению этой основной задачи, рассмотрим каким образом от задачи с ограничением неравенств можно перейти к ОЗЛП и обратно.

2.4. Переход от задачи линейного программирования с ограничениями типа неравенств к ОЗЛП и обратно

Пусть имеется задача линейного программирования с n переменными $x_1, x_2, \ldots, x_n$ в которой ограничения, наложенные на переменные имеют вид линейных неравенств. Представим неравенства в стандартной форме

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + b_1 \geq 0, \\ &a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + b_2 \geq 0, \\ &\vdots \\ &a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + b_n \geq 0. \end{aligned} \tag{6}$$

Требуется найти такую совокупность неотрицательных значений $x_1, x_2, \ldots, x_n$, которая удовлетворяла бы неравенству (6) и обращало бы в минимум линейную функцию

$$W = c_1x_1 + c_2x_2 + ... + c_nx_n \to \min \tag{7}$$

От такой задачи легко перейти к основной ЗЛП. Введем обозначение:

$$\begin{aligned} &y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n} + b_1, \\ &y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n} + b_2, \\ &\vdots \\ &y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn} + b_n, \end{aligned} \tag{8}$$

где $y_1, y_2, \ldots, y_m$ новые параметры, которые будем называть добавочными.

Согласно условию (7) эти переменные $x_1, x_2, \ldots, x_n$ должны быть неотрицательными.

Таким образом, возникает задача линейного программирования в следующей постановке.

Найти неотрицательные значения переменных $x_1, x_2, \ldots, x_n$, $y_1, y_2, \ldots, y_m$ чтобы они удовлетворяли условию (6) и обращали бы W в минимум (максимум) [1].

Рассмотрим пример. Имеется задача линейного программирования с ограниченными неравенствами.

$$\begin{aligned}
&2x_1 - x_2 + 3x_3 \le 6,\\
&x_3 - 3x_2 \le -1,\\
&x_5 - 2x_4 + x_1 \ge -1,\\
&x_5 - x_1 \le 0.
\end{aligned} \qquad (9)$$

Найти неотрицательные значения переменных x_1, x_2, x_3, x_4, x_5, которые обращают в минимум линейную функцию:

$$W = x_1 - 2x_2 - 3x_3 \to \min. \qquad (10)$$

Приводим неравенство (9) к стандартной форме.

$$\begin{aligned}
&-2x_1 + x_2 - 3x_3 + 6\\
&-1 - x_3 + 3x_2 \ge 0,\\
&1 + x_5 - 2x_4 + x_1 \ge 0\\
&x_1 - x_5 \ge 0.
\end{aligned} \qquad (11)$$

Вводим дополнительные переменные:

$$\begin{aligned}
&y_1 = -2x_1 + x_2 - 3x_3\\
&y_2 = -1 - x_3 + 3x_2,\\
&y_3 = 1 + x_5 - 2x_4 + x_1\\
&y_4 = x_1 - x_5.
\end{aligned} \qquad (12)$$

Теперь необходимо определить неотрицательные переменные $x_1, x_2, x_3, x_4, x_5, y_1, y_2, y_3, y_4$ удовлетворяющие уравнениям (12) и обращающие в минимум (максимум) линейную функцию (10).

Возникает и обратный переход от ОЗЛП к задаче с ограниченными параметрами.

Пример. Имеется ОЗЛП:

$$\begin{aligned}
&x_1 + x_2 = 1,\\
&x_2 - 2x_3 = -3,\\
&x_3 - x_4 + x_5 = 1.
\end{aligned} \qquad (13)$$

И минимизируем функцию $W = -x_1 - x_2 + x_5$. (14)

Так как $m = 3$, $n = 5$, и $n - m = 2$, то выбираем какие-то две переменные в качестве свободных. Переменные x_1 и x_2 в качестве свободных выбирать нельзя, так как они связаны с первым уравнением системы (13). По этой же причине в качестве свободных нельзя выбирать переменные x_2 и x_3. Они связаны уравнением (2). Поэтому выбираем в качестве свободных переменных x_1 и x_4 и выразим через них все остальные x_2, x_3, x_5 .

$$x_2 = 1 - x_1,$$
$$x_3 = -\frac{1}{2}x_1 + 2, \qquad (15)$$
$$x_5 = \frac{1}{2}x_1 + x_4 - 1.$$

Так как $x_2 \geq 0,\ x_3 \geq 0,\ x_5 \geq 0$, условия (15) могут быть заменены неравенством:

$$1 - x_1 \geq 0,$$
$$-\frac{1}{2}x_1 + 2 \geq 0, \qquad (16)$$
$$\frac{1}{2}x_1 + x_4 - 1 \geq 0.$$

Перейдем в выражение линейной функции W к свободным переменным x_1 и x_4.

Подставляя в W вместо x_2 и x_5 выражение (16), имеем:

$$W = -x_1 + x_1 - 1 + \frac{1}{2}x_1 + x_4 - 1 = \frac{1}{2}x_1 + x_4 - 2.$$

Получили задачу ЛП с ограниченными неравенствами.

Методы нахождения оптимальных решений ЗЛП [3, 6]

Имеется задача ЛП:

$$-5x_1 - x_2 + 2x_3 \leq 2,$$
$$-x_1 + x_3 + x_4 \leq 5, \qquad (18)$$
$$-3x_1 + 5x_4 \leq 7.$$

Требуется оптимизировать функцию $W = 5x_1 - 2x_3$. (19)

Приведем (18) к стандартной форме

$$2 + 5x_1 + x_2 - 2x_3 \geq 0,$$
$$5 + x_1 - x_3 - x_4 \geq 0, \qquad (20)$$
$$7 + 3x_1 - 5x_4 \geq 0.$$

Введем добавочные переменные

$$y_1 = 2 + 5x_1 + x_2 - 2x_3,$$
$$y_2 = 5 + x_1 - x_3 - x_4, \qquad (21)$$
$$y_3 = 7 + 3x_1 - 5x_4.$$

Число переменных $n = 7$ $x_1, x_2, x_3, x_4, y_1, y_2, y_3$, а количество уравнений $m = 3$. Тогда $n - m = 4$, которые выбираем в качестве свободных переменных. Выбираем в качестве таковых x_1, x_2, x_3, x_4. Получаем опорное решение:

$x_1 = x_2 = x_3 = x_4 = 0$, $y_1 = 2$, $y_2 = 5$, $y_3 = 7$. Однако из уравнения (19) видно, что $W = 0$. Это решение не является оптимальным, так как в выражении (19) x_3 отрицательно.

Следовательно, увеличивая x_3, можно уменьшить W. Посмотрим, какая переменная y_1 и y_2 раньше обращаются в нуль при увеличении x_3.

Очевидно, что $y_1 = 0$, при $x_3 = 1$, а $y_2 = 0$ при $x_3 = 1$.

Поэтому переменную y_1 вводим в число свободных членов вместо x_3.

Перепишем первое уравнение в (20) в виде: $x_3 = \frac{5}{2}x_1 + \frac{x_2}{2} - \frac{y_1}{2} + 1$.

Подставим выражение x_3 во второе уравнение (21) и получим:

$$y_2 = 5 + x_1 - \frac{5}{2}x_1 - \frac{x_2}{2} + \frac{y_1}{2} + x_4 - 1 = -\frac{3}{2}x_1 - \frac{x_2}{2} + \frac{y_1}{2} - x_4 + 4.$$

Третье уравнение (19) не содержит x_3 и оно не изменяется.

Получили новую систему уравнений:

$$\begin{aligned} x_3 &= \frac{5}{2}x_1 + \frac{x_2}{2} - \frac{y_1}{2} + 1, \\ y_2 &= -\frac{3}{2}x_1 - \frac{x_2}{2} + \frac{y_1}{2} - x_4 + 4, \\ y_3 &= 7 + 3x_1 - 5x_4. \end{aligned} \tag{22}$$

Со свободными переменными x_1, x_2, x_4, y_1 и базисными x_3, y_2, y_3.

Выразим линейную функцию W через новые свободные переменные:

$W = -x_2 + y_1 - 2$.

Снова положим свободные переменные $x_2 = y_1 = 0$ (23). Получаем $W = -2$.

Это лучше, чем $W = 0$. Но это решение также не является оптимальным, так как x_2 в уравнении (23) отрицательно.

Обменяем местами x_2 и y_2, т.е. первую выведем из свободных, а вторую введем.

Для этого разрешим второе уравнение системы (21) относительно x_2 и подставим выражение для x_2 в первое уравнение. Получим еще один вид системы:

$$\begin{aligned} x_3 &= x_1 - y_2 - x_4 + 5, \\ x_2 &= -3x_1 - 2y_2 + y_1 - 2x_4 + 8, \\ y_3 &= 3x_1 - 5x_4 + 7. \end{aligned} \tag{24}$$

Выразим W через новые свободные переменные $W = 3x_1 + 2y_2 + 2x_4 - 10$ (25). Полагаем, что $x_1 = x_4 = y_2 = 0$ и получаем $W = -10$.

Это решение является оптимальным, так как все свободные переменные в выражении (25) не отрицательны. Следовательно, оптимальное решение ОЗЛП является $x_1^* = 0, x_2^* = 8, x_3^* = 5, x_4^* = 0, y_1^* = 0, y_2^* = 0, y_3^* = 7$.

2.5. Транспортная задача линейного программирования

Задача ставится следующим образом [2, 5]. Имеется m пунктов отправления (ПО) $A_1, A_2, ..., A_m$, в которых сосредоточены запасы однородных грузов в количествах соответственно $a_1, a_2, ..., a_m$ единиц. Имеется n пунктов назначения (ПН) $B_1, B_2, ..., B_n$, подававших заявки соответственно на $b_1, b_2, ..., b_n$ единиц. Сумма всех заявок равна сумме всех запасов:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j . \qquad (26)$$

Известны стоимости C_{ij} перевозки единицы грузов от каждого ПО A_i до каждого ПН B_j. Все числа C_{ij} образуют прямоугольную таблицу:

$$\begin{pmatrix} C_{11} & C_{12} \dots & C_{1n} \\ C_{21} & C_{22} \dots & C_{2n} \\ \vdots & \vdots & \vdots \\ C_{m1} & C_{m2} & C_{mn} \end{pmatrix} \qquad (27)$$

Требуется составить такой план перевозок, чтобы все заявки были выполнены, а общая стоимость всех перевозок была бы минимальной. Поставим эту задачу линейного программирования. Обозначим x_{ij} – количество единиц груза, отправляемого из i-го ПО в j-й ПН.

Совокупность чисел x_{ij} будем называть планом перевозок, а сами величины x_{ij} – перевозками. При этом должны удовлетворяться следующие условия:

1. Суммарное количество грузов, не направляемого из каждого ПО во все ПН, должно быть равно запасу груза в должном пункте это даст m условий-равенств:

$$\begin{aligned} & x_{11} + x_{12} + ... + x_{1n} = Q_1, \\ & x_{21} + x_{22} + ... + x_{2n} = Q_2, \\ & \vdots \\ & x_{m1} + x_{m2} + ... + x_{mn} = a_m. \end{aligned} \qquad (28)$$

2. Суммарное количество груза, доставляемого в каждый ПН из всех ПО должно быть равно заявке, поданной данным пунктом. Это даст нам n условий равенств:

$$x_{11} + x_{12} + ... + x_{1n} = b_1,$$
$$x_{21} + x_{22} + ... + x_{2n} = b_2, \qquad (29)$$
$$\vdots$$
$$x_{m1} + x_{m2} + ... + x_{mn} = b_n.$$

3. Суммарная стоимость всех перевозок должна быть минимальной:

$$W = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij} \to \min, \qquad (30)$$

где знак двойной суммы означает, что суммирование производится по всем комбинациям индексов i и j, т.е. по всем параметрам ПО-ПН. Мы видим, что перед ними задача линейного программирования. Особенностью этой задачи является то, что все коэффициенты в условиях (29) и (30) равны единице.

