Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Кемеровский государственный университет»
Новокузнецкий институт (филиал)

Факультет информатики, математики и экономики
Кафедра математики, физики и математического моделирования

ДИСКРЕТНЫЕ И ВЕРОТНОСТНЫЕ МАТЕМАТИЧЕСКИЕ МОДЕЛИ

(Вероятностные математические модели)

Методические указания к организации самостоятельной работы студентов для обучающихся по направлению подготовки 01.04.02 Прикладная математика и информатика, профиль «Математическое моделирование»

Новокузнецк

2020

УДК [378.147.88:004.42](072)
ББК 74.484(2Рос-4Кем)я73+32.973я73
Р47

Р47 Дискретные и вероятностные математические модели (вероятностные математические модели): методические указания к организации аудиторной и самостоятельной работы студентов факультета информатики, математики и экономики, обучающихся по направлению подготовки 01.04.02 Прикладная математика и информатика, профиль «Математическое моделирование» / Новокузнецкий ин-т (фил.) Кемеров. гос. ун-та. – Новокузнецк : НФИ КемГУ, 2021 – 97 с.

Методические указания содержат краткие теоретические сведения и задания для самостоятельной работы; список основной и дополнительной литературы.
Методические указания предназначены для наиболее рациональной организации аудиторной и внеаудиторной самостоятельной работы студентов.

Рекомендовано на заседании
кафедры математики, физики и математического моделирования
Протокол № 3 от 22 октября 2020г.
Заведующий кафедрой

_______________ / Е.В. Решетникова

УДК [378.147.88:004.42](072)
ББК 74.484(2Рос-4Кем)я73+32.973я73
Р47

Текст представлен в авторской редакции

ОГЛАВЛЕНИЕ

Введение 5
1 Метод статистических испытаний 10
1.1 Идея метода 10
1.2 Единичный жребий 11
1.3 Разыгрывание случайных событий и величин 13
1.4 Работа по теме «Метод статистических испытаний» 16
2 Линейные статистические модели 18
2.1 Выборочное уравнение прямой линии регрессии 18
2.2 Работа по теме «Линейные статистические модели» 23
3 Модели систем массового обслуживания 27
3.1 Марковский случайный процесс 27
3.2 Уравнения Колмогорова для вероятностей состояний 32
3.3 Классификация СМО и типов управления 37
3.4 Многоканальная СМО с отказами 39
3.5 Одноканальная СМО с ожиданием и неограниченной очередью 44
3.6 Одноканальная СМО с ожиданием и ограниченной очередью 50
3.7 Работа по теме «Модели СМО» 56
4 Модели управления запасами 61
4.1 Исследование операций в экономике 61
4.2 Задача о дилижансах 64
4.3 Анализ чувствительности решения к условиям задачи 67
4.4 Задача управления запасами с вогнутой функцией затрат 68
4.5 Алгоритм оптимизации модели. Числовой пример 71
4.6 Анализ чувствительности решения к длительности планового периода 73
4.7 Статическая модель 74
4.8 Случай единовременного штрафа 77
3.8 Работа по теме «Модели управления запасами» 80
5 Имитационное моделирование 81
5.1 Область применения 82
5.2 Целевая установка 85
5.3 Этапы построения и использования имитационной модели 86
5.4 Оценка надежности систем 87
5.5 Оценка абсолютной пропускной способности СМО 89
5.6 Оценка затрат, связанных с управлением запасов 92
5.7 Работа по теме «Имитационное моделирование» 95
6 Список литературы 97

Введение

Математическая модель – это приближенное описание какого-либо класса явлений внешнего мира, выраженное с помощью математической символики. Анализ математической модели позволяет проникнуть в сущность изучаемого явления. Математическая модель – мощный метод познания внешнего мира, а также прогнозирования и управления протекающих процессов.

При построении модели реальное явление неизбежно упрощается. Чем лучше модель отражает характерные черты явления, тем полезнее будут вытекающие из нее выводы и рекомендации. Вместе с тем, модель должна быть по возможности простой, не «засоренной» мелкими второстепенными факторами. Поэтому всегда присутствуют две опасности: первая – усложнить модель подробностями, «не увидев из-за деревьев леса», и вторая – слишком упростить явление, «выплеснув с водой ребенка».

Все вышесказанное относится и к вероятностным математическим моделям, в которых дополнительно присутствует фактор неопределенности. Часто в этих моделях используется некоторый показатель эффективности f, который зависит от трех групп факторов: заданных, заранее известных α; элементов искомого решения x; неизвестных факторов ξ. Например, мы собираемся в путешествие и складываем в чемодан вещи. Размеры чемодана и имеющийся набор вещей заданы (факторы α), а погода в районе путешествия заранее неизвестна (факторы ξ). Какие предметы одежды (решение x) надо взять с собой?

Для ответа на этот вопрос необходимо построить модель, позволяющую определить сумму приятных впечатлений от путешествия (f) в зависимости от факторов α, ξ и решения x.

Следует отметить, что разные люди будут строить модель поразному: если молодой человек, скорее всего, обратит внимание на экзотику и уникальность путешествия, то пожилой путешественник, пожалуй, предпочтет безопасность и комфорт. В любом случае в этой модели необходимо учитывать ретроспективные статистические данные о вероятной погоде в районе путешествия.

Другой пример: проектируется разработка рудного месторождения в зоне повышенной сейсмичности и тектонической активности (Горная Шория). Ни моменты активизации этих процессов, ни места их проявления заранее неизвестны. А проектировать все-таки нужно, и никакая неопределенность не избавит нас от этой обязанности.

Для того, чтобы строить модели таких явлений не наобум, а трезво, современная наука обладает рядом приемов. Каким из них воспользоваться – зависит от того, какова природа неизвестных факторов ξ, откуда они возникают и кем контролируются. Другими словами, с какой неопределенностью мы в данной модели сталкиваемся? Дело в том, что неопределенности можно классифицировать.

Наиболее благоприятным для моделирования видом неопределенности является случай, когда неизвестные факторы представляют собой **случайные величины** (или случайные функции), статистические характеристики которых нам известны или могут быть получены методом статистического оценивания. Такие модели **называются стохастическими**. Например, моделируется процесс обслуживания покупателей с целью повышения пропускной способности обслуживания и сокращения длины очередей. Нам в точности неизвестно, какое количество покупателей придет за рабочий день, когда именно они появятся, какие покупки сделают, и сколько времени будет продолжаться обслуживание каждого из них. Однако характеристики этих случайных величин могут быть получены статистическим путем.

Рассмотрим более подробно этот вид неопределенности. Самым простым было бы заменить случайные факторы ξ их средними значениями (математическими ожиданиями). Тогда модель становится детерминированной и может быть исследована обычными математическими методами. Такой подход применим в большинстве моделей физики, механики, в которых пренебрегается случайностью ряда параметров (давление, теплоемкость, трение) и принимаются их средние значения. Весь вопрос в том, насколько случайны эти параметры: если они мало отклоняются от своих математических ожиданий, так поступать можно и нужно. Тот же прием будет уже опрометчивым, если влияние случайности на интересующие нас процессы будет существенно. Возьмем модель обслуживания покупателей. Пренебрежем случайностью момента прихода покупателей и времени их обслуживания. Тогда окажется, что обслуживание покупателей, спланированное без учета случайности, попросту не будет справляться со своей задачей: в одни моменты времени длина очереди покупателей будет неограниченно возрастать, а в другие – служба обслуживания будет простаивать из-за отсутствия покупателей. Другими словами, встречаются (и очень часто) модели, в которых случайность входит по существу, и свести модель к детерминистической не удается.

Рассмотрим модель, в которой случайные факторы ξ входят по существу и заметно влияют на показатель эффективности f, который тоже «существенно случаен». Рассмотрим другой прием, согласно которому в качестве показателя эффективности выбирается среднее значение (математическое ожидание) этой случайной величины $\overline{f} = M(f)$, а в качестве x – такое решение, при котором этот усредненный показатель обращается в максимум $\overline{f} = M(\alpha, x, \xi) \Rightarrow \max$. В большинстве случаев такой подход вполне оправдан. Действительно, оптимизируя эффективность «в среднем», мы в конечном счете после многократного применения полученного решения x выиграем больше, чем если бы совсем не пользовались расчетом. Однако в каждом конкретном случае применения решения x эффективность f может сильно отличаться от ожидаемой как в большую, так, к сожалению, и в меньшую сторону.

Такая «оптимизация в среднем» применима в стохастических моделях, обладающих свойством повторяемости, при которой «недостача» показателя эффективности в одном случае компенсируется его «избытком» в другом. Например, если мы каждый день применяем одну и ту же стратегию управления запасами предприятия с целью минимизации затрат, то ежедневные затраты суммируются, «минус» в одном случае покрывается «плюсом» в другом.

Однако если модель не обладает свойством повторяемости, то «оптимизация в среднем» не применима. Например, в службе неотложной медицинской помощи большого города применяется стратегия обслуживания заявок с целью минимизации времени ожидания врача. Но время ожидания врача отдельным больным – величина случайная, которая **не суммируется**. Если применить «оптимизацию в среднем», то при малом времени ожидания «в среднем» некоторые больные будут ожидать врача очень долго! Чтобы избежать таких неприятностей необходимо дополнить показатель эффективности условием, чтобы фактическое время T ожидания врача было бы не больше какого-то предельного значения t_0. Однако время T – случайная величина, для которой нельзя просто потребовать выполнения условия $T \le t_0$. Можно потребовать, чтобы оно выполнялось **с очень большой вероятностью** $\boldsymbol{\beta}$. Таким образом, добавочное условие приобретает вид

$$P(T \le t_0) \ge \beta .$$

Например, в диспетчерском пункте г. Нью-Йорка штат операторов, обслуживающих срочные вызовы по телефону 911, выбирается с помощью этого условия, в котором T – время ожидания свободного оператора, t_0 = 20 сек, а β = 0,95. Это условие называется **стохастическим ограничением**. Наличие такого условия сильно усложняет модель.