В силу особой структуры транспортной задачи при ее решении приходится долго решать систему уравнений. Все операции по нахождению оптимального плана сводятся к манипуляции непосредственно с таблицей, где в определенном порядке записываются условия транспортной задачи: перечень ПО и ПН заявки и запасы, а также стоимости перевозок. Транспортная таблица состоит из m строк и n столбцов. В правом верхнем углу каждой клетки будет ставить стоимости перевозок единицы груза, а в центре помещаем саму перевозку X_{ij}.

Для наглядности рассмотрим пример со следующими исходными данными:

$A_1 = 30$ $B_1 = 18$ $C_{11} = 13$ $C_{21} = 11$ $C_{31} = 6$ $C_{41} = 14$

$A_2 = 48$ $B_2 = 27$ $C_{12} = 7$ $C_{22} = 8$ $C_{32} = 10$ $C_{12} = 8$

$A_3 = 20$ $B_3 = 42$ $C_{13} = 14$ $C_{23} = 12$ $C_{33} = 10$ $C_{43} = 10$

$A_4 = 30$ $B_4 = 15$ $C_{14} = 7$ $C_{27} = 6$ $C_{34} = 8$ $C_{44} = 10$

$B_5 = 26$ $C_{15} = 5$ $C_{25} = 8$ $C_{35} = 11$ $C_{45} = 15$

Пример транспортной таблицы, где приведены условия задачи и стоимости перевозок, представлен в табл. 1.

Общая транспортная модель с m пунктами отправления n пунктами назначения имеет $m+n$ ограничений в виде равенств, по одному на каждый пункт отправления и назначения. Поскольку транспортная модель всегда сбалансирована (сумма предложений равна сумме спроса) одно из этих равенств должно быть избыточным. Таким образом, транспортная модель имеет $m+n-1$ независимых ограничений.

Специальная структура транспортной модели позволяет применять следующие методы решения [1, 4, 6]:

1. Метод северо-западного угла.

2. Метод наименьшей стоимости.

Таблица 1

Пример транспортной таблицы

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы, a_i
A_1	13	7	14	7	5	30
A_2	11	8	12	6	8	48
A_3	6	10	10	8	11	20
A_4	14	8	10	10	15	30
Заявки, b_j	18	27	42	15	26	128

При использовании метода северо-западного угла заполнение таблицы начинается с левого верхнего угла табл. 2. Пункт B_1 подал заявку на 18 единиц груза. Удовлетворим ее из запасов пункта A_1. После этого в нем останется еще $30-18=12$ единиц груза. Отдадим их пункту B_2. Но заявка пункта B_2 ещё не удовлетворена. Выделим недостающие 15 единиц из запасов пункта A_2 и т.д. Рассуждая таким же образом, заполним до конца транспортную таблицу (табл. 2).

Для метода северо-западного угла стоимость перевозок будет равна:
$L=18\cdot 13+12\cdot 7+15\cdot 8+22\cdot 12+11\cdot 6+20\cdot 10+4\cdot 10+26\cdot 15=1398$.

Метод наименьшей стоимости находит лучшее решение, чем метод северо-западного угла, поскольку заполняются клетки табл. 3 с минимальными стоимостями перевозок. Пункт назначения B_5 заказал 26 единиц груза. Удовлетворим это из запасов A_1, а оставшиеся в A_1 четыре единицы груза выделим B_4, который заказал 15 единиц груза. Необходимые для B_4 одиннадцать единиц груза выделим из запасов A_2. Аналогично заполняем до конца остальные клетки табл. 3.

Для рассматриваемого случая стоимость перевозок плана будет равна:
$L=4\cdot 7+26\cdot 5+27\cdot 8+10\cdot 12+11\cdot 6+20\cdot 10+18\cdot 14+12\cdot 10=1132$.

Таким образом, использование метода наименьшей стоимости позволило уменьшить стоимость перевозок на $1398-1132=266$ единиц.

Таблица 2

Заполнение транспортной таблицы по методу северо-западного угла

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы, a_i
A_1	13 18	7 12	14	7	5	30
A_2	11	8 15	12 22	6 11	8	48
A_3	6	10	10 20	8	11	20
A_4	14	8	10	10 4	15 26	30
Заявки, b_j	18	27	42	15	26	128

Таблица 3

Заполнение транспортной таблицы по методу наименьшей стоимости

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы, a_i
A_1	13	7	14	7 4	5 26	30
A_2	11	8 27	12 10	6 11	8	48
A_3	6	10	10 20	8	11	20
A_4	14 18	8	10 12	10	15	30
Заявки, b_j	18	27	42	15	26	128

Глава 3. Динамическое программирование
3.1. Метод динамического программирования

Динамическое программирование представляет собой математический метод оптимизации решений, приспособленный к многошаговым (многоэтапным) операциям [3, 5, 10].

К таким задачам относятся:

1. Прокладка наивыгоднейшего пути между двумя пунктами (для ГА, РЖД и автомобильного транспорта).
2. Распределение ресурсов.
3. Составление перспективных и текущих планов отрасли, корпораций, предприятий.
4. Расчет радиоэлектронных средств.

Пусть исследуемая операция Q представляет собой процесс, развивающийся во времени и распадающийся на ряд шагов или этапов. Пример расчета РЭС распадается на расчет отдельных каскадов. Процесс, о котором идет речь, является управляемым, т.е. на каждом шаге/этапе применяется какое-то решение, от которого зависит успех в целом.

Предположим, что выигрыш W за всю операцию складывается из выигрышей на отдельных шагах:

$$W = \sum_{i=1}^{n} w_i ,$$

где w_i − выигрыш на i-м шаге.

Операция представляет собой управляемый процесс, т.е. мы можем выбирать какие-то параметры, влияющие на вход и исход операции, причем на каждом шаге выбирается какое-то решение, от которого зависит выигрыш на данном шаге и выигрыш на операцию в целом. Будет называть это решение “шаговым управлением”. Обозначим его буквой x, а шаговые управления − буквами $x_1, x_2, \ldots, x_m$:

$$x = x_1, x_2, \ldots, x_m \; .$$

Требуется найти такое управление x, при котором выигрыш W обращается в максимум, т.е.

$$W^* = \sum_{i=1}^{n} w_i \to \max .$$

Управление x^*, при котором этот максимум достигается, будем называть оптимальным управлением. Оно состоит из совокупности оптимальных шаговых управлений

$$x^* = x_1^*, x_2^*, \ldots, x_m^* \; .$$

Поставленную задачу можно решать разными способами [10]:

1. Искать оптимальное управление для всего периода.

2. Строить его постепенно, шаг за шагом, таким образом, что на каждом этапе расчета оптимизируется только один шаг процесса.

Обычно второй способ оптимизации оказывается существенно проще первого, особенно при большом числе шагов процесса. Реализация такого подхода составляет основное содержание метода динамического программирования.

Планируя многошаговую операцию нужно выбирать управления на каждом шаге с учетом его будущих последствий на еще предстоящих шагах. Управление на i - м шаге выбирается не так, чтобы выигрыш именно на данном шаге был максимален, а так, чтобы была максимальна сумма выигрышей на всех оставшихся до конца шагах плюс данный.

Однако среди всех шагов есть один, который может планироваться без оглядки на будущее. Это последний шаг. Поэтому процесс динамического программирования обычно разворачивается от конца к началу, в результате чего находится условное управление и соответствующий оптимальный выигрыш на последнем m-м шаге.

На этом шаге делаются предположения (гипотезы) относительно предпоследнего $m-1$ шага. Теперь можно оптимизировать управление на $m-1$ шаге. Снова сделаем все возможные предположения, о том, чем кончится предыдущий $m-2$ - й шаг. Далее двигаясь назад, оптимизируем остальные шаги, пока не дойдем до первого.

Таким образом, определили условное оптимальное управление и условный оптимальный выигрыш для всех шагов, начиная от первого и до конца процесса.

Предположим, что все условные оптимальные управления и условные оптимальные выигрыши на всех шагах нам известны. Тогда можно определить уже не условное, а безусловное оптимальное управление и безусловный выигрыш.

В самом деле, предположим, что известно начальное состояние операции, обозначим его S_0. Знаем, что делать на первом шаге – нужно использовать то условие управления, которое выработано для первого шага. В результате выполнения этого управляющего воздействия оптимизируемая система после первого шага в некоторое состояние S_2, но для этого состояния уже известно условное оптимальное управление на втором шаге и т.д. Таким образом будет найдено безусловное оптимальное управление операции.

Проведенное описание показывает, что в процессе оптимизации управления методом динамического программирования многошаговый процесс выполняется дважды:

- первый раз он идет от конца к началу, в результате чего находятся условные оптимальные управления на всех шагах и условный оптимальный выигрыш на каждом шаге.

- второй раз он проходит от начала к концу, при этом определяются безусловные управления на всех шагах операции и тем самым находится безусловное оптимальное управление всей операции в целом.

Первый этап условной оптимизации несравненно сложнее и длительнее второго. Второй этап почти не требует дополнительных вычислений.

Рассмотрим пример использования динамического программирования для расчета усилителя низкой частоты (УНЧ). На вход усилителя подается напряжение U_1, а на выходе должны получить напряжение U_m. Коэффициент усиления:

$$K_У = \frac{U_m}{U_1} = K_{У1} K_{У2} \cdots K_{Уm},$$

где $K_{У1}, K_{У2}, \ldots, K_{Уm}$ – коэффициент усиления первого, второго, последнего каскада УНЧ.

Коэффициент усиления каждого каскада:

$$K_{У2} = S \cdot R_Э,$$

где S – крутизна вольтамперной характеристики транзистора каскада; $R_Э$ – эквивалентное сопротивление нагрузки каскада:

$$R_Э = \frac{R_Н \cdot R_{ВХ}}{R_Н + R_{ВХ}},$$

где $R_Н$ – сопротивление нагрузки; $R_{ВХ}$ – входное сопротивление следующего каскада.

Если расчет осуществлять от первого каскада, то необходимо сделать предположение о входном сопротивлении второго каскада, о котором еще ничего не известно. Поэтому ошибки расчета $K_У$ каждого каскада будут большими.

При расчете усилителя с последнего каскада известно U_m и сопротивление нагрузки. На основании этих данных выбирается транзистор, а по результатам расчета каскада определяется $R_{ВХ}$ и напряжение на входе последнего каскада, т.е. напряжение на выходе предпоследнего каскада. Поступая аналогично, рассчитываются все каскады усилителя.

Рассмотренный метод также можно использовать для расчетов приемников, передатчиков и других устройств РЭС радиолокации, радионавигации, посадки и связи.

3.2. Прокладка наивыгоднейшего пути между двумя пунктами

Необходимо соорудить путь, соединяющий два пункта A и B, из которых второй лежит к северо-востоку от первого. Для простоты допустим, что сооружение пути состоит из ряда шагов и на каждом шаге можно двигаться либо строго на север, либо строго на восток. Любой путь из A в B представляет собой ступенчатую ломанную линию, отрезки которой параллельны одной из координатных осей. Затраты на сооружение каждого из таких отрезков известны. Требуется проложить такой путь из A в B, при котором суммарные затраты минимальны.

Можно поступить одним из двух способов:

- либо перебрать все возможные варианты и выбрать тот, на котором затраты минимальны;
- либо разделить процесс перехода из A в Bна отдельные шаги и оптимизировать управление по шагам.

Оказывается второй способ предпочтительней.

Продемонстрируем, как это сделать. Разделим расстояние от A до B в восточном направлении на 7 частей, а в северном − на 5 частей. Тогда любой путь из A в B состоит из $m = 7 + 5 = 12$ отрезков. Проставим на каждом из отрезков стоимость прокладки пути по этому отрезку (табл. 4)

Процедуру условной оптимизации будет разворачиваться от конца точки B к точке A. Прежде всего произведем условную оптимизацию на последнем шаге. У нас два пути на север (10 ед.) или на восток (14 ед.). Двигаемся на север до точки D. С этой точки можем двигаться на восток (13 ед.) или север (11 ед.). Двигаемся на север (точка E) и так далее. Условные минимальные затраты

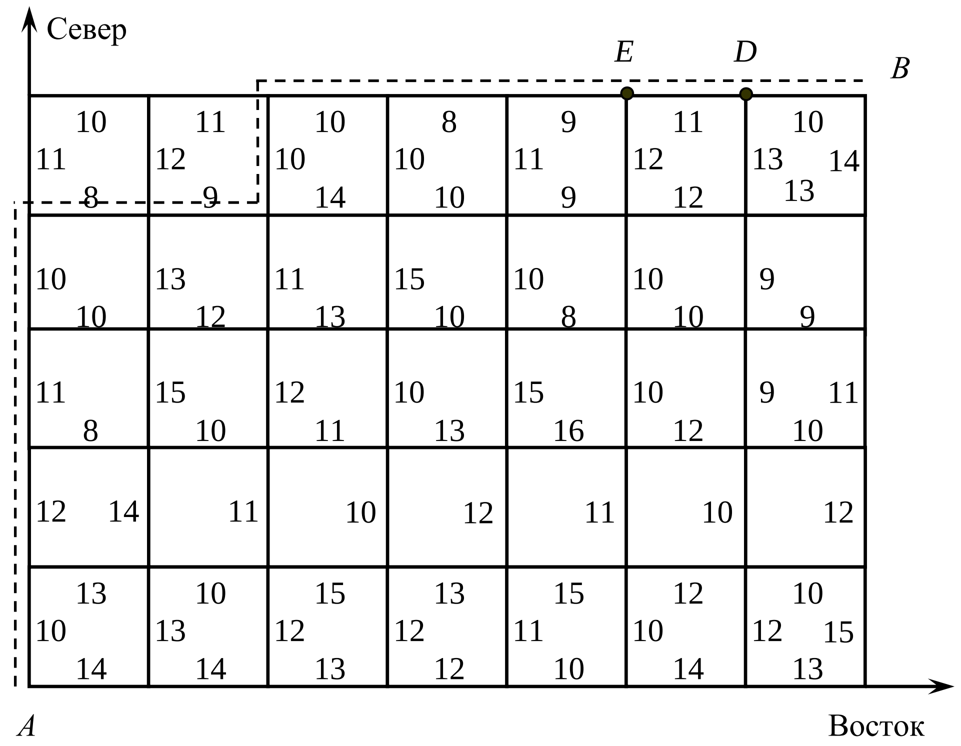
$$L = 10 + 11 + 9 + 8 + 10 + 10 + 9 + 8 + 10 + 11 + 12 + 10 = 118$$

Теперь остается построить безусловное оптимальное управление – траекторию ведущую из A в B самым дешевым способом. Для этого нужно “слушаться стрелок”, т.е. что они предписывают, что делать на каждом шаге.

Соответственно безусловное оптимальное управление будет $X = C,C,C,C,B,B,C,B,B,B,B,B$, т.е. первые 4 шага на север, следующие 2 шага на восток, затем 1 на север, остальные 5 на восток, $W = 118$ усл. ед.

Таблица 4

Иллюстрация решения задачи определения оптимального пути



3.3. Общая постановка задачи динамического программирования

Имеется некоторая динамическая система S, которая с течением времени может менять свое состояние. Мы можем управлять этим процессом, т.е. тем или иным способом влиять на состояние системы.

Из множества возможных управлений найти такое оптимальное управление, которое переводит систему S из начального состояния в конечное состояние, так чтобы при этом показатель качества W обращался в минимум или максимум.

Запишем принципиальную структуру обоих стадий оптимизации (с начала в конец и с конца в начало). Введем обозначение.

Пусть $W_i(S)$ – условный оптимальный выигрыш, получаемый на всех последующих шагах, начиная с i-го и до конца, $x_i(S)$ – условное оптимальное управление на i- м шаге.

Поставим задачу: определить функции $W_i(S)$ и $x_i(S)$, т.е. условный оптимальный выигрыш и условное оптимальное управление для всех шагов $i=1,2,\ldots,m$.

Рассмотрим i- й шаг процесса управления. Пусть в результате $i-1$ предыдущих шагов процесса управления система пришла в состояние S и мы выбираем некоторое управление X_i на i- м шаге. Если мы его применим, то на i- м шаге получим выигрыш $W_i = W(S, x_i)$.

Под влиянием управления X_i на i- м шаге система из состояния S перейдет в состояние S'. Это новое состояние $S' = \varphi_i(S, x_i)$.

Запишем выигрыш, который получим на всех шагах i- го , если на i- м шаге было примерно любое (вообще говоря не оптимальное) управление X_i, а на всех последующих шагах от $i+1$ до m-го примерно оптимальное управление. Этот выигрыш будет равен выигрышу w_i на i- м шаге плюс условный оптимальный выигрыш на всех последующих шагах. Обозначим такой полуоптимальный выигрыш через $\hat{w}_i = (S, X_i)$:

$$\hat{W}(S, X_i) = w_i(S, X_i) + w_i+1(S^1)$$

или

$$\hat{W}_i(S, X_i) = w_i(S, X_i) + w_i+1\left[\varphi(S, X_i)\right]$$

Теперь в соответствии с принципом оптимизации, мы должны выбрать управление, при котором $\hat{W}_i(S, X_i)$ максимален и достигает значение:

$$W_i(S) = \max\left(w_i(S, X_i) + W_{i+1}\left[\varphi(S, X_i)\right]\right). \qquad (31)$$

Таким образом, управление при котором этот максимум достигает и есть условное оптимальное управление на i- м шаге, а сама величина $W_i(S)$ – условный оптимальный выигрыш на всех шагах начиная с i-го и до конца.

Формула (31) есть основное функциональное уравнение динамического программирования [4, 10]. Что касается функции $W_m\ S$ (условный оптимальный выигрыш на последнем шаге), то она может быть определена очень просто. Действительно за последним шагом отсутствует другой шаг. Поэтому

$$W_m\ S\ = \max\left[w_m\ S, X_m\ \right].$$

Зная $W_m\ S, X_m$, можно по формуле (31) , полагая в ней $i+1=m$, найти функцию W_{m-1} и соответствующее оптимальное управление, затем $W_{m-2}\ S$ и $X_{m-2}\ S$ и так далее вплоть до последнего шага.

Таким образом, порядок решения задачи динамического программирования сводится к следующему:

1. Выбрать способ описания процесса, т.е. параметры характеризующие состояние системы.

2. Расчленить операцию на шаги (этапы):

- шаги должны быть достаточно мелким, чтобы процедура оптимизации была достаточно простой;

- шаги не должны быть слишком мелкими, чтобы не производить ненужных расчетов поиска оптимального решения;

3. Записать выигрыш на i - м шаге в зависимости от состояния системы S в начале этого шага и управление X_i:

$$w_i = w\ S, x_i\ .$$

4. Записать для i - го шага функцию, выражающую изменение состояния системы от S к S' под влиянием управления X_i:

$$S' = \varphi\ S, X_i\ .$$

5. Записать основное функциональное уравнение

$$W_i\ S\ = \max\ w_i\ S, X_i\ + W_{i+1}\left[\varphi\ S, X_i\ \right]\ .$$

6. Найти функцию $W_m\ S$ (условный оптимальный выигрыш) для последнего шага:

$$W_m\ S\ = \max\left[w_m\ S, X_m\ \right].$$

7. Зная $W_m\ S$ и пользуясь уравнением для $W_i\ S\ \ n,5$ при конкретном виде функций $w_i\ S, X_i$ и $\varphi_i\ S, X_i$ найти одну за другой $W_{m-1}\ S\ , W_{m-2}\ S\ , \ldots, W_1\ S$ и соответствующие им условные оптимальные управления $X_{m-1}\ S\ , X_{m-2}\ S\ , \ldots, X_1\ S\ .$

3.4. Задача распределения ресурсов

Постановка задачи. Имеется определенное начальное количество средств K_0, которые должны распределить в течение m лет между двумя отраслями производства I и II. Средства, вложенные в каждую отрасль, приносят за год определенный доход, зависящий от объема вложений. Если вложим средства X в отрасль 1, то за год получим доход, равный $f(X)$, при этом вложенные средства частично уменьшатся так, что к концу года останется какая-то часть $f(X) < X$.

Аналогично средства Y, вложенные в отрасль 2, приносят за год доход $q(Y)$ и уменьшаются до $\psi(Y) < Y$. По истечению года, оставшиеся от K_0 средства заново распределяются между отраслями I и II. Новых средств извне не поступает, в производство вкладываются оставшиеся средства. Доход в производство не вкладывается, а накапливается отдельно. Требуется найти такой способ управления ресурсами (какие средства, в какие годы и в какую отрасль вкладывать), при котором суммарный доход от обеих отраслей за m лет будет максимальным. Решение осуществляется по стандартной схеме.

1. Система S – две отрасли, с вложенными в них средствами. Она характеризуется двумя параметрами X и Y, которые выражают количество средств, вложенных в отрасли 1 и 2 соответственно. Естественным этапом процесса является хозяйственный год. В процессе управления величины X и Y меняются в зависимости от двух причин:

- перераспределением средств между отраслями в начале каждого года;
- уменьшением средств за год.

Управлением на i- м шаге будет количество средств X_i и Y_i, вкладываемые в отрасли 1 и 2 на этом шаге. Требуется найти такое оптимальное управление, при котором суммарный доход, приносимый обеими отраслями за пять лет, был максимальным.

1. Состояние системы перед i- м шагом характеризуется одним параметром K-количеством средств, сохранившихся после предыдущих $i-1$ шагов. Управление на i- м шаге будет состоять в том, что мы выделим в отрасль 1 средств X_i. Тогда $Y_i = K - X_i$.

Выигрыш на i- м шаге равен:

$$w_i(K, X_i) = f(X_i) + q(K - X_i).$$

1. Под влиянием этого управления на i- м шаге система перейдет из состояния K в состояние

$$K = \varphi(X_i) + \psi(K - X_i).$$

2. Основное функциональное уравнение имеет вид:

$$W_i = \max_{0 \le X_i \le K} \{ f(X_i) + q(K - X_i) + W_{i+1}(\varphi(X_i) + \psi(k - X_i)) \}.$$

3. Условный оптимальный выигрыш на последнем шаге будет:

$$W_i = \max_{0 \le X_i \le K} \quad f \quad X_m \quad + q \quad K - X_m$$

Если соответствует условное оптимальное управление $X_m \quad K$, при котором этот максимум достигается.