Рассмотрим вероятностные модели, которые не являются стохастическими. Это бывает в двух случаях:

1) распределение вероятностей для параметров ξ в принципе существует, но к моменту принятия решения не может быть получено;

2) распределение вероятностей для параметров ξ вообще не существует.

Ситуация типа 1) встречается в случае, когда к моменту принятия решения еще нет ретроспективных данных о случайных параметрах β, а решение принимать надо! В этом случае некоторые элементы решения $\boldsymbol{x}$ оставляются свободными, изменяемыми. Затем выбирается для начала какой-то вариант решения, хотя заведомо известно, что он не самый лучший. Далее, по мере накопления опыта, целенаправленно изменяются свободные параметры решения для того, чтобы эффективность не уменьшалась, а увеличивалась. Такие модели, совершенствующиеся в процессе применения, называются **адаптивными**. Преимущество адаптивных моделей в том, что они не только избавляют нас от предварительного сбора статистики, но и перестраиваются в ответ на изменения обстановки.

Ситуация типа 2) встречается в случае, когда параметры ξ не обладают устойчивостью при однородных опытах. Такой случай называется «дурной неопределенностью». Например, моделируется прибыль некоторой торгово-промышленной компании, которая зависит от того, юбки какой длины ξ будут носить женщины через два года. Распределение вероятностей для величины ξ в принципе не может быть получено ни из каких статистических данных. Просто величина ξ не обладает устойчивостью во времени.

В настоящее время нет полноценной научной теории для таких ситуаций. Обычно проводятся предварительные расчеты, в ходе которых для разных значений параметра ξ определяются сильные и слабые стороны решений x. Они помогают заранее отбросить те решения, которые при **любых** значениях ξ уступают другим, т. е. оказываются неконкурентоспособными. В ряде случаев это позволяет существенно сузить класс допустимых решений, сведя его к небольшому числу вариантов, которые легко могут быть просмотрены в поисках удачного компромисса.

При рассмотрении моделей с «дурной неопределенностью» всегда полезно применять разные подходы и сталкивать в споре разные точки зрения. Одна из них называется «позицией крайнего пессимизма» и сводится к тому, что надо всегда рассчитывать на худшее и принимать то решение, которое дает максимальный эффект в наихудших условиях. Если в этих условиях мы получим выигрыш, то в любых других он будет не меньшим («принцип гарантированного результата»). Такой подход оправдан в так называемых «конфликтных ситуациях», когда оба противника руководствуются стратегией «наибольшей зловредности». В более нейтральных ситуациях этот принцип не является единственно возможным. Действительно, если военачальник всегда будет принимать решения исходя из гипотезы, что его противник необычайно умен, то вряд ли ему будет сопутствовать удача.

1 Метод статистических испытаний

1.1 Идея метода

Пусть требуется построить некоторую аналитическую модель со стохастической неопределенностью. Эта модель позволяет установить аналитическую (формальную) зависимость между исходными данными, элементами решения и результатом (исходом), который характеризуется одним или несколькими показателями. Польза и желательность построения аналитических моделей (хотя бы приближенных) сомнению не подлежат. Беда в том, что их удается построить только для самых простых «незатейливых» ситуаций, и, самое главное, они требуют допущения о марковском характере случайного процесса, что далеко не всегда соответствует действительности. В случаях, когда аналитические методы неприменимы, приходится прибегать к универсальному методу статистического моделирования или, как его часто называют, **методу статистических испытаний**.

Метод статистических испытаний позволяет заменить решение задачи аналитическими методами (дифференциального и интегрального исчисления) «разыгрыванием» случайного явления с помощью специально организованной процедуры, включающей в себя случайность и дающей случайный результат. Так же как в действительности в результате статистического моделирования («розыгрыша») мы получаем каждый раз новую реализацию случайного процесса [3, 4, 15].

Для чего нужна такая реализация? Сама по себе она не имеет никакого значения. Так же как один случай излечения больного с помощью какого-то лекарства (или несмотря на это лекарство). Другое дело, если таких реализаций получится много. Это множество реализаций можно использовать как некий искусственно полученный статистический материал, который может быть обработан обычными методами математической статистики. После такой обработки могут быть получены любые интересующие нас характеристики: вероятности событий, математические ожидания и дисперсии случайных величин и т. д.

При моделировании случайных явлений методом статистических испытаний мы пользуемся самой случайностью как аппаратом исследования, заставляя его работать на нас.

В сущности, методом статистических испытаний может быть решена любая вероятностная задача. Но оправданным он становится только тогда, когда процедура розыгрыша проще, а не сложнее аналитического расчета. Такая ситуация складывается в случаях, когда необходимо проверить приемлемость основных допущений аналитической модели с помощью контрольных расчетов. В этих случаях, образно говоря, метод статистических испытаний играет роль своеобразного ОТК.

Дело в том, что статистические модели не требуют серьезных допущений и упрощений. В них могут быть использованы любые законы распределения, любая сложность системы, множественность ее состояний. Современные вычислительные средства позволяют оперативно получить любое число реализаций, необходимое для нахождения искомых параметров с приемлемой точностью. Однако результаты статистического моделирования гораздо труднее осмыслить, чем расчеты по аналитическим моделям. Поэтому труднее оптимизировать полученное решение (его приходится «нащупывать» вслепую). Правильное сочетание аналитических и статистических методов – дело искусства и опыта исследователя. Нередко аналитическими методами удается описать какие-то составные части модели, а затем из таких частей, как из «кирпичиков», строить здание большой модели [3].

1.2 Единичный жребий

Основным элементом, из множества которых складывается статистическая модель, является **одна случайная реализация** моделируемого явления. Например, один случай работы автомобиля до его отказа, один день работы фирмы, одна эпидемия гриппа и т. д. Реализация – это как бы один «экземпляр» случайного явления со всеми присущими ему случайностями. Реализации отличаются друг от друга вследствие этих случайностей.

Отдельная реализация разыгрывается с помощью специально разработанной процедуры, в которой важную роль играет «бросание жребия». Каждый раз, когда в ход явления вмешивается случай, его влияние учитывается не расчетом, а **жребием**.

Поясним понятие «жребия». Пусть в ходе процесса наступил момент, когда его дальнейшее развитие (а значит и результат) зависит от того, произошло или нет какое-то событие A? Например, попал ли в цель снаряд? Сдал ли студент экзамен? Выигран ли футбольный матч? Тогда нужно «бросанием жребия» решить вопрос: произошло событие или нет. Как можно осуществить этот жребий? Нужно привести в действие какой-то механизм случайного выбора (например, бросание монеты или игральной кости). Если жребий бросается для того, чтобы узнать, произошло ли событие A, то нужно организовать его так, чтобы условный результат розыгрыша имел ту же вероятность, что и событие A. Как это делается – мы рассмотрим ниже.

Кроме случайных событий объектом розыгрыша могут быть случайные величины. Например, время до первого отказа технического устройства; время прилета самолета; вес поезда, прибывающего на станцию; координаты очага землетрясения и т. п. Таким образом, с помощью жребия можно разыгрывать системы нескольких случайных величин.

Условимся называть «единичным жребием» любой опыт со случайным исходом, который отвечает на один из следующих вопросов:

- произошло или нет событие А;
- какое из событий А1, А2, … , Аn произошло;
- какое значение приняла случайная величина Х;
- какую совокупность значений приняла система случайных величин $X_1, X_2, \dots, X_n$?

Любая реализация случайного явления методом статистических испытаний строится из цепочки единичных жребиев, перемежающихся с обычными расчетами. Ими учитывается влияние исхода жребия на дальнейший ход событий (в частности, на условия, в которых будет разыгран следующий жребий).

Единичный жребий может быть разыгран разными способами. Но есть один стандартный механизм, с помощью которого можно осуществить любую разновидность жребия: для каждой из них достаточно уметь получать случайное число R, все значения которого равномерно распределены в интервале (0, 1). Условимся кратко называть величину R – «случайным числом (0, 1)». Покажем, что с помощью такого числа можно разыграть любой из четырех видов единичного жребия.

1.3 Разыгрывание случайных событий и величин

Чтобы ответить на вопрос, **произошло или нет событие A**, надо знать вероятность p события A. Разыграем случайное число $R(0, 1)$. Если оно окажется меньше p, будем считать, что событие произошло, а если больше p – не произошло.

А как быть, – спросит читатель, – если число R окажется в точности равным p? Вероятностью такого совпадения можно пренебречь.

Пусть события $A_1, A_2, \dots, A_n$ несовместны и образуют полную группу. Тогда сумма их вероятностей $p_1, p_2, \dots, p_n$ равна единице. Разделим интервал (0, 1) на n участков длиной $p_1, p_2, \dots, p_n$ (рисунок 1.1). На какой из участков попало число R, такое событие (A_2) и появилось.

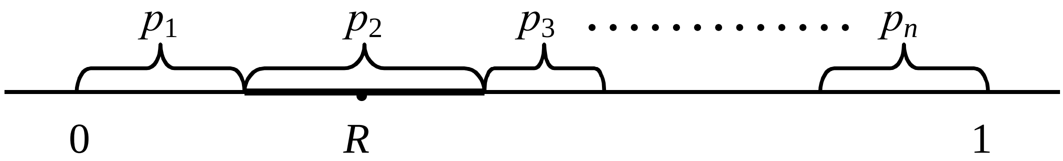

Рисунок 1.1 – Разыгрывание полной группы событий

Если **случайная величина X дискретна**, т. е. принимает возможные значения $x_1, x_2, \dots, x_n$ с вероятностями $p_1, p_2, \dots, p_n$, то, очевидно, случай сводится к предыдущему. Теперь рассмотрим случай, когда **случайная величина является непрерывной** и имеет заданную плотность вероятности $f(x)$. Чтобы разыграть ее значение, достаточно осуществить следующую процедуру: перейти от плотности вероятности $f(x)$ к функции распределения $F(x)$ по формуле

$$F(x) = \int_{-\infty}^{x} f(x)\,dx,$$

затем найти для функции F обратную ей функцию G. Затем разыграть случайное число $R(0, 1)$ и для него найти такое значение $\boldsymbol{X}$, при котором $\boldsymbol{X} = G(R)$. Можно доказать, что полученное значение $\boldsymbol{X}$ имеет как раз нужное нам распределение $f(x)$.

На практике часто приходится разыгрывать значение случайной величины, имеющей **нормальное распределение**. Для нее, как для любой непрерывной случайной величины, правило розыгрыша остается справедливым. Но можно поступить иначе, воспользовавшись центральной предельной теоремой Ляпунова: при сложении достаточно большого числа независимых случайных величин с одинаковыми распределениями получается случайная величина, имеющая приближенно нормальное распределение. На практике, чтобы получить нормальное распределение, достаточно сложить шесть экземпляров случайного числа $R(0, 1)$. Сумма этих шести чисел $Z = R_1 + R_2 + \ldots + R_6$ имеет распределение, настолько близкое к нормальному, что в большинстве практических задач им можно заменить нормальное. Для того, чтобы математическое ожидание и среднее квадратическое отклонение этого нормального распределения были равны заданными (m, σ), нужно подвергнуть величину Z линейному преобразованию и вычислить

$$X = \sigma\sqrt{2}(Z-3)+m.$$

Это и будет нужная нам нормально распределенная случайная величина.

Если разыгрывается совокупность значений **системы случайных величин** $X_1, X_2, \ldots, X_n$, которые независимы, то достаточно n раз повторить вышеописанную процедуру. Если же они зависимы, то разыгрывать каждую последующую нужно на основе ее условного закона распределения при условии, что все предыдущие приняли те значения, которые дал розыгрыш.

Мы рассмотрели все четыре варианта единичного жребия и убедились, что все они сводятся к **розыгрышу случайного числа** $R(0, 1)$. А как же разыгрывается это число? Существует целый ряд разновидностей так называемых «датчиков случайных чисел», решающих эту задачу. Самый простой из них – это вращающийся барабан, в котором перемешиваются занумерованные шарики с номерами 1, 2, … , 1000. Приведем барабан во вращение и после остановки выберем первый попавшийся шарик, прочтем его номер и разделим на 1000.

Можно поступить иначе: вместо 1000 шариков заложить в барабан только 10, с номерами 0, 1, … , 9. Вынув один шарик, прочтем первый десятичный знак дроби. Вернем его обратно, снова покрутим барабан и возьмем второй шарик – это будет второй десятичный знак и т. д. Легко доказать, что полученная таким образом десятичная дробь будет иметь равномерное распределение в интервале (0, 1). Преимущество этого способа в том, что он никак не связан с числом знаков, с которым мы хотим знать R.

Отсюда один шаг до рационализаторского предложения: не разыгрывать число R каждый раз, когда это понадобится, а сделать это заранее, т. е. составить достаточно обширную таблицу, в которой все цифры 0, 1, … , 9 встречаются случайным образом и с одинаковой вероятностью (частотой). Такие таблицы действительно составлены и применяются на практике. Они называются **таблицами случайных чисел** (*приложение Е*).

При ручном применении метода статистических испытаний таблицы случайных чисел – наилучший способ разыгрывания случайного числа $R(0, 1)$. Если же моделирование осуществляется на ЭВМ, то применяются встроенные функции рандомизации, позволяющие вычислить так называемые «псевдослучайные числа». Приставка «псевдо» означает, что фактически эти числа случайными не являются, но практически ведут себя как случайные: все значения в интервале (0, 1) встречаются в среднем одинаково часто и, кроме того, связь между последовательными значениями получаемых чисел практически отсутствует. Существует ряд алгоритмов получения псевдослучайных чисел, различающихся между собой по простоте, равномерности и другим признакам. Одним из самых простых алгоритмов является выбор двух произвольных n-значных двоичных чисел. Затем перемножают их и в полученном произведении берут n средних знаков. Затем перемножают полученное число с одним из предыдущих чисел и в произведении снова берут n средних знаков и т. д. Полученная таким образом последовательность чисел является практически случайной.

1.4 Работа по теме «Метод статистических испытаний»

1. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [0, 4], проверив гипотезу о форме распределения.
2. Смоделировать выборку 100 значений случайной величины x, распределенной нормально с параметрами (20, 10), проверив гипотезу о форме распределения.
3. Смоделировать выборку 45 значений случайной величины x, распределенной по биномиальному закону с параметрами n = 40, p = 0,1, проверив гипотезу о форме распределения.
4. Смоделировать выборку 35 значений случайной величины x, распределенной по биномиальному закону с параметрами n = 20, p = 0,2, проверив гипотезу о форме распределения.
5. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [–3, 2], проверив гипотезу о форме распределения.
6. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [2, 4], проверив гипотезу о форме распределения.
7. Смоделировать выборку 100 значений случайной величины x, распределенной нормально с параметрами (20, 15), проверив гипотезу о форме распределения.
8. Смоделировать выборку 45 значений случайной величины x, распределенной по биномиальному закону с параметрами n = 30, p = = 0,1, проверив гипотезу о форме распределения.
9. Смоделировать выборку 35 значений случайной величины x, распределенной по биномиальному закону с параметрами n = 20, p = = 0,3, проверив гипотезу о форме распределения.
10. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [–1, 2], проверив гипотезу о форме распределения.

11. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [1, 4], проверив гипотезу о форме распределения.

12. Смоделировать выборку 100 значений случайной величины x, распределенной нормально с параметрами (20, 5), проверив гипотезу о форме распределения.

13. Смоделировать выборку 40 значений случайной величины x, распределенной по биномиальному закону с параметрами $n = 40$, $p = 0{,}1$, проверив гипотезу о форме распределения.

14. Смоделировать выборку 30 значений случайной величины x, распределенной по биномиальному закону с параметрами $n = 20$, $p = 0{,}3$, проверив гипотезу о форме распределения.

15. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [–3, 0], проверив гипотезу о форме распределения.

16. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [–2, 4], проверив гипотезу о форме распределения.

17. Смоделировать выборку 100 значений случайной величины x, распределенной нормально с параметрами (10, 5), проверив гипотезу о форме распределения.

18. Смоделировать выборку 40 значений случайной величины x, распределенной по биномиальному закону с параметрами $n = 20$, $p = 0{,}3$, проверив гипотезу о форме распределения.

19. Смоделировать выборку 30 значений случайной величины x, распределенной по биномиальному закону с параметрами $n = 25$, $p = 0{,}4$, проверив гипотезу о форме распределения.

20. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [–2, 3], проверив гипотезу о форме распределения.

21. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [–2, 3], проверив гипотезу о форме распределения.

22. Смоделировать выборку 100 значений случайной величины x,

распределенной нормально с параметрами (15, 8), проверив гипотезу о форме распределения.

23. Смоделировать выборку 50 значений случайной величины x, распределенной по биномиальному закону с параметрами n = 15, p = = 0,2, проверив гипотезу о форме распределения.

24. Смоделировать выборку 40 значений случайной величины x, распределенной по биномиальному закону с параметрами n = 25, p = = 0,2, проверив гипотезу о форме распределения.

25. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [–7, 0], проверив гипотезу о форме распределения.

26. Смоделировать выборку 80 значений случайной величины x, имеющей равномерное распределение на промежутке [–2, 5], проверив гипотезу о форме распределения.

27. Смоделировать выборку 100 значений случайной величины x, распределенной нормально с параметрами (12, 5), проверив гипотезу о форме распределения.

2 Линейные статистические модели

2.1 Выборочное уравнение прямой линии регрессии

В данном разделе рассмотрим задачу установления зависимости изучаемой случайной величины *У* от одной или нескольких других величин. Для этого изучим сначала зависимость *У* от одной случайной величины *X*. Эта зависимость может быть трех типов:

1. **Функциональная** зависимость имеет место, когда каждому значению независимой величины *X* соответствует одно определенное значение зависимой величины *У*.

2. **Статистическая** зависимость имеет место, когда изменение одной из величин влечет изменение распределения другой.

3. **Корреляционная** зависимость имеет место, когда изменение одной из величин влечет изменение условного среднего другой.

Условным средним $\overline{y}_x$ называют среднее арифметическое наблюдаемых значений *У*, соответствующих $X = x$.

Пусть изучается система двух количественных признаков (*X, У*). Корреляционная зависимость этих признаков считается линейной, если нет никакой дополнительной информации о виде этой зависимости. В противном случае применяется метод выравнивания, который сводит нелинейную зависимость к линейной с помощью необходимой замены переменных.