4. Зная функцию $W_m \quad K$, находим из п.3 условные оптимальные выигрыши на двух последних, на трех последних и т.д. шагах:

$$W_{m-1} \quad K \quad = \max_{0 \le X_{m-1} \le K} \quad f \quad X_{m-1} \quad + q \quad K - X_{m-1} \quad + W_m \left[\varphi \quad X_{m-1} \quad + \varphi \quad K - X_{m-1} \quad \right],$$

$$W_{m-2} \quad K \quad = \max_{0 \le X_{m-2} \le K} \quad f \quad X_{m-2} \quad + q \quad K - X_{m-2} \quad + W_m \left[\varphi \quad X_{m-2} \quad + \varphi \quad K - X_{m-2} \quad \right],$$

$$W_1 \quad K \quad = \max_{0 \le X_1 \le K} \quad f \quad X_1 \quad + q \quad K - X_1 \quad + W_2 \left[\varphi \quad X_1 \quad + \varphi \quad K - X_2 \quad \right]$$

6. Начальное состояние K_0 задано, поэтому максимальный доход будет:

$W_{\max} = W_1 \quad K_0 \quad .$

Оптимальное уравнение на первом шаге:

$X_1 = X_1 \quad K_0 \quad .$

Состояние системы после первого шага:

$K_1^* = f \quad X_1 \quad + \psi \quad K - X_1 \quad .$

Оптимальное уравнение на втором шаге:

$X_2 = X_2 \quad K_1^*$ и т.д.

Состояние системы после i-х шагов:

$K_i^* = \varphi \quad X_i \quad + \psi \quad K_{i-1}^* - X_i \quad .$

Оптимальное уравнение на i-м шаге:

$X_i = X_i \quad K_{i-1}^* \quad .$

Далее оптимальное уравнение строится по цепочке:

$$K_0^* \to X_1 \quad K_0 \quad \to K_1^* \to X_2 \quad K_1^* \quad \to \ldots \to K_{m-1}^* \to X_m \quad K_{m-1}^* \quad \to K_m^* .$$

Величина K_m^* будет представлять собой количество средств, оставшихся (при оптимальном управлении) после последнего шага.

Совокупность средств, вложенных по годам в отрасль 1 $X = \quad X_1, X_2, \ldots, X_m$, будет представлять собой оптимальное управление.

Количество средств, вложенных в отрасль 2 по годам:

$$Y = Y_1, Y_2, \ldots, Y_m = \quad K_0 - X_1, K_1^* - X_2, \ldots, K_{m-1}^* - X_m$$

Глава 4. Исследование операций по схеме марковских случайных процессов

4.1. Марковский случайный процесс с дискретным состоянием и дискретным временем

До сих пор мы рассматривали детерминированные задачи исследования операций. Однако многие операции, которые приходятся анализировать, развиваются как случайные процессы, исход которых зависит от ряда случайных факторов. Для математического описания операций, развивающихся в форме случайного процесса, применяется математический аппарат, разработанный в теории вероятностей для так называемых марковских случайных процессов [7, 8, 9].

Случайный процесс, протекающий в системе, называется марковским случайным процессом, если он обладает следующим свойством: для каждого момента времени t_0 вероятность любого состояния системы в будущем (при $t > t_0$) зависит только от ее состояния в настоящем (при $t = t_0$) и не зависит от того, когда и каким образом система пришла в это состояние (т.е. как развивался процессов прошлом). Другими словами, в марковском случайном процессе будущее развитие зависит только от настоящего состояния и не зависит от предыстории процесса.

Случайный процесс называется с дискретным состоянием, если возможные состояния системы $S_1, S_2, S_3, \ldots$ можно перечислить (пронумеровать) одно за другим, а сам процесс состоит в том, что время от времени система S скачком перескакивает из одного состояния в другое.

Пример. Техническое средство может находиться в одном из пяти возможных состояний:

S_1 – работоспособна и исправна;

S_2 – неработоспособна и неисправна;

S_3 – оценка технического состояния;

S_4 – ремонтируется;

S_5 – списывается.

Процесс, протекающий в системе, состоит в том, что система случайным образом переходит из состояния в состояние. Переходы $S_1 \to S_2$, $S_2 \to S_3$и т.д., но не $S_1 \to S_3$.

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

Состояние системы изображается прямоугольниками (или кругами), а возможные переходы из состояния в состояние – стрелками, соединяющими в прямоугольники для рассматриваемого примера, граф состояния приведен на рис. 1.

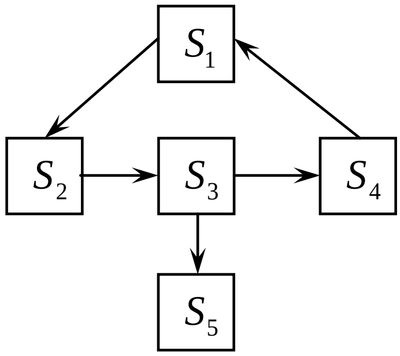


Рис.1. Граф состояний системы, включающий 5 состояний

Следующий пример. Система представляет собой техническое устройство, состоящее из двух узлов 1 и 2 (т.е. резервированное устройство), каждый из них может отказать в какое-то время. Отказавший узел начинается немедленно ремонтироваться.

Возможные состояния системы:

S_1 – оба узла работоспособны;

S_2 – первый узел неработоспособный (отказал) и восстанавливается, а второй работоспособный;

S_3 – первый узел работоспособный, а второй неработоспособный и восстанавливается;

S_4 – оба узла неработоспособны и восстанавливаются.

Граф состояний приведен на рис. 2.

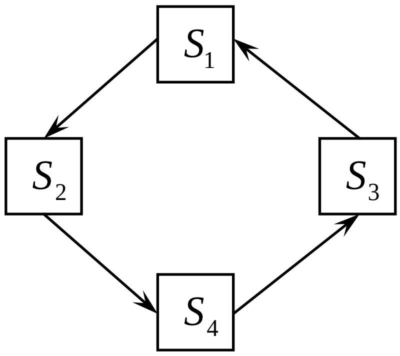


Рис.2. Граф состояний резервированного устройства

Случайный процесс называется процессом с дискретным временем, если переходы системы из состояния в состояние возможны только в строго определенные моменты времени $t_1, t_2, \ldots$. В промежутки времени между этими моментами система S сохраняет свое состояние.

Пусть имеется система S, которая может находиться в состояниях $S_1, S_2, \ldots, S_n$, причем переходы системы из состояния в состояние возможны только в моменты времени $t_1, t_2, \dots, t_k$. Будем называть эти моменты “шагами” или “этапами” процесса и рассматривать случайный процесс, происходящий в системе S, как функцию целочисленного аргумента $1, 2, \dots, k$ (номер шага).

Случайный процесс, происходящий в системе, состоит в том, что в последовательные моменты времени $t_1, t_2, t_3 \dots$ система S оказывается в тех или иных состояниях, например $S_1 \to S_2 \to S_3 \to S_4 \to S_1 \dots$.

В общем случае в моменты $t_1, t_2, \dots$ система может не только менять состояние, но оставаться в прежнем $S_1 \to S_2 \to S_3 \to S_3 \to S_2 \to S_3 \to S_4 \to S_1 \dots$. Предположим, что для каждого состояния известна вероятность перехода в любое другое состояние за один шаг. Обозначим P_{ij} вероятность перехода из состояния S_i в состояние S_j, P_i – вероятность задержки системы в состоянии S_i. Запишем переходные вероятности в P_{ij} в виде матрицы:

$$\left\| P_{ij} \right\| = \begin{Vmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1j} & \cdots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \cdots & P_{2j} & \cdots & P_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{i1} & P_{i2} & P_{i3} & \cdots & P_{ij} & \cdots & P_{in} \\ P_{n1} & P_{n2} & P_{n3} & \cdots & P_{nj} & \cdots & P_{nn} \end{Vmatrix}.$$

Некоторые из переходных вероятностей P_{ij} могут быть равны нулю. Это означает, что за один шаг переход системы из i-го состояния в j-е невозможен. По главной диагонали матрицы переходных вероятностей стоит вероятность P_{ij} того, что система не выйдет из состояния S_i, а останется в нем.

При рассмотрении марковских цепей удобно пользоваться графом состояний, на котором у стрелок проставлены соответствующие переходные вероятности. Такой граф назовем размеченным графом состояний [3, 7].

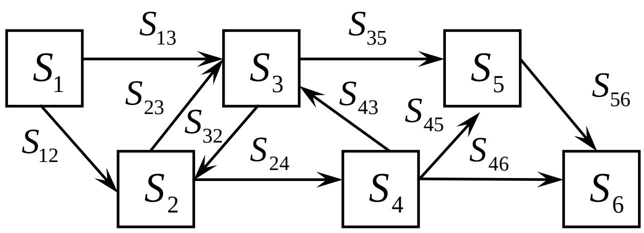


Рис.3. Размеченный граф состояний

На рис. 3 представлены только те переходные вероятности, которые не равны нулю и меняют состояние системы. Тогда $P_{11} = 1 - P_{12} + P_{13}$;

$P_{22} = 1 - \ P_{23} + P_{24}\ $; $P_{33} = 1 - \ P_{35} + P_{32}\ $; $P_{44} = 1 - \ P_{45} + P_{46}\ $; $P_{55} = 1 - P_{56}$; $P_{66} = 1 - P_{62}$.

Если из состояния S_i не исходит не одной стрелки (не возможен переход из него ни в какое другое состояние), соответствующая вероятность задержки P_{ij} равна 1.

Предположим, что в начальный момент (перед первым шагом) система находится в состоянии S_1. Тогда для начального момента ((0) момента) будем иметь $P_1\ 0\ = 1$, $P_2\ 0\ = 1$, $P_3\ 0\ = 0$, $P_4\ 0\ = 0$, $P_5\ 0\ = 0$, $P_6\ 0\ = 0$.

Вероятность состояния после первого шага, берутся из первой строки вероятности перехода $P_1\ 1\ = P_{11}$, $P_2\ 1\ = P_{12}$, $P_3\ 1\ = P_{13}$, $P_4\ 1\ = P_{14}$, $P_5\ 1\ = P_{15}$, $P_6\ 1\ = P_{16}$.

Найдем вероятностью состояний после 2-го, 3-го и так далее шагов.

Будем их вычислять по формулам полной вероятности. Опуская промежуточные выкладки, запишем вероятности после второго шага:

$$P_i\ 2\ = \sum_{j=1}^{6} P_j\ 1\ \cdot P_{ji}.$$

4.2. Случайный марковский процесс с дискретным состоянием и непрерывным временем

Случайный процесс называется процессом с непрерывным временем, если переход системы из состояния в состояние возможен в любой, наперед неизвестный случайным момент времени. Для определения вероятностей состояний необходимо знать характеристики процесса, аналогичные переходным вероятностям для процесса с дискретным состоянием и дискретным временем. Вместо переходных вероятностей P_{ij} вводится в рассмотрение плотность вероятности перехода λ_{ij}.