Выборочное уравнение прямой линии среднеквадратичной регрессии *У* на *X* имеет вид:

$$\overline{y}_x = \overline{y} + r_{\text{в}} \frac{\overline{\sigma}_y}{\overline{\sigma}_x}(x - \overline{x}) \text{, где } r_{\text{в}} = \frac{\sum n_{uv} uv - n\overline{u}\overline{v}}{n\overline{\sigma}_u \overline{\sigma}_v}.$$

Рассмотрим пример отыскания выборочного уравнения прямой линии среднеквадратичной регрессии *У* на *X* по несгруппированным данным двумерной выборки [7].

▪ <u>*Пример* 1.</u> Найти выборочное уравнение прямой линии регрессии *У* на *X* согласно протоколу наблюдений двумерной выборки (x_i, y_i), $i = 1, \ldots, 100$, приведенному в таблице 2.1. Для этого необходимо: провести группировку данных таблицы 2.1, выбрав не менее пяти интервалов по каждой переменной, содержащих не менее пяти точек; составить корреляционную таблицу; методом произведений вычислить выборочный коэффициент парной регрессии, а также выборочные средние и средние квадратические отклонения этих переменных; проверить гипотезу об отсутствии корреляционной связи.

Таблица 2.1

X	*У*	*X*	*У*	*X*	*У*	*X*	*У*	*X*	*У*
21,47	20	32	32	40	36	25	23,93	34	48,01
42	45,48	24,14	13	28,64	46,61	34,03	36,02	30	26,91
24,23	18	19,85	18,02	26	23,01	23,74	17,85	29,24	43,37
41	53,39	18,53	19,46	30,07	43,17	28,5	36,29	36,81	45,36
30	46,96	35	41,89	32,4	37,49	31,17	38	35,28	57,47
24	27,11	29,95	30	40	55,72	29,95	38,19	28	40
24,13	12	28,45	35	35,18	40,14	26	28,62	23,83	15,5
30	31,25	29	32,38	40	49,12	30,66	29	26	24,51
31	35,26	28	31,64	28	36,74	29	35	38	40

25	25,3	28	39	28,5	32,09	30,25	29	32	35,33
40	34,94	39	57,79	29,8	33,71	32	33	35,47	46
29,5	31,62	30,97	30	39	38	19,52	18,32	41	35
36,09	46	41	53	29	38	26,49	23,73	38,5	42,65
31,13	27,67	31	38,74	28,84	29	39,5	45	35,89	45
32,06	22,49	34,9	44,15	41	45,67	31	32	33,71	45,63
39	32,91	30	36	27,5	39,26	30,33	38,59	31	40
31,84	39	32,19	39,54	40,36	33,36	30,43	28,75	40,5	40,32
30	39,56	34,19	35,6	38	36,47	40	45	33,18	43,88
30,5	38,23	36,38	44,08	30,5	39,66	22,73	20,05	29	40
25	29,35	28,58	30,05	35,59	47,63	36,21	47,95	42	56

Решение. Сгруппируем данные таблицы 2.1. Хотя в этом процессе исходные данные теряются, но сохраняется удовлетворительное общее представление об искомой зависимости. Наблюдения двумерной выборки в виде n = 100 точек с координатами (x_i, y_i) изображаются на координатной плоскости и заключаются в прямоугольник, который разбивается на прямоугольные клетки со сторонами h_x и h_y , подобно тому как разбивается отрезок на k интервалов в примере 1 п. 2.1. На рисунке 2.1 выбрано k = 5, так как оно удовлетворяет неравенствам $ln(n) \le k \le \sqrt{n}$, а стороны клеток h_x = 5, h_y = 10 являются целыми числами.

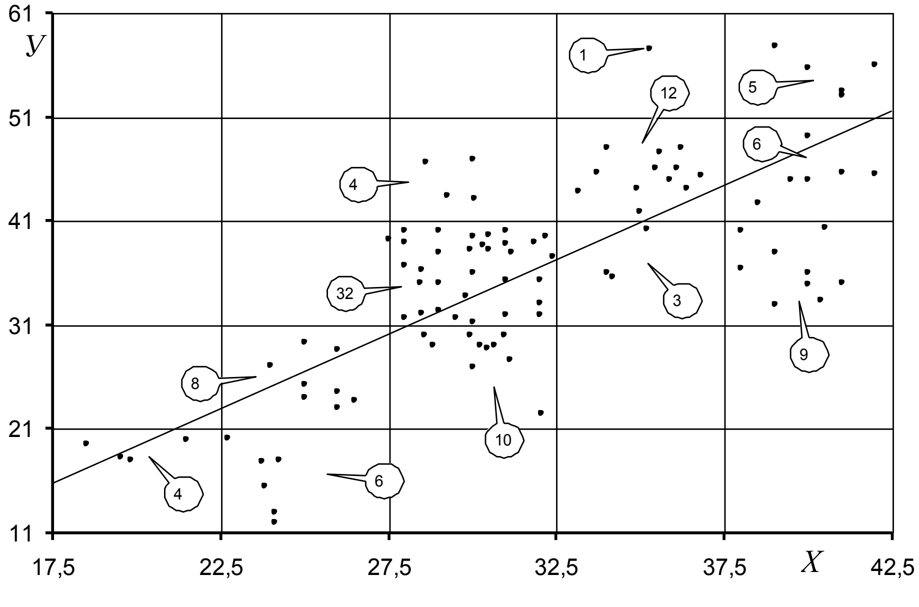

Рисунок 2.1 – Группировка наблюдений двумерной выборки

Если точка расположена на границе, разделяющей две клетки, то она относится к той клетке, в которой находится больше точек. Найденные таким образом значения частот приведены в клетках **корреляционной** таблицы 2.2, обведенных двойной линией.

Таблица 2.2

У	*X*					n_y
	20	25	30	35	40	
16	4	6				10
26		8	10			18
36			32	3	9	44
46			4	12	6	22
56				1	5	6
n_x	4	14	46	16	20	$n = 100$

Сверху и слева указаны соответствующие координаты центров клеток (варианты *X, У*), а снизу и справа – суммарные частоты вариант.

Для каждой случайной величины *X* и *У* вводятся условные варианты u и v. Составим корреляционную таблицу 2.3 в условных вариантах.

Таблица 2.3

v	u					n_v	$U=\Sigma n_{uv}u$	vU
	–2	–1	0	1	2			
–2	–8 4	–6 6				10	–14	28
–1		–8 8	0 10			18	–8	8
0			0 32	3 3	18 9	44	21	0
1			0 4	12 12	12 6	22	24	24
2				1 1	10 5	6	11	22
n_u	4	14	46	16	20	100	$\Sigma\, vU = 82$	

Методом произведений найдем выборочные средние:

сначала $\overline{u}=0{,}34$ и $\overline{u^2}=1{,}26$; затем $\overline{v}=-0{,}04$ и $\overline{v^2}=1{,}04$.

Выборочные средние квадратические отклонения условных вариант u и v найдем по формулам:

$$\overline{\sigma}_u=\sqrt{\overline{u^2}-(\overline{u})^2}=1{,}07;\quad \overline{\sigma}_v=\sqrt{\overline{v^2}-(\overline{v})^2}=1{,}02.$$

Затем проверяется гипотеза о нормальном распределении генеральных совокупностей *X* и *У* (см. пример 2, подраздел 1.5). Только после этого отыскивается выборочный коэффициент корреляции.

Произведение частоты n_{uv} на варианту u , то есть $n_{uv}u$, записывают в правом верхнем углу клетки таблицы 2.3, содержащей частоту n_{uv}. Например, в правых верхних углах клеток первой строки записаны произведения: 4(–2) = –8; 6(–1) = –6. В предпоследнем столбце приведены значения $U = \Sigma n_{uv}u$, которые получаются суммированием по строкам чисел в правых верхних углах. В последнем столбце приведены произведения числа U и соответствующего числа v. Суммируя числа из последнего столбца, получим $\Sigma n_{uv}uv = \Sigma vU = 82$.

Выборочный коэффициент корреляции найдем по формуле

$$r_{\text{в}}=\frac{\Sigma n_{uv}uv-n\overline{u}\overline{v}}{n\overline{\sigma}_u\overline{\sigma}_v}=\frac{82-100\cdot 0{,}34\cdot(-0{,}04)}{100\cdot 1{,}07\cdot 1{,}02}=0{,}76.$$

Поскольку значение $r_{\text{в}} > 0{,}196$ (*приложение К*), то с уровнем значимости 0,05 можно утверждать о зависимости *У* от *X*. Если это условие не выполняется, то можно принять гипотезу об отсутствии корреляционной зависимости с отмеченным уровнем значимости (см. пример 5, подраздел 4.6).

Найдем выборочные средние $\overline{x}$ и $\overline{y}$, учитывая, что шаги $h_x = 5$, $h_y =$ $= 10$, а ложные нули $C_x = 30$; $C_y = 36$:

$$\overline{x}=\overline{u}\cdot h_x+C_x=0{,}34\cdot 5+30=31{,}7;\quad \overline{y}=\overline{v}\cdot h_y+C_y=-0{,}04\cdot 10+36=35{,}6.$$

Затем найдем выборочные средние квадратические отклонения:

$$\sigma_x = h_x\,\sigma_u = 5\cdot 1{,}07 = 5{,}35;\ \sigma_y = h_y\,\sigma_v = 10\cdot 1{,}02 = 10{,}2.$$

Подставим найденные величины в уравнение регрессии *Y* на *X*

$$\overline{y}_x - \overline{y} = r_{в}\frac{\sigma_y}{\sigma_x}(x-\overline{x}). \quad \text{Получим} \quad \overline{y}_x - 35{,}6 = 0{,}76\frac{10{,}2}{5{,}35}(x-31{,}7)$$

или окончательно $\overline{y}_x = 1{,}45x - 10{,}36$. График полученной корреляционной зависимости приведен на рисунке 2.1. Этот график разбивает облако данных на две «симметричные» части и наилучшим образом приближает эти данные.

2.2 Работа по теме «Линейные статистические модели»

Таблица Б.1

j	*i*						
	1	2	3	4	5	6	7
9						***30***	***29***
8	***22***	***23***	***24***	***25***	***26***	***27***	***28***
7	***21***	***20***	***19***	***18***	***17***	***16***	
6	***11***	***12***	***13***	***14***	***15***		
5	***10***	***9***	***8***	***7***			
4	***4***	***5***	***6***				
3	***3***	***2***					
2	***1***						

Наблюдаемые значения случайных величин **Х** , **У** определяются по правилу: $\boldsymbol{x = Z_i}$, $\boldsymbol{y = Z_j}$ **,** где индексы ***i, j*** определяются номером варианта с помощью табл. 1. Например, для варианта № ***10*** выбираем клетку в таблице Б.1 с цифрой ***10***. Соответствующие индексы $\boldsymbol{i} = 1$, $\boldsymbol{j} = 5$ – номера столбца и строки, проходящей через рассматриваемую клетку.

Значения $\boldsymbol{Z_i}$, $\boldsymbol{Z_j}$ приведены в таблице Б.2. Необходимо:

1. Построить облако данных на плоскости ***x, y*** и выполнить графически группировку наблюдений величин.

2. Составить корреляционную таблицу и вычислить выборочные средние, дисперсии, стандартные отклонения, асимметрии и эксцессы, коэффициент корреляции случайных величин ***x, y*** с помощью метода произведений и пакета программ «Статистика».

3. С помощью метода произведений и пакета программ «Статистика» найти выборочное уравнение прямой линии регрессии ***y*** на ***x***, изобразить линию регрессии на облаке данных и проверить гипотезу об отсутствии корреляционной связи между величинами ***x, y***.

4. Найти выборочное уравнение $x = b_0 + b_1 y_1 + b_2 y_2 + ... + b_8 y_8$ множественной регрессии, сравнить эмпирические и расчетные значения функции ***x***. Определить с помощью пакета программ «Статистика» или приложения «Excel», какие независимые переменные этой связи являются значимыми, и найти выборочное уравнение только с этими переменными.

Таблица Б.2

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
24,82	16,15	41,79	21,6	36,32	21,47	44,91	18,67	20,28
16,6	33,42	45,39	17,89	42,13	29,93	45,48	32,85	18,76
28,05	14,47	42,53	18,41	34,8	24,23	55,06	35,16	16,17
19,31	12,83	37,65	17,73	35,67	20,52	53,39	33,62	18,78
27,25	23,05	27,59	17,49	36,59	15,82	46,96	23,79	16,32
19,56	7,28	34,82	15,57	30,89	18,9	27,11	18,96	18,37
20,45	33,94	30,84	18,9	36,21	24,13	29,24	24,11	16,48
15,46	15,01	30,58	18,43	36,42	18,12	31,25	19,41	17,57
20,77	20,1	37,31	23,57	37,59	25,67	35,26	27,39	13,21

Продолжение таблицы Б.2

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
21,95	18,55	29,47	11,31	34,48	31,75	25,3	22,14	16,9
25,6	17,68	53,67	21,56	40,37	25,47	32,7	29,71	17,92
28,27	21,2	54,85	12,57	34,19	24,14	57,62	32,24	10,63
19,7	28,69	35,75	14,44	38,61	19,85	18,02	33,67	11,56
18,09	8,21	30,48	13,31	35,6	18,53	19,46	23,43	15,32
13,28	14,92	28,71	12,07	33,51	16,41	41,89	32,51	13,84
33,17	17,92	60,19	21,54	20,76	29,95	50,27	21,34	42,73
30,93	28,55	48,22	22,04	16,7	38,45	61,9	22,24	26,07

27,66	26,99	52,96	18,42	27,65	31,41	32,38	20,02	11,45
28,21	30,93	39,99	20,66	31,83	29,54	31,64	16,46	29,24
34,15	23,68	70,26	18,08	32,19	26,01	50,78	27,29	23,93
26,79	25,77	61,73	18,67	19,63	34,61	40,43	18,31	37,17
26,31	12,04	41,59	16,89	21,29	28,64	46,61	18,04	30,55
25,44	14,29	47,74	23,42	24,96	22,32	23,01	22,9	21,99
32,01	18,23	30,94	20,42	32,05	30,07	43,17	16,82	39,68
31,43	25,86	69,48	15,7	20,5	32,4	37,49	20,21	36,66
32,54	29,25	81,55	23,96	31,21	37,03	55,72	29,16	51,35
31,25	27,31	69,16	20,71	26,04	35,18	40,14	20,51	31,58
34,97	24,43	67,96	25,58	28,77	28,78	49,12	24,9	34,92
30,12	23,96	50,6	28,68	32,08	28,71	36,74	24,39	21,89
38,48	28,3	71,92	25,58	23,08	24,67	32,09	25,79	18,76
33,34	26,62	66,99	27,3	26,64	21,32	23,93	17,78	39,84
28,64	32,03	55,27	31,93	31,08	34,03	36,02	23,66	29,23
32,64	25,54	53,74	19,43	35,65	23,74	17,85	23,12	24,36
22,87	20,18	26,67	18,17	22,94	21,46	36,29	15,35	21,16
28,94	30,72	62,33	20,67	38,82	31,17	41,5	19	37,62
34,9	22,46	48,67	29,56	32,27	29,95	38,19	19,45	25,11
32,2	30,77	71,91	20,54	33,65	41,91	28,62	22,11	28,74
34,54	28,88	35,67	21,38	31,79	30,66	66,97	23,48	30,04
22,28	16,34	49,72	16,14	21,82	27,1	41,56	17,92	22,99
40,06	17,01	72,94	30,66	26,52	30,25	67,33	17,06	34,48
24,72	29,77	46,69	22,69	32,67	28,14	48,01	20,8	18,45
27,85	18,64	55,97	11,28	26,75	20,05	26,91	18,61	26,39

Продолжение таблицы Б.2

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
34,67	18,71	62,15	16,68	32,01	29,24	43,37	18,34	36,38
26,77	24,29	52,72	17,18	28,25	36,81	45,36	21,89	23,76
37,41	37,33	83,81	23,92	38,12	35,28	57,47	29,01	48,8
30,77	19,9	29,83	20,44	34,46	25,73	41,51	21,84	25,04
32,21	24,81	49,88	16,81	29,89	23,83	15,5	21,81	25,31
26,9	19,04	46,02	17	13,28	22,72	24,51	18,65	27,06
28,97	23,97	44,86	22,23	33,16	32,64	31,23	17,81	24,87
29,96	15,46	67,43	20,06	33,72	32	35,33	26,04	21,17

28,03	15,21	29,18	15,84	18,91	30,78	29,35	18,06	30,22
32,5	24,32	36,14	23	29,8	35,65	34,94	21,59	40,48
28,67	16,6	20,35	24,02	25,86	21,13	31,62	17,93	14,81
32,97	25,84	50,7	28,05	31,86	36,09	58,02	22,67	27,98
29,1	16,56	54,01	15,49	20,18	31,13	27,67	18,71	40,35
35,62	20,28	55,15	22,85	30,38	32,06	22,49	29,69	26,72
26,97	26,55	40,13	9,2	28,38	32,69	32,91	22,99	23,89
34,78	26,83	27,58	24,3	28,53	31,84	41,19	21,56	16,61
27,97	25	59,34	18,31	27,3	20,07	39,56	25,4	10,09
27	18,12	18,18	16,85	33,01	25,2	38,23	21,7	11,27
28,63	28,09	60,99	22,02	25,93	26,24	57,79	23,55	29,95
27,05	27,97	42,98	20,78	23,65	30,97	58,97	18,54	40,03
32,19	22,13	76,33	25,88	36,01	23,95	63,53	26,14	19,83
31,17	13,66	72,15	16,97	33,15	26,95	38,74	21,36	39,13
25,57	14,96	27,73	20,98	34,25	34,9	44,15	22,03	14,25
17,93	16,92	24,07	25,8	30,84	27,05	41,28	16,68	31,64
27,73	21,48	66,77	25,5	34,66	32,19	39,54	18,99	37,07
35,05	23,22	52,39	32,85	43,38	34,19	35,6	18,42	39,75
27,38	25,18	70,58	25,45	27,96	36,38	44,08	24,06	12,51
31,7	26,76	59,88	22,09	25,66	28,58	30,05	22,82	40,04
28,43	19,59	54,44	11,25	23,95	21,62	33,71	21,27	11,4
30,85	18,87	41,31	25,3	26,81	29,8	45,37	19,56	30,62
32,22	23,13	36,84	21,44	27,39	23,24	46,92	20,84	27,34
33,96	21,92	66,23	17,12	30,32	28,84	59,08	23,47	18,04
30,81	21,05	61,97	20,03	34,65	30,18	45,67	15,05	29,29
31,35	22,66	52,32	15,05	29,67	27,5	39,26	19,69	29,2