Назовем плотность вероятности перехода λ_{ij} предел отношения вероятности перехода системы за время Δt из состояния S_i в состояние S_j к длине промежутка Δt, т.е.

$$\lambda_{ij} = \lim_{\Delta t} \frac{P_{ij}\ \Delta t}{\Delta t}, \tag{32}$$

где $P_{ij}\ \Delta t$ – вероятность того, что система, находившая в момент времени t в состоянии S_i за время Δt перейдет из него в состояние S_j.

Из формулы (32) следует, что при малом Δt вероятность перехода

$$P_{ij}\ \Delta t\ \cong \lambda_{ij} \cdot \Delta t.$$

Предположим, что известны плотности вероятностей перехода λ_{ij} для всех пор состояний S_i, S_j. Построим граф состояний системы S и против каждой стрелки проставим соответствующую плотность вероятности перехода, т.е. получим размеченный граф.

Примером процесса с дискретным состоянием и непрерывным временем является техническое устройство S, состоящее из двух узлов, каждый из которых в случайный момент времени может перейти в неработоспособное состояние, после чего мгновенно начинается его ремонт.

Возможные состояния системы:

S_0 – оба узла работоспособны;

S_1 – первый узел ремонтируется, второй работоспособен;

S_2 – второй узел ремонтируется, первый работоспособен;

S_3 – оба узла ремонтируются.

Вероятность нахождения в состояниях S_0, S_1, S_2, S_3 обозначим через P_0, P_1, P_2, P_3 соответственно. Размеченный граф состояний представлен на рис. 4.

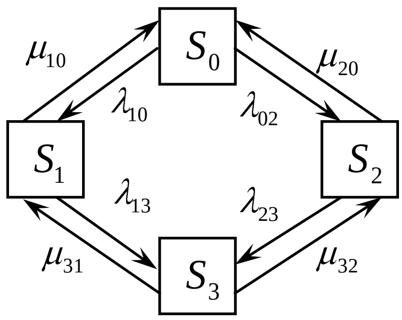


Рис.4. Граф состояний процесса с дискретным состоянием и непрерывным временем

Пусть система находится в состоянии S_0. Поток отказов первого узла λ_{01} переводит систему в состояние S_1. При этом поток отказов первого узла λ_{01} равен единице, деленной на среднее время безотказной работы первого узла, т.е. $\lambda = \dfrac{1}{T_0}$. Поток восстановления работоспособности узла μ_{10} равен единице, деленное на среднее время восстановления, т.е. $\mu_{10} = \dfrac{1}{T_B}$ переводит систему из S_1 в S_0. Вычисление интенсивности потоков событий, переводящих систему по всем стрелкам размеченного графа, проводится аналогично случаю, приведенному на рис. 1.

Имея размеренный граф состояний, можно найти все вероятности состояний $P_i\;t$ как функции времени. Для этого составляются и решаются уравнения Колмогорова – дифференциальные уравнения, в которых неизвестными функциями являются вероятности состояния. Эта система 4-х линейных дифференциальных уравнений с четырьмя неизвестными функциями P_0, P_1, P_2, P_3.

Сформулируем общее правило составления уравнений Колмогорова, в левой части каждого уравнения стоит производная вероятности какого-то (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, умноженную на вероятность данного состояния. Пользуясь этим правилом, запишем уравнения Колмогорова для рассматриваемого случая:

$$\begin{aligned}
\frac{dP_0}{dt} &= \mu_{10}P_1 + \mu_{20}P_2 - (\lambda_{01} + \lambda_{02})P_0,\\
\frac{dP_1}{dt} &= \mu_{01}P_0 + \mu_{31}P_3 - (\lambda_{13} + \mu_{10})P_1,\\
\frac{dP_2}{dt} &= \lambda_{02}P_0 + \mu_{32}P_3 - (\lambda_{23} + \mu_{20})P_2,\\
\frac{dP_3}{dt} &= \lambda_{13}P_1 + \lambda_{23}P_2 - (\mu_{31} + \mu_{32})P_3.
\end{aligned} \tag{33}$$

Для уравнений (33) необходимо задать начальные условия. Так, например, уравнения (33) решать при начальных условиях $P_0(0)=1$, $P_1(0)=P_2(0)=P_3(0)=0$, т.е. в начальный момент оба узла работоспособны. Как решаются подобные уравнения. Вообще говоря, линейные дифференциальные уравнения с постоянными коэффициентами можно решать аналитически, но это удобно только когда число уравнений не превосходят двух (или иногда трех). В других случаях их решают численно – вручную или на ЭВМ.

Рассмотрим, что будет происходить с вероятностями состояний при $t \to \infty$. Будут ли $P_i(t)$ $(i = 0,1,2,3)$ стремится к каким-то пределам.

Если эти пределы существуют и не зависят от начального состояния системы, то они называются финальными вероятностями состояний.

Предположим, что это условие выполнено и финальные вероятности существуют:

$$\lim_{t\to\infty} P_i(t) = P_i, \quad (i = 0,1,2,3).$$

Финальную вероятность состояния S можно истолковывать как среднее относительное время пребывания системы в этом состоянии. Если финальные вероятности $P_i(t)$ постоянны, то левые части уравнений Колмогорова будут равны нулю. Тогда получим систему линейных уравнений.

$$
\begin{aligned}
&\lambda_{01}+\lambda_{02}\;\; P_0=\mu_{10}P_1+\mu_{20}P_2,\\
&\lambda_{13}+\mu_{10}\;\; P_1=\lambda_{01}P_0+\mu_{31}P_3,\\
&\lambda_{23}+\mu_{20}\;\; P_2=\lambda_{02}P_1+\mu_{20}P_3,\\
&\mu_{31}+\mu_{32}\;\; P_3=\lambda_{13}P_1+\lambda_{22}P_2.
\end{aligned}
\tag{34}
$$

Одновременно необходимо учитывать нормировочное условие, т.е.

$$P_0+P_1+P_2+P_3=1. \tag{35}$$

1. Нерезервированные средства

Нерезервированные средства имеют следующие состояния:

S_0 – работоспособное средство; S_1 – не работоспособное средство (ремонтируется).

Размеченный граф состояний приведен на рис. 5.

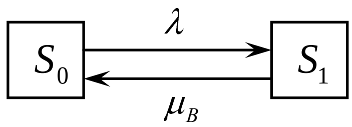


Рис. 5. Граф состояний для нерезервированной системы

Составим и решим алгебраические уравнения для финальных вероятностей:

$$P_0\cdot\lambda=P_1\cdot\mu_B \text{ или } P_1=P_0\cdot\frac{\lambda}{\mu_B},$$

где P_0 – вероятность нахождения средства в состоянии S_0; P_1 – вероятность нахождения средства в состоянии S_1; λ – интенсивность отказов, $\lambda=\dfrac{1}{T_0}$; μ_B – интенсивность восстановления, $\mu_B=\dfrac{1}{T_B}$. Нормировочное уравнение: $P_0+P_1=1$ или $P_0\cdot\dfrac{\lambda}{\mu_B}=P_0\left(1+\dfrac{\lambda}{\mu_B}\right)=1$.

Тогда $P_0=\dfrac{1}{1+\dfrac{\lambda}{\mu_B}}=\dfrac{T_0}{T_0+T_B}=K_Г$, $P_1=1-P_0=1-\dfrac{T_0}{T_0+T_B}=\dfrac{T_B}{T_0+T_B}$.

Таким образом, финальные вероятности определены. Физический смысл $K_Г$ – вероятность пребывания РЭС в работоспособном состоянии.

2. Горячее резервирование средств

Система имеет следующие состояния:

S_0 – оба комплекта работоспособны;

S_1 – первый комплект ремонтируется, второй работоспособен;

S_2 – второй комплект ремонтируется, первый работоспособен;

S_3 – оба комплекта ремонтируются.

Граф состояний представлен на рис. 6.

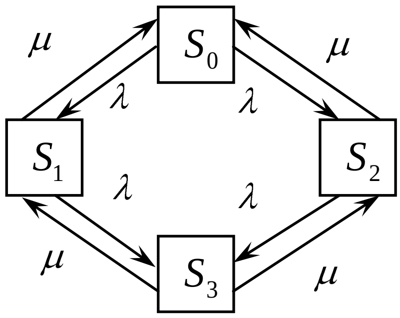


Рис.6. Граф состояний системы, находящейся в горячем резерве

Алгебраические уравнения Колмогорова для рассматриваемого случая записываются в виде:

(1) $2\lambda P_0 = \mu \;\; P_1 + P_2 \;\;$,

(2) $\;\lambda + \mu \;\; P_1 = \lambda P_1 + \mu P_3$, (36)

(3) $\;\lambda + \mu \;\; P_2 = \lambda P_0 + \mu P_3$,

(4) $2\mu P_3 = \lambda \;\; P_1 + P_2 \;\;$.

Из уравнений (2) и (3) системы (36) имеем, что $P_1 = P_2$. Тогда уравнение (1) записывается в виде:

$P_0 \lambda = P_1 \mu$ или $P_1 = \dfrac{\lambda}{\mu} P_0$.

Из уравнения (4) получим:

$\mu P_3 = \lambda P_1$, $P_3 = \dfrac{\lambda}{\mu} P_1 = \dfrac{\lambda^2}{\mu^2} P_0$.

Полученные выражения для P_1, P_2 и P_3 подставим в нормированное уравнение. Тогда

$$P_0 + 2P_0 \frac{\lambda}{\mu} + P_0 \frac{\lambda^2}{\mu^2} = 1 \text{ или } P_0 \left(1 + 2P_0 \frac{\lambda}{\mu} + \frac{\lambda^2}{\mu^2} \right) = 1 \text{ или } P_0 = \frac{1}{\left(1 + \dfrac{\lambda}{\mu} \right)^2} = K_Г^2 .$$

Таким образом, мы определили все вероятности состояний P_0, P_1, P_2, P_3.

Анализ показал, что вероятность P_0 при использовании горячего резервирования системы меньше, чем для нерезервированной системы.

Следовательно, T_0 для нерезервированной системы больше, чем для резервированной системы.

3. Холодное резервирование с переключением

Резервированные средства имеют следующие состояния:

S_0 – оба полукомплекта работоспособны;

S_1 – первый комплект работоспособен, второй ремонтируется;

S_2 – второй комплект работоспособен, а первый ремонтируется;

S_3 – оба комплекта ремонтируются.

Граф состояний приведен на рис. 7.

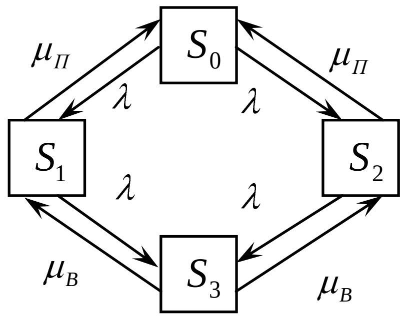


Рис.7. Граф состояний системы, находящейся в холодном резерве

Для рассматриваемого случая линейные алгебраические уравнения Колмогорова записываются в виде:

(1) $2\lambda P_0 = \mu_B \;\; P_1 + P_2 \;\;$,

(2) $\;\; \lambda + \mu_B \;\; P_1 = \lambda P_0 + \mu_П P_3,$ (37)

(3) $\;\; \lambda + \mu_B \;\; P_2 = \lambda P_0 + \mu_П P_3,$

(4) $2\mu_П P_3 = \lambda \;\; P_1 + P_2 \;\;$,

где λ и μ_B – интенсивности отказа и восстановления; $\mu_П$ – интенсивность переключения.