Продолжение таблицы Б.2

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
32,32	21,79	63,24	22,79	35,97	40,36	33,36	20,69	30,6
26,98	23,15	34,08	28,46	36,78	33,25	36,47	15,03	16,03
33,02	16,36	47,65	19,37	27,02	20,33	39,66	17,99	23,3
35,71	22,24	80,98	17,11	35,56	35,59	47,63	24,77	28,02
31,76	15,92	69,93	20,6	19,8	25,38	42,51	23,7	17,99
31,27	23,26	47,82	22,7	18,15	19,52	18,32	16,84	29,12
27,43	30,1	59,33	26,86	29,6	26,49	23,73	30,04	37,86
24,73	18,56	29,82	12,85	22,93	32,49	38,32	21,6	11,91

33,76	14,85	38,82	21,94	32,35	25,25	45,85	20,02	22,4
29,59	30,56	56,89	28,96	29,3	30,33	38,59	26,82	40,43
29,46	26,6	80,41	15,53	35,22	30,43	28,75	21,87	17,02
39,89	34,55	71,77	16,44	39,42	37,59	41,04	23,67	34,89
27,75	26,36	46,16	18,02	28,83	22,73	20,05	21,76	32,24
26,8	20,52	38,9	24,66	26,82	36,21	47,95	21,19	10,97
31,5	27,8	70,25	24,49	36,12	35,47	29,1	20,04	32,7
39,09	27,61	41,65	31,17	41,43	35,5	40,66	25,17	27,59
30,65	16,76	69,96	25,86	32,92	38,5	42,65	16,03	37,32
31,44	26,9	72,32	15,38	33,91	35,89	28,13	22,06	31,97
29,35	19,07	33,83	26,01	37,01	33,71	45,63	24,44	32,26
23,4	23,14	53,51	22,85	26,54	21,92	47,02	17,97	18,96
28,49	15,35	32,62	22,16	24,64	35,98	40,32	15,27	18,82
30,97	20,25	59,31	20,69	22,68	33,18	43,88	24,31	35,23
39,5	28,42	75,67	22,28	36,07	24,65	72,47	23,08	25,18
24,88	55,51	15,54	27,62	29,05	54,94	18,65	16,88	29,79

3 Модели систем массового обслуживания

3.1 Марковский случайный процесс

Примерами систем массового обслуживания (СМО) являются телефонные станции, ремонтные мастерские, билетные кассы, справочные бюро, магазины, парикмахерские и т. п. Каждая СМО состоит из некоторого числа обслуживающих единиц, которые называются **каналами** обслуживания. Каналами могут быть: линии связи, лифты, кассиры, продавцы, кареты скорой помощи и др. СМО могут быть одноканальными и многоканальными.

Каждая СМО предназначена для обслуживания своего **потока заявок**, поступающих в случайные моменты времени. Для обслуживания заявки требуется какое-то случайное время, после чего канал освобождается и готов к приему следующей заявки. Случайный поток заявок и времен обслуживания приводят к тому, что в какие-то периоды времени на входе СМО скапливается большое число заявок, стоящих в очереди либо покидающих СМО необслуженными. В другие же периоды

СМО будет работать с недогрузкой или вообще простаивать.

Процесс работы СМО является случайным с дискретными состояниями и непрерывным временем. Состояние СМО меняется скачком в моменты появления новой заявки или окончания обслуживания. Поэтому необходимо построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, правила работы, характер потока заявок) с интересующими нас характеристиками – показателями эффективности СМО, описывающими ее способность справляться с потоком заявок. В качестве таких показателей используются: среднее число заявок, обслуживаемых СМО в единицу времени; среднее число занятых каналов; среднее число заявок в очереди, среднее время ожидания обслуживания и т. д. Исследованию СМО посвящены работы [2, 3–6, 9–11, 13, 16, 20, 23].

Математический анализ работы СМО очень облегчается, если процесс этой работы – марковский. Пусть имеется некоторая физическая система, которая с течением времени меняет свое состояние, причем заранее неизвестным, случайным образом. Тогда говорят, что в системе протекает **случайный процесс**.

Следует отметить, что большинству процессов, протекающих в реальных системах, свойственны черты случайности. Рассмотрим, например, космический корабль, выводимый на заданную орбиту. Процесс вывода неизбежно сопровождается случайными ошибками, отклоняющими корабль от заданной орбиты. Поэтому обязательно проводится коррекция орбиты, а процесс вывода является случайным процессом.

Такими же чертами обладает процесс полета пассажирского самолета. Из-за турбулентных потоков в атмосфере Земли и других случайных факторов полет сопровождается возмущениями, колебаниями, которые отчетливо ощущаются пассажирами как «болтанка» полета.

Труднее привести пример «неслучайного» процесса, чем случайного. Даже такая характеристика точной, выверенной работы, как «работает, как часы» подвержена сомнению, поскольку любые часы либо уходят вперед, либо отстают. Пока эти возмущения несущественны, мы можем ими пренебречь и рассматривать процесс как неслучайный.

Случайный процесс, протекающий в некоторой системе, называется **марковским** [3], если для любого момента времени вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент и не зависят от того, когда и как система пришла в это состояние. Поэтому для марковского случайного процесса легче предсказывать будущее, так как для этого достаточно учитывать только настоящее состояние, забыв «предысторию» процесса.

На практике марковские процессы в чистом виде обычно не встречаются. Но часто приходится иметь дело с процессами, для которых влиянием «предыстории» можно пренебречь. При изучении таких процессов можно с успехом применять марковские модели.

Большое значение имеют так называемые марковские процессы с дискретными состояниями и непрерывным временем. Процесс называется **процессом с дискретными состояниями**, если его возможные состояния можно заранее перечислить, и переход системы из состояния в состояние происходит «скачком», практически мгновенно. Процесс называется **процессом с непрерывным временем**, если моменты возможных переходов из состояния в состояние не фиксированы заранее, а неопределенны, т. е. переход может осуществляться в любой момент времени. Примером такого процесса является техническое устройство, состоящее из двух узлов, каждый из которых в случайный момент может выйти из строя, после чего мгновенно начинается ремонт узла. Ремонт тоже продолжается в случайное заранее неизвестное время. Перечислим все возможные состояния устройства: S_0 – оба узла исправны; S_1 – первый узел ремонтируется, а второй исправен; S_2 – второй узел ремонтируется, а первый исправен; S_3 – оба узла ремонтируются. Переходы системы из состояния в состояние происходят практически мгновенно.

Графом состояний называется геометрическая схема, в которой возможные состояния изображены прямоугольниками, соединенными стрелками возможных переходов. На рисунке 3.1 изображен граф для рассмотренного выше примера.

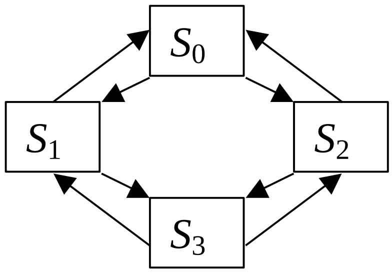

Рисунок 3.1 – Граф системы

Стрелка, направленная из S_0 в S_1 означает переход в момент отказа первого узла; стрелка, направленная обратно, означает переход в момент окончания ремонта этого узла. Остальные стрелки объясняются аналогично.

Потоком событий называется последовательность однородных событий, следующих друг за другом в случайные моменты времени. Например, поток покупателей в магазин; поток железнодорожных составов, поступающих на сортировочную станцию; поток вызовов неотложной скорой помощи и т. д.

На рисунке 3.2 изображена какая-то одна реализация потока.

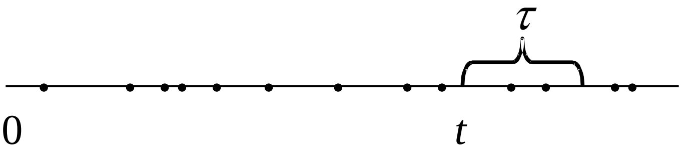

Рисунок 3.2 – Реализация потока случайных событий

Поток событий можно наглядно изобразить рядом точек на оси времени, положение каждой из которых случайно. Говоря о потоке событий нужно иметь в виду, что, в отличие от теории вероятностей, где каждое событие имело некоторую вероятность, события, образующие поток, сами по себе вероятностями не обладают. Вероятности рассматриваются для других, производных от них событий. Например, вероятность того, что, начиная с момента времени t за время длительностью τ, произойдут два события, обозначается символом $P_\tau(t, 2)$.

Важной характеристикой потока событий является его **интенсивность** $\boldsymbol{\lambda}$ – среднее число событий, происходящих за единицу времени. Интенсивность потока может быть как постоянной, так и переменной, зависящей от времени. Например, поток автомашин, двигающихся по улице, днем интенсивнее, чем ночью.

Поток событий называется **стационарным**, если его вероятностные характеристики не зависят от времени t, т. е. $P_\tau(t, 2) = P_\tau(2)$. В частности интенсивность стационарного потока должна быть постоянной. Это отнюдь не значит, что фактическое число событий, появляющееся в единицу времени, постоянно. Нет, поток неизбежно имеет какие-то случайные сгущения и разряжения (рисунок 3.2). Важно, что для стационарного потока эти сгущения и разряжения не носят закономерного характера: на один участок длины 1 может попасть больше, а на другой – меньше событий, но среднее число событий, приходящееся на единицу времени, постоянно и от времени не зависит.

Как правило, отклонения от стационарности имеют физические причины. Если поток событий имеет тенденцию к явно выраженным сгущениям и разряжениям (особенно периодическим), нужно заподозрить физическую причину и постараться ее выявить.