Нормировочное уравнение $P_0 + P_1 + P_2 + P_3 = 1$. Из уравнений (2) и (3) системы (37) видно, что $P_1 = P_2$.

Тогда уравнение (1) записывается в виде:

$P_0\lambda = \mu_B P_1$ или $P_1 = \dfrac{\lambda}{\mu_B} P_0$.

Уравнение (4) имеет вид:

$$P_3 = \frac{\lambda}{\mu_П} P_1 = \frac{\lambda^2}{\mu_П \mu_B} P_0.$$

Перепишем нормированное уравнение в виде

$$P_0\left(1+\frac{\lambda}{\mu_П}+\frac{\lambda}{\mu_В}+\frac{\lambda^2}{\mu_П\mu_В}\right)=1.$$

Откуда $P_0=\dfrac{T_0^2}{T_0^2+2T_ПT_0+T_ПT_В}$.

Так как $2T_ПT_0>T_ПT_В$, то получим

$$P_0=\frac{T_0}{T_0+2T_П},\ P_1=\frac{T_П}{T_0}P_0,\ P_2=\frac{T_П}{T_0}P_0,\ P_3=\frac{T_ПT_В}{T_0}P_0.$$

Таким образом, мы определили все вероятности состояний P_0,P_1,P_2,P_3.

Среднее время безотказной работы резервированной системы

$$T_0'=\frac{P_0T_В}{1-P_0}.$$

Упрощенно T_0' увеличивается по сравнению T_0 средства во столько раз, во сколько $T_В$ больше $2T_П$.

4.3. Потоки событий

Простейший поток и его свойства. При рассмотрении случайных процессов, протекающих в системах с дискретными состояниями, часто приходится встречаться с так называемыми потоками событий.

Потоком событий называется последовательность одноразовых событий, следующих одно за другим в какие-то, в общем случае, случайные моменты времени [1, 3]. Основными свойствами потоков событий являются стационарность, регулярность, ординарность и отсутствие последействия.

Стационарность потока означает, что его вероятные характеристики не меняются во времени. Регулярность потока означает, что события следуют одно за другим через равные промежутки времени. Ординарность означает, что в определенный момент времени появляется одно событие, а не группами по несколько раз. Отсутствие последействия в потоке означает что события, образующие потоки, появляются независимо друг от друга и не зависят от того, сколько их поступило в предыдущий момент времени.

Поток событий называется простейшим, если он обладает сразу тремя свойствами: стационарен, ординарен и не имеет последействия. Название “простейший” связано с тем, что процессы имеют наиболее простое математическое описание. Для простейшего потока с интенсивностью λ интервал T между соседними событиями имеет показательное (экспоненциальное) распределение с плотностью $f\ t\ =\lambda\cdot e^{-\lambda t}$.

График изменения плотности распределения представлен на рис. 8.

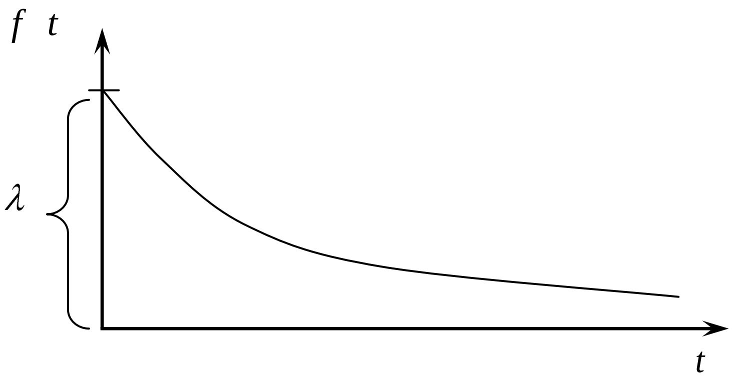


Рис.8. Плотность распределения простейшего потока

Для случайной величины T математическое ожидание $M_T = \frac{1}{\lambda}$, дисперсия $D_T = \frac{1}{\lambda^2}$, среднеквадратическое отклонение $\sigma_T = \sqrt{D_T}\frac{1}{\lambda}$, коэффициент вариации (мера случайности) $V_T = \frac{\sigma_T}{M_T}$.

Таким образом, для простейшего потока коэффициент вариации интервалов между событиями равен 1.

Потоки Эрланга. Поток событий называется потоком Пальма (иначе рекуррентный поток) если он стационарен, ординарен, а интервалы времени между событиями T_1, T_2 представляют собой независимые случайные величины с одинаково произвольным распределением. Приведем пример потока Пальма.

Некоторый элемент радиосистемы (например, транзистор) работает непрерывно до своего отказа. Отказавший элемент мгновенно заменяется новым. Если отдельные экземпляры элементов выходят из строя независимо друг от друга, то поток отказов представляет собой поток Пальма.

Важным для практики образования потоков Пальма являются так называемые потоки Эрланга. Эти потоки образуются в результате “просеивания простейших” потоков.

Рассмотрим на оси $0t$ простейший поток и сохраним в нем не все точки, а только каждую вторую (рис. 9).

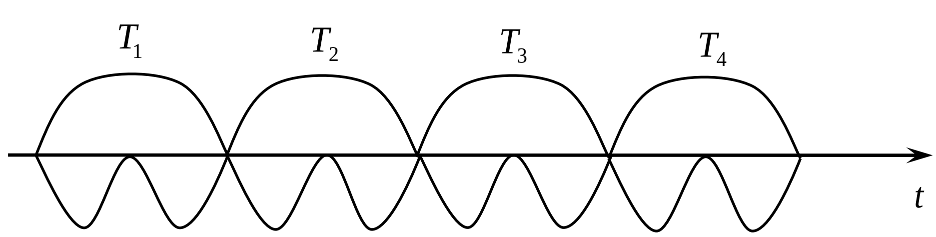


Рис. 9. Иллюстрация простейшего потока событий

В результате такой операции “прореживание” или “просеивание” образуют снова поток событий, он называется потоком Эрланга. Вообще потоком Эрланга k-го порядка $Э_k$ называется поток, полученный из простейшего, если в последнем сохранить каждую k-ю точку, а остальные выбросить. Например, потоком Эрланга 4-го порядка.

Простейший поток. Очевидно, что простейший поток представляет собой поток Эрланга 1-го порядка. Интервал времени T между соседними событиями в потоке Эрланга k-го порядка представляет собой сумму к независимых случайных величин простейшего потока:

$$T = T_1 + T_2 + T_3 + \ldots + T_k = \sum_{i=1}^{k} T_i .$$

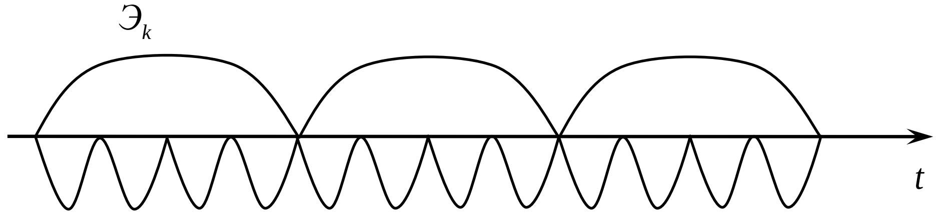


Рис. 10. Поток Эрланга k-го порядка

Каждая из этих случайных величин $T_1, T_2, ..., T_k$ распределена по экспоненциальному закону

$$f_1(t) = \lambda e^{-\lambda t} (t > 0) .$$

Закон распределения интервала T между соседними событиями в потоке называется законом Эрланга k-го порядка, имеющий вид

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

Очевидно, при $k = 1$ получаем экспоненциальное распределение $f_1(t) = \lambda e^{-\lambda t}$.

Параметры закона распределения Эрланга:

$m_t^{(k)} = \dfrac{k}{\lambda}$ – математическое ожидание; $D_t^{(k)} = \dfrac{k}{\lambda^2}$ – дисперсия; $\sigma_t = \dfrac{\sqrt{k}}{\lambda}$ - среднеквадратическое отклонение.

Заметим, что как закон распределения $f_k(t)$, так и все его характеристики выражены не через интенсивность самого потока Эрланга, а через интенсивность λ простейшего потока, который подвергается прореживанию.

Выразим их через поток Эрланга. Обозначим λ_k интенсивность самого потока Эрланга. Тогда

$$\lambda_k = \frac{\lambda}{k!}.$$

Так как из исходного простейшего потока с интенсивностью λ берется только k-я часть.

Подставив выражение λ через λ_k, получим

$$f_k(t) = \frac{(k\lambda_k)^k}{(k-1)!} t^{k-1} e^{-k\lambda_k t}.$$

Соответственно, $m_t^{(k)} = \dfrac{1}{\lambda_k}$, $D_t^{(k)} = \dfrac{1}{k\lambda_k^2}$, $\sigma_t^{(k)} = \dfrac{1}{\sqrt{k}\lambda_k}$.

Пусть $\lambda_k = \Lambda = const$ и меняем только порядок закона Эрланга. Математическое ожидание остается постоянным $m_t = \dfrac{1}{\Lambda}$, а дисперсия и среднеквадратическое отклонение будут меняться

$$D_t^{(k)} = \frac{1}{k\Lambda^2},\ \sigma_t^{(k)} = \frac{1}{\sqrt{k}\Lambda} 0,3.$$

Следовательно, при $k \to \infty$ D_t^k и σ_t^k стремятся к нулю, т.е. поток Эрланга приближается к регулярному потоку с постоянным интервалом между событиями:

$$T = \frac{1}{\Lambda} = const.$$

Это свойство потоков Эрланга удобно для практического использования. График плотности распределения представлен на рис. 11.

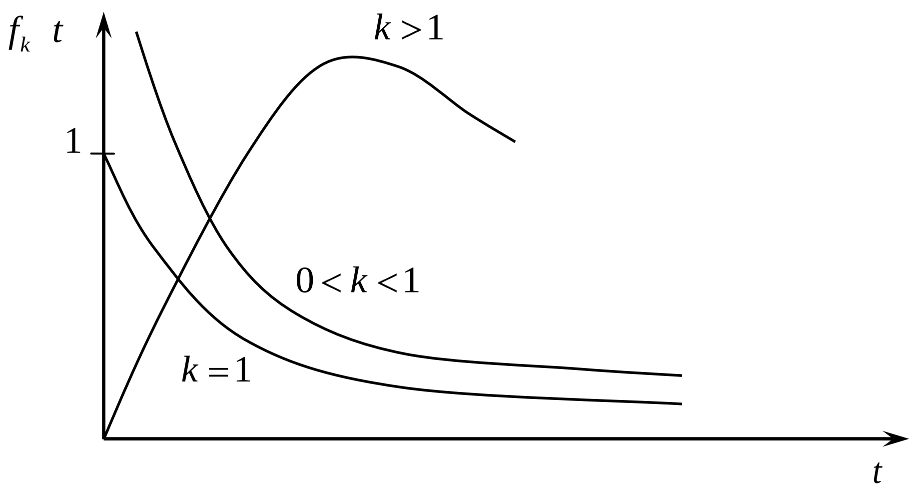


Рис.11. Плотность распределения для потока Эрланга

Оно дает возможность, задаваясь различными k, получать потоки, обладающие различным последействием – от полного отсутствия последействия $k = 1$ до жесткой функциональной связи между моментами появление событий $k = \infty$. Таким образом, порядок потока Эрланга k может служить “мерой последействия”.

Пример. В результате статической обработки интервалов времени между событиями в некотором потоке получены следующие характеристики:

- среднее значение интервала $m_t = 2$ мин.

- среднее квадратическое отклонение интервала $\sigma_t = 0,9$.

Требуется подобрать поток Эрланга, обладающий приблизительно теми же характеристиками, найти его интенсивность λ и порядок k.

Решение.

$$\lambda = \frac{1}{m_t} = 0,5 \ \textit{соб./мин},$$

$$k = \left(\frac{1}{\sigma_t \lambda}\right)^2 = \left(\frac{1}{0,9 \cdot 0,5}\right)^2 = 4,9.$$

Ближайшее целое число 5, т.е. данный поток можно приблизительно заменить потоком Эрланга 5-го порядка с плотностью

$$f_S(t) = \frac{(5 \cdot 0,5)^5}{4!} t^4 e^{-0,5 \cdot 5t} = 4,1 t^4 \cdot e^{-2,5t}.$$

График $f(t)$ представлен на рис. 12.

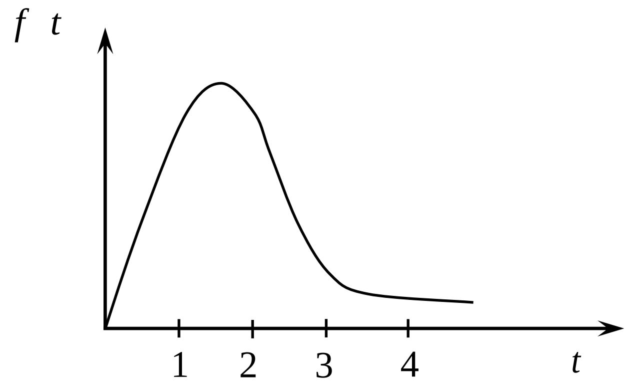


Рис. 12. График плотности распределения потока Эрланга 5-го порядка

Глава 5. Теория массового обслуживания
5.1. Классификация систем массового обслуживания

При исследовании операций часто приходится сталкиваться с работой систем, называемых системой массового обслуживания (СМО) [3, 8]. Примерами таких систем являются телефонные станции, ремонтные мастерские, включая билетные кассы, магазины и так далее, т.е. там, где наблюдаются очереди на обслуживание.

Каждая СМО состоит из какого-то числа каналов обслуживаний, в качестве которых могут быть линии связи, рабочие точки, кассиры и т.д. СМО могут быть одноканальными и многоканальными.

Всякая СМО предназначена для обслуживания потока заявок, поступающих в случайные моменты времени. Обслуживание заявки продолжается тоже случайное время $T_{обс}$. После обслуживания канал освобождается и готов к приему следующей заявки. Случайный характер потока заявок и времени обслуживания может привести к тому, что в какие-то периоды времени на входе СМО скапливается излишне большое количество заявок, поэтому СМО разделяется на СМО с отказами и СМО с очередью.

В СМО с отказами заявка, поступившая в момент, когда все каналы заняты, получает отказ и покидает СМО. Пример телефония. В СМО с очередью заявка, пришедшая в момент, когда все каналы заняты, не уходит, а становится в очередь и ожидает возможность быть обслуживаемой.

СМО могут быть однородными, т.е. состоящими из одинаковых устройств и неоднородными (когда заявка обслуживается на одном этапе или фазе, и многофазными, когда заявки, обслуживают в начале одной, а затем в другой фазах).

Примеры. Однофазная СМО – продавец помогает выбрать товар, работает на кассе и упаковывает. Покупатель только оплачивает. Многофазная СМО – продавец помогает выбирать товар, кассир выбивает чек, покупатель оплачивает товар и получает его на контроле.

Кроме этих признаков, СМО делятся на два класса – открытые и замкнутые. В открытой СМО характеристики потока заявок не зависит от того, в каком состоянии находится СМО (сколько каналов заняты). В замкнутой – зависит.

Важнейшим критериями СМО являются:

- вероятность пропуска (задержки в обслуживании) заявки;
- математическое ожидание числа пропущенных (задержанных) заявок за фиксированное время;
- математическое ожидание числа занятых каналов;
- математическое ожидание длины очереди.

Процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем. Потоки событий в СМО включают в себя поток заявок и время обслуживания.

Поток заявок на обслуживание рассматривается во времени. Он может быть стационарным, когда его характеристики во времени не меняются, и нестационарным, когда его характеристики меняются во времени. Он может быть без последействия, когда число поступающих заявок не зависит от того, сколько их поступило в предыдущий момент времени, и с последействием, когда число поступающих в данный момент времени заявок зависит от того сколько их поступило в предыдущий момент времени.

Наконец, потоки могут быть ординарные, когда в один момент времени не может поступить несколько заявок и неординарные.

5.2. Формула Литтла

Получим формулу, связывающую среднее число заявок, находящихся в системе массового обслуживания, т.е. обслуживаемых или стоящих в очереди, $L_{сист}$ и среднее время пребывания заявки в системе $W_{сист}$ обозначим:

$x\ t$ – число заявок, пребывающих в СМО до момента t;

$y\ t$ – число заявок, покинувших СМО до момента t.

И та и другая функция являются случайными и меняются скачком в моменты прихода заявок $x\ t$ и уходов заявок $x\ t$. Вид функции $x\ t$ и $y\ t$ показаны на рис. 13.

Очевидно, что для любого момента времени t их разность $z\ t = x\ t - y\ t$ есть не что иное, как число заявок, находящихся в СМО. Когда линии $x\ t$ и $y\ t$ сливаются, в системе нет заявок.

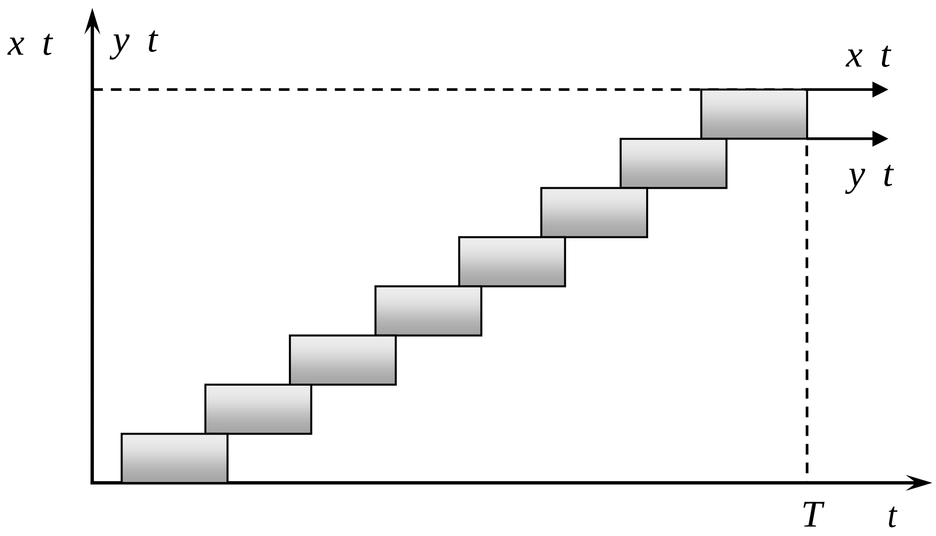


Рис.13. Функции числа заявок, прибывающих и покинувших СМО

Рассмотрим очень большой промежуток времени T и вычислим число заявок, находящихся в СМО. Оно будет равно интегралу от функции $z\ t$ на этом промежутке, деленному на длину интервала T:

$$L_{сист} = \frac{1}{T}\int_0^T z\ t\ dt.$$

Но этот интеграл представляет собой не что иное, как площадь заштрихованной фигуры (рис. 13). Можно считать, что

$$\int_0^T z\ t\ dt = \sum_i t_i,$$

где t_i – время пребывания соответствующей заявки в системе. Тогда

$$L_{сист} = \frac{1}{T}\sum_i t_i \, . \tag{38}$$

Разделим и умножим правую часть уравнения (38) на λ. Тогда $L_{сист} = \frac{1}{T\lambda}\sum_i t_i \lambda_i$.

Величина $T\lambda$ – среднее число заявок, прошедших за время T. Если разделим сумму всех времен t_i на среднее число заявок, то получим среднее время пребывания заявки в системе $-W_{сист}$. Итак, $L_{сист} = \lambda W_{сист}$, откуда $W_{сист} = \frac{1}{\lambda}L_{сист}$, т.е. среднее время пребывания заявки в системе равно среднему числу заявок в системе, деленному на интенсивность потока заявок.

Аналогично выводится вторая формула Литтла, связывающая среднее время пребывания заявки в очереди $W_{оч}$ и среднее число заявок в очереди $L_{оч}$:

$$W_{оч} = \frac{1}{\lambda}L_{оч}.$$

5.3. Схема гибели и размножения

Граф состояний для схемы гибели и размножения имеет вид, показанный на рис. 14 [1, 5]. Особенности этого графа в том, что все состояния системы можно вытянуть в одну цепочку, в которой каждое из средних состояний $S_1, S_2, \ldots, S_{n-1}$ связано с прямой и обратной стрелками с каждым из соседних состояний, а крайние состояния S_0, S_k – только с одним соседним состоянием. Пользуясь графом состояний, составим и решим алгебраические уравнения для финальных вероятностей.

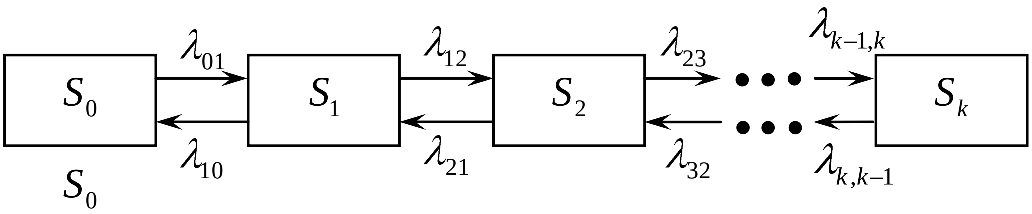


Рис.14. Граф состояний схемы гибели и размножения

Для первого состояния имеем

$$\lambda_{01}P_0 = \lambda_{10}P_1 \, . \tag{39}$$

Для второго состояния $\lambda_{12} + \lambda_{10} \;\; P_1 = \lambda_{01}P_0 + \lambda_{21}P_1$

$$\lambda_{12}P_1 = \lambda_{21}P_2 \, . \tag{40}$$

Далее совершенно аналогично $\lambda_{23}P_2 = \lambda_{32}P_3$ и вообще

$$\lambda_{k-1,k}P_{k-1} = \lambda_{k,k-1}P_k \, . \tag{41}$$

Кроме того, надо учесть нормировочное условие

$P_0 + P_1 + P_2 + \ldots + P_n = 1$.