Поток событий называется **потоком без последействия**, если число событий, попадающих на любой участок времени τ (рисунок 3.2), не зависит от того, сколько событий попало на предыдущие участки времени. Например, поток покупателей в магазин, расположенный в густонаселенном районе, практически не имеет последействия. А вот поток студентов, идущих на экзамен, уже имеет последействие. Если до часу дня на экзамен уже пришли все студенты, то после часа больше никто не придет.

Поток событий называется **ординарным**, если события в нем появляются поодиночке, а не группами по нескольку сразу. Например, поток больных, направляющихся к зубному врачу, обычно ординарен, чего нельзя сказать о молодоженах, направляющихся в загс для регистрации брака. Поток поездов, подходящих к станции, ординарен, а поток вагонов – неординарен. Если поток событий ординарен, то вероятностью попадания на малый участок времени $\Delta\tau$ двух или более событий можно пренебречь.

Поток событий называется **простейшим** (пуассоновским), если он обладает сразу тремя свойствами: стационарен, ординарен и не имеет последействия. Название «простейший» связано с тем, что процессы, моделируемые простейшими потоками, имеют наиболее простое математическое описание.

Простейший поток играет среди других потоков особую роль, в чем-то подобную роли нормального распределения среди других законов распределения: при сложении достаточно большого числа независимых стационарных и ординарных потоков (сравнимых между собой по интенсивности) получается поток, близкий к простейшему. Причем тем ближе, чем больше число слагаемых потоков.

Легко доказать, что для простейшего потока с интенсивностью λ вероятность того, что за время длительностью τ произойдут k событий, вычисляется по **формуле Пуассона** $P_\tau(k) = \frac{(\lambda\tau)^k}{k!} e^{-(\lambda\tau)}$, а интервал времени T между соседними двумя событиями имеет **показательное распределение** с плотностью $f(t) = \lambda e^{-(\lambda\tau)}$. Математичес-кое ожидание случайной величины T, есть величина, обратная интенсивности λ, $M(T) = 1/\lambda$. Другими словами, среднее число событий, происходящих в единицу времени, обратно среднему интервалу времени между соседними событиями.

В расчетах, связанных с потоками событий, очень удобно пользоваться понятием «элемента вероятности». Рассмотрим на оси времени простейший поток с интенсивностью λ и произвольно расположенный элементарный (очень маленький!) участок времени $\Delta\tau$. **Элементом вероятности** называется вероятность попадания на этот участок хотя бы одного события потока. Легко доказать, что элемент вероятности (с точностью до величин более высокого порядка малости по сравнению с $\Delta\tau$) равен $\lambda\Delta\tau$, т. е. для простейшего потока элемент вероятности равен интенсивности потока, умноженной на длину элементарного участка. Очевидно, что элемент вероятности, в силу отсутствия последействия, совершенно не зависит от того, сколько событий и когда появились ранее.

3.2 Уравнения Колмогорова для вероятностей состояний

Рассмотрим марковский процесс для некоторого устройства, которое переходит из состояния в состояние под действием каких-то потоков событий. Например, поток отказов узлов, поток восстановлений этих узлов. Если все потоки событий, переводящие устройство из состояния в состояние, – простейшие, то процесс, протекающий в устройстве, будет марковским. Действительно, простейший поток не обладает последействием. Поэтому в нем «будущее» не зависит от «прошлого».

Если устройство находится в состоянии S_i, из которого есть непосредственный переход в состояние S_j, то мы себе это будем представлять так, как будто на устройство, пока оно находилось в состоянии S_i, действует простейший поток событий, переводящий его по стрелке $S_i \to S_j$. Как только появится первое событие этого потока, происходит «перескок» устройства из S_i в S_j. Для наглядности очень удобно на графе состояний у каждой стрелки проставлять интенсивность того потока событий, который переводит устройство по данной стрелке. Такой граф называется **размеченным**.

Построим размеченный граф состояний для устройства из двух узлов (рисунок 3.1). Интенсивность потоков событий, переводящих устройство из состояния в состояние, будем вычислять, предполагая, что среднее время ремонта узла не зависит от того, ремонтируется ли один узел или оба сразу. Найдем все интенсивности потоков событий, переводящих устройство из состояния в состояние. Пусть устройство находится в состоянии S_0. Какой поток событий переводит его в состояние S_1? Очевидно, поток отказов первого узла. Его интенсивность λ_1 обратна среднему времени безотказной работы первого узла. Какой поток событий переводит его обратно из S_1 в S_0? Очевидно, поток «окончаний ремонтов» первого узла. Его интенсивность μ_1 обратна среднему времени ремонта первого узла. Аналогично вычисляются интенсивности потоков событий, переводящих устройство по всем стрелкам графа на рисунке 3.3.

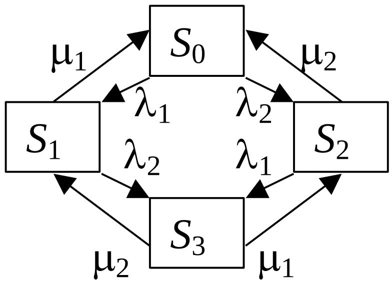

Рисунок 3.3 – Размеченный граф состояний устройства

Имея в распоряжении размеченный граф состояний рассматриваемого устройства, легко построить математическую модель данного процесса. В самом деле, назовем **вероятностью *i*-го состояния** вероятность $p_i(t)$ того, что в момент t устройство будет находиться в состоянии S_i. Очевидно, что для любого момента времени сумма вероятностей всех состояний равна единице. Кроме этого, вероятности состояний удовлетворяют системе дифференциальных уравнений Колмогорова, отражающих размеченный граф состояний. Покажем на примере рассматриваемого устройства, как составляется это уравнение для первого узла.

Пусть $p_1(t)$ – вероятность того, что в момент t устройство будет находиться в состоянии S_1. Придадим t малое приращение Δt и найдем $p_1(t + \Delta t)$ – вероятность того, что в момент $t + \Delta t$ устройство будет находиться в состоянии S_1. Как это может произойти? Очевидно, двумя способами: либо 1) в момент t устройство уже было в состоянии S_1, а за время Δt не вышло из него; либо 2) в момент t устройство было в состояниях S_0 или S_3, а за время Δt перешло в состояние S_1. Найдем вероятность первого варианта.

Суммарный поток событий, выводящий устройство из состояния S_1, тоже будет простейшим, с интенсивностью $\lambda_2 + \mu_1$. Значит, вероятность того, что элемент выйдет из состояния S_1, равна элементу вероятности этого потока $(\lambda_2 + \mu_1)\,\Delta t$; вероятность противоположного события, что не выйдет из состояния S_1: $1 - (\lambda_2 + \mu_1)\,\Delta t$. Вероятность того, что в момент t устройство было в состоянии S_1, равна $p_1(t)$. Согласно теореме умножения вероятностей эту вероятность нужно умножить на вероятность того, что, находившись в момент t в состоянии S_1, устройство за время Δt не перейдет из него в другое состояние, то есть на $1 - (\lambda_2 + \mu_1)\ \Delta t$. Таким образом, вероятность первого варианта равна $p_1(t)[1-(\lambda_2+\mu_1)\Delta t]$. Найдем вероятность второго варианта.

Она равна сумме вероятностей того, что, находившись в момент t в состояниях S_0 или S_3, устройства за время Δt перешло в состояние S_1, т. е. она равна $p_0(t)\lambda_1\Delta t + p_3(t)\mu_2\Delta t$.

Складывая вероятности обоих вариантов, согласно формуле полной вероятности получим уравнение

$$p_1(t+\Delta t) = p_1(t)[1-(\lambda_2+\mu_1)\Delta t] + p_0(t)\lambda_1\Delta t + p_3(t)\mu_2\Delta t.$$

Раскроем квадратные скобки, перенесем $p_1(t)$ в левую часть и разделим обе части на Δt:

$$\frac{p_1(t+\Delta t)-p_1(t)}{\Delta t} = p_0(t)\lambda_1 + p_3(t)\mu_2 - p_1(t)(\lambda_2+\mu_1).$$

Устремим Δt к нулю; слева получим в пределе производную функции $p_1(t)$. Таким образом, получается дифференциальное уравнение для $p_1(t)$:

$$\frac{dp_1(t)}{dt} = p_0(t)\lambda_1 + p_3(t)\mu_2 - p_1(t)(\lambda_2+\mu_1).$$

Рассуждая аналогично для всех остальных состояний устройства, получим еще три дифференциальных уравнения.

Сформулируем теперь общее правило составления уравнений Колмогорова. В левой части каждого из них стоит производная вероятности какого-то 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 устройство из данного состояния*, умноженная на вероятность данного состояния.

Уравнения Колмогорова дают возможность найти все вероятности состояний как функции времени. Поставим вопрос иначе: что будет происходить с вероятностями состояний при $t \to \infty$? В теории случайных процессов доказывается, что в случае конечного числа состояний их вероятности имеют пределы, которые называются финальными вероятностями состояний.

Как понимать эти финальные вероятности? Финальную вероятность состояния S_i можно истолковать как среднее относительное время пребывания устройства в этом состоянии.

Финальные вероятности обозначаются теми же буквами, что и сами вероятности состояний, но понимая под ними уже не переменные величины, а постоянные числа.