Решим эту систему уравнений из уравнения (39):

$$P_1 = \frac{\lambda_{01}}{\lambda_{10}} P_0.$$

Из уравнения (40):

$$P_2 = \frac{\lambda_{12}}{\lambda_{21}} P_1 = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} P_0.$$

Из уравнения (41):

$$P_2 = \frac{\lambda_{23}\lambda_{12}\lambda_{01}}{\lambda_{32}\lambda_{21}\lambda_{10}} P_0$$

И вообще для любого k:

$$P_k = \frac{\lambda_{k-2,k} \cdots \lambda_{12}\lambda_{01}}{\lambda_{k,k-1} \cdots \lambda_{21}\lambda_{10}} P_0.$$

Таким образом, вероятности $P_1, P_2, \ldots, P_k$ выражены через P_0. Подставим выражения в нормировочное условие. Получим

$$P_0 \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \cdots + \frac{\lambda_{k-2,k} \cdots \lambda_{12}\lambda_{01}}{\lambda_{k,k-1} \cdots \lambda_{21}\lambda_{10}} \right) = 1.$$

Тогда $P_0 = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \cdots + \frac{\lambda_{k-2,k} \cdots \lambda_{12}\lambda_{01}}{\lambda_{k,k-1} \cdots \lambda_{21}\lambda_{10}} \right)^{-1}$.

Таким образом, все финальные вероятности определены.

5.4. n-канальная СМО с отказами (задача Эрланга)

Эта задача возникла из практических нужд телефонии. Однако с ее помощью можно определить количество рабочих мест для ремонта технических средств в АТБ и ЭРТОС. Задача ставится так: имеется k каналов, на которые поступает поток заявок с интенсивностью λ. Поток обслуживаний имеет интенсивность $\mu = \frac{1}{T_{обс}}$, где $T_{обс}$ – среднее время обслуживания. Определить финальные вероятности состояний СМО, а также характеристики её эффективности:

A – абсолютную пропускную способность, т.е. среднее число заявок, обслуживаемых в единицу времени;

Q – относительную пропускную способность, т.е. среднюю долю прошедших заявок, обслуживаемых системой;

$P_{отк}$ – вероятность отказа того, что заявка покинет СМО необслуживаемой;

k – среднее число занятых каналов.

Решение. Состояния системы S (СМО) будем нумеровать по числу заявок, находящихся в системе:

S_0 – в СМО нет ни одной заявки;

S_1 – в СМО находится одна заявка (один канал занят);

S_k – в СМО находится k заявок (все k каналов заняты).

Граф состояний СМО соответствует схеме гибели и размножения. Размеченный граф состояний представлен на рис. 15.

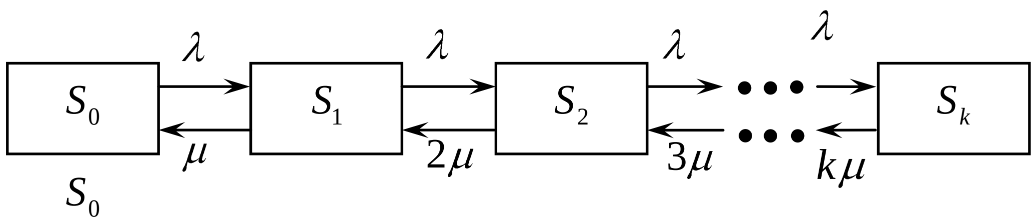


Рис.15. Граф состояний СМО

Из рис. 15 видно, что поток заявок с интенсивностью λ (верхние стрелки) переводит систему из любого левого состояния в соседнее правое состояние.

Проставим интенсивности у нижних стрелок. Пусть система находится в состоянии S_1 (работает один канал). Он производит μ обслуживаний в единицу времени. Проставляем у стрелки $S_1 \to S_0$ интенсивность μ. Пусть система находится в состоянии S_2 (работают два канала). Чтобы ей перейти в S_1 нужно, чтобы закончил обслуживание либо первый канал, либо второй суммарная интенсивность их потоков обслуживаний равна 2μ. Суммарный поток обслуживаний k каналами – $k\mu$.

Зная все интенсивности, воспользуемся формулами схемы гибели и размножения и получим $P_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2!\mu^2} + \frac{\lambda^3}{3!\mu^3} + \cdots + \frac{\lambda^k}{k!\mu^k}\right)^{-1}$.

Обозначим $\frac{\lambda}{\mu} = \rho$, тогда

$$P_0 = \left(1 + \rho + \frac{\rho^2}{2!} + \frac{\rho^3}{3!} + \cdots + \frac{\rho^k}{k!}\right)^{-1},$$

$$P_1 = \rho P_0,\ P_2 = \frac{\rho^2}{2!}P_0,\ P_3 = \frac{\rho^3}{3!}P_0,\ P_k = \frac{\rho^k}{k!}P_0.$$

Эти формулы для финальных вероятностей называются формулами Эрланга. По ним определяются характеристики эффективности СМО. Сначала найдем $P_{отк}$ – вероятность того, что заявка получит отказ. Для этого необходимо, чтобы все k каналов были заняты:

$$P_{отк} = \frac{\rho^k}{n!} P_0 .$$

Отсюда находим относительную пропускную способность – вероятность того, что заявка будет обслужена:

$$Q = 1 - P_{отк} = 1 - \frac{\rho^k}{k!} P_0 .$$

Абсолютная пропускная способность

$$\Lambda = \lambda Q = \lambda \left(1 - \frac{\rho^k}{k!} P_0 \right).$$

Среднее число занятых каналов

$$k = \frac{\Lambda}{\mu} = \rho \left(1 + \frac{\rho^k}{k!} P_0 \right).$$

Таким образом, определены все характеристики эффективности СМО.

5.5. Одноканальная СМО с неограниченной очередью

На практике довольно часто встречаются одноканальные СМО с очередью (врач, обслуживающий пациентов; продавец, обслуживающий покупателей; специалист по ТО и ремонту в АТБ и ЭРТОС).

Пусть имеется одноканальная СМО с очередью, на которую поступает поток заявок с интенсивностью λ, поток обслуживаний имеет интенсивность μ. Требуется найти финальные вероятности, а также характеристики её эффективности:

$L_{сист}$ – среднее число заявок в системе; $W_{сист}$ – среднее время пребывания заявки в системе; $L_{оч}$ – среднее число заявок в очереди; $W_{оч}$ – среднее время пребывания заявки в очереди; $P_{зан}$ – вероятность того, что канал занят.

Состояние системы, как и раньше, будем нумеровать по числу заявок в СМО:

S_0 – канал свободен, S_1 – канал занят, очереди нет, S_2 – канал занят, одна заявка стоит в очереди, S_k – канал занят, $k-1$ заявок стоят в очереди.

График состояний имеет вид, показанный на рис. 16.

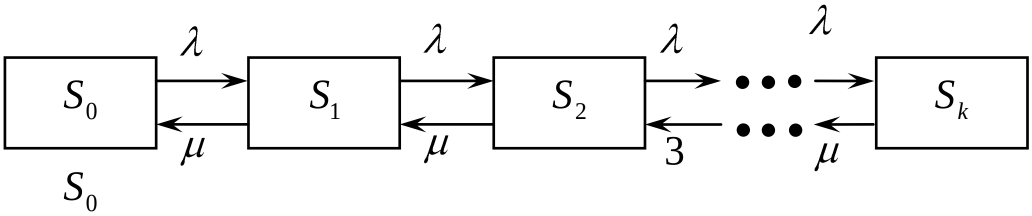


Рис.16. График состояний одноканальной СМО с очередью

Из рис. 16 видно, что эта схема однотипна схеме гибели и размножения. Поэтому будем использовать выражения для финальных вероятностей, но с условием, что число состояний будет бесконечным:

$$P_0 = \left[1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \ldots + \left(\frac{\lambda}{\mu}\right)^k + \ldots\right]^{-1} = \left[1 + \rho + \rho^2 + \ldots + \rho^k + \ldots\right]^{-1}. \quad (42)$$

Ряд в формуле (42) представляет собой геометрическую прогрессию, который сходится при $\rho < 1$. Суммируя прогрессию, в (42) имеем

$$1 + \rho + \rho^2 + \ldots + \rho^k + \ldots = \frac{1}{1-\rho}.$$

Тогда $P_0 = 1 - \rho$, $P_1 = \rho \cdot \rho_0$, $P_2 = \rho^2 \cdot \rho \cdot \ldots$. Или $P_1 = \rho \ 1-\rho$, $P_2 = \rho^2 \ 1-\rho$, … $P_k = \rho^k \ 1-\rho$.

Теперь определим характеристики эффективности. Отпуская промежуточные выкладки, запишем выражение для определения среднего числа заявок в системе

$$L_{сист} = \frac{\rho}{1-\rho}.$$

Используя формулу Литтла, получим среднее время пребывания заявки в системе:

$$W_{сист} = \frac{\rho}{\lambda \ 1-\rho}.$$

Вероятность того, что канал занят:

$P_{зан} = 1 - P_0 = \rho$.

Предположим, что это условие выполняется. Тогда финальные вероятности определяются выражениями:

$$P_0 = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \ldots + \frac{\rho^k}{k!} + \ldots \frac{\rho^{n+1}}{n! \ n-\rho}\right]^{-1},$$

$$P_1 = \frac{\rho}{1!} P_0, \ \ldots, \ P_k = \frac{\rho^k}{k!} P_0, \ P_n = \frac{\rho^n}{n!}, \ P_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} P_0, \ P_{n+r} = \frac{\rho^{n+r}}{n^r \cdot n!} P_0.$$

Теперь определяем характеристики эффективности СМО.

Среднее число занятых каналов

$$k = \frac{\lambda}{\mu} = \rho.$$

Опуская промежуточные выкладки, получим выражение для среднего числа заявок в очереди

$$L_{оч} = \frac{\rho^{n+1} P_0}{n \cdot n! \left(1 - \frac{\rho}{n}\right)}.$$

Прибавляя к $L_{оч}$ среднее число заявок под обслуживанием (оно же – среднее число занятых каналов), получим выражение для среднего числа заявок в системе:

$$L_{сист} = L_{оч} + \rho .$$

Средние времена пребывания заявки в очереди и в системе получаются делениями $L_{оч}$ и $L_{сист}$ на λ, то есть

$$W_{ист} = \frac{L_{сист}}{\lambda},\ W_{оч} = \frac{L_{оч}}{\lambda} .$$

Среднее число заявок под обслуживанием $L_{об} = \rho$. Отсюда $L_{оч} = L_{сист} - \rho = \dfrac{\rho}{1-\rho} - \rho = \dfrac{\rho^2}{1-\rho}$.

Среднее время пребывания заявки в очереди:

$$W_{оч} = \frac{\rho^2}{1-\rho} .$$

Таким образом, все характеристики СМО определены.

Литература

1. Вентцель Е.С. Исследование операций. Задачи, принципы, методология. - М.: КноРус, 2010.
2. Вентцель Е.С. Исследование операций. - М.: Высшая школа, 2010.
3. Таха Х.А. Введение в исследование операций. - М.: Вильямс, 2005.
4. Вагнер Г. Основы исследования операций. - М.:Мир, 1972.
5. Зайченко Ю.П. Исследование операций. - Киев: Вища школа, 1979.
6. Лунгу К.Н. Линейное программирование. Руководство к решению задач. - М.: Физматлит, 2005.
7. Тихонов В.И., Миронов М.А. Марковские процессы. - М.: Сов. Радио, 1977.
8. Боровков А.А. Вероятностные процессы в теории массового обслуживания. - М.: Наука, 1972.
9. Дынкин Е.Б., Юшкевич А.А. Теоремы и задачи о процессах Маркова. - М.: Наука, 1967.
10. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования.- М.: Наука, 1965.