Как же вычислить финальные вероятности? Очень просто. Если вероятности постоянны, то их производные равны нулю. Значит, чтобы найти финальные вероятности, нужно все левые части в уравнениях Колмогорова занулить и решить полученную систему уже не дифференциальных, а линейных алгебраических уравнений. Можно и не писать уравнений Колмогорова, а прямо по графе состояний написать систему линейных алгебраических уравнений. Если перенести отрицательные члены каждого уравнения из правой части в левую, то получим сразу систему уравнений, где слева стоит финальная 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.

Пользуясь этим правилом, составим уравнения для финальных вероятностей состояний устройства, граф состояний которого приведен на рисунке 3.3:

$$\left.\begin{aligned}(\lambda_1+\lambda_2)p_0 &= \mu_1 p_1 + \mu_2 p_2,\\ (\lambda_2+\mu_1)p_1 &= \lambda_1 p_0 + \mu_2 p_3,\\ (\lambda_1+\mu_2)p_2 &= \lambda_2 p_0 + \mu_1 p_3,\\ (\mu_1+\mu_2)p_3 &= \lambda_2 p_1 + \lambda_1 p_2.\end{aligned}\right\}$$

Полученные уравнения являются линейно зависимыми, т. е. одно из них можно удалить, а вместо него необходимо добавить нормировочное условие равенства единице суммы финальных вероятностей всех состояний:

$$p_0 + p_1 + p_2 + p_3 = 1.$$

▪ *Пример* 1. Найти финальные вероятности всех состояний рассмотренного выше устройства со следующими значениями интенсивностей потоков: $\lambda_1 = 1$, $\lambda_2 = 2$, $\mu_1 = 2$, $\mu_2 = 3$.

Решение. Заменим четвертое уравнение системы нормировочным условием. При заданных значениях интенсивностей они примут вид:

$$\left.\begin{aligned} 3p_0 &= 2p_1 + 3p_2, \\ 4p_1 &= p_0 + 3p_3, \\ 4p_2 &= 2p_0 + 3p_3, \\ p_0 + p_1 + p_2 + p_3 &= 1. \end{aligned}\right\}$$

Решая полученную систему, получим:

$p_0 = 6/15 = 0{,}4$; $p_1 = 3/15 = 0{,}2$; $p_2 = 4/15 \approx 0{,}27$; $p_3 = 2/15 \approx 0{,}13$,

Полученное решение имеет следующий смысл: в предельном режиме устройство в среднем 40 % времени будет находиться в состоянии S_0, когда оба узла исправны, 20 % – в состоянии S_1, когда первый узел ремонтируется, а второй работает, 27 % – в состоянии S_2, когда второй узел ремонтируется, а первый работает и 13 % – в состоянии S_3 полной негодности, когда оба узла ремонтируются.

Знание этих финальных вероятностей помогает оценить эффективность работы устройства. Пусть устройство в состоянии S_0, когда оба узла исправны, приносит в единицу времени доход C_0 денежных единиц, в состояниях S_1, S_2 – C_1, C_2 денежных единиц соответственно, а в состоянии S_3 – вообще не приносит дохода. Тогда средний доход в единицу времени будет

$$W = 0{,}4\ C_0 + 0{,}2\ C_1 + 0{,}27\ C_2\ .$$

3.3 Классификация СМО и типов управления

Системы массового обслуживания делятся на типы по ряду признаков. Первое деление: СМО **с отказами** и СМО **с очередью**. В СМО с отказами заявка, поступившая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем процессе обслуживания не участвует. Примеры СМО с отказами встречаются в телефонии, где короткие гудки красноречиво говорят о том, что в этой ситуации ждать чего-то бесполезно. В СМО с очередью заявка, пришедшая в момент, когда все каналы заняты, не уходит, а становится в очередь и ожидает возможности быть обслуженной. На практике чаще встречаются СМО с очередью. Недаром теория массового обслуживания имеет второе название: «теория очередей».

СМО с очередью подразделяются на разные виды в зависимости от того, как организована очередь – ограничена она или не ограничена. Ограничения могут касаться как длины очереди, так и времени ожидания (так называемые «СМО с нетерпеливыми заявками»).

При анализе СМО должна также учитываться «дисциплина обслуживания» – заявки могут обслуживаться либо в порядке поступления (раньше пришла – раньше обслуживается), либо в случайном порядке. Нередко встречается так называемое **обслуживание с приоритетом** – некоторые заявки обслуживаются вне очереди. Приоритет может быть как **абсолютным** – когда заявка с более высоким приоритетом «вытесняет» из-под обслуживания заявку с низшим (например, пришедший в парикмахерскую клиент высокого ранга прогоняет с кресла обыкновенного недобритого клиента), так и относительным – когда начатое обслуживание доводится до конца, а заявка с более высоким приоритетом имеет лишь право на лучшее место в очереди.

Существуют СМО с так называемым **многофазовым** обслуживанием, состоящим из нескольких последовательных этапов или «фаз» (например, покупатель, пришедший в магазин, должен сначала выбрать товар, затем оплатить в кассе, затем получит его на контроле).

Кроме этих признаков, СМО делятся на два класса: «открытые» и «замкнутые». В **открытой** СМО характеристики потока заявок не зависят от того, в каком состоянии сама СМО (сколько каналов занято). В **замкнутой** СМО – зависят. Например, если один рабочий обслуживает группу станков, время от времени требующих наладки, то интенсивность потока «требований» со стороны станков зависит от того, сколько станков уже неисправно и ждет наладки. Действительно, если почти все станки вышли из строя, то интенсивность потока заявок падает до нуля.

Управление СМО может производиться под разными углами зрения: с точки зрения организаторов (или владельцев) СМО или с точки зрения обслуживаемых клиентов. С первой точки зрения желательно «выжать все, что возможно» из СМО и добиться того, чтобы ее каналы были предельно загружены. С точки зрения клиентов желательно всемерное уменьшение очередей, которые зачастую становятся настоящим «бичом быта», приводя к бессмысленной трате сил и времени и, в конечном итоге, к понижению производительности труда. Поэтому необходим системный подход к управлению СМО, обеспечивающий полноту рассмотрения всех последствий каждого решения. Например, с точки зрения клиентов СМО желательно увеличение числа каналов обслуживания; но ведь работу каждого канала надо оплачивать, что удорожает обслуживание. Построение математической модели позволяет выбрать разумное число каналов с учетом всех «за» и «против».

3.4 Многоканальная СМО с отказами

Рассмотрим одну из первых по времени, «классических» задач, возникшую из практических нужд телефонии. Эта задача была решена в начале прошлого века датским математиком Эрлангом. Однако и сегодня эта задача не утратила актуальности. Например, диспетчерский пункт приема срочных вызовов по телефону 911 в Нью-Йорке обслуживает поток заявок с интенсивность 840 вызовов в час, причем каждый вызов обслуживается в среднем 80 секунд. Если все операторы заняты обслуживанием этих вызовов, то система 911 бесполезна. Поэтому штат операторов и соответствующее количество линий связи в этом пункте подбираются из условия, чтобы вероятность того, что все операторы будут заняты более 20 секунд, не превосходила 5 %.

- *Задача Эрланга*. Имеется n линий связи [13], которые называются каналами обслуживания и на которые поступает поток заявок с интенсивностью λ. Поток обслуживаний имеет интенсивность μ. Найти финальные вероятности системы, а также характеристики ее эффективности:

A – **абсолютную пропускную способность**, то есть среднее число заявок, обслуживаемых в единицу времени;

q – **относительную пропускную способность**, то есть среднюю долю пришедших заявок, обслуживаемых системой;

$P_{\text{отк}}$ – **вероятность отказа**, то есть того, что заявка покинет систему необслуженной;

$\bar{k}$ – **среднее число занятых каналов**.

Решение. Состояния системы S_0, S_1, … , S_n нумеруются по числу заявок, находящихся в системе. Это число совпадает с количеством занятых каналов. Размеченный граф системы приведен на рисунке 3.4.

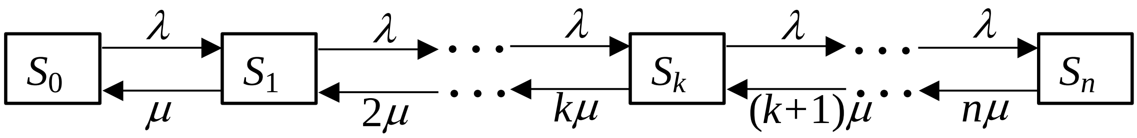

Рисунок 3.4 – Размеченный граф n канальной системы с отказами

Действительно, пусть система находится в состоянии S_1 (работает один канал). Он производит μ обслуживаний в единицу времени. Поэтому нижняя стрелка, соединяющая состояния S_0 и S_1, отмечена интенсивностью μ. Теперь представим себе, что система находится в состоянии S_2 (работают два канала). Чтобы ей перейти в S_1, нужно, чтобы либо закончил обслуживание первый канал, либо второй; суммарная интенсивность их потоков обслуживаний равна 2μ, т. к. каналы обслуживают заявки одновременно. Поэтому нижняя стрелка, соединяющая состояния S_1 и S_2, отмечена этой интенсивностью. Суммарный поток обслуживаний, даваемый k каналами, имеет интенсивность $k\mu$, n каналами – $n\mu$. Отмечаем нижние стрелки соответствующими интенсивностями.

Пользуясь графом рисунка 3.4, составим алгебраические уравнения для финальных вероятностей состояний. Их существование вытекает из того, что из каждого состояния можно перейти в каждое другое, и число состояний конечно.

$$\left.\begin{array}{l}\lambda p_0 = \mu p_1,\\ (\lambda+\mu)p_1 = \lambda p_0 + 2\mu p_2,\\ .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\\ (\lambda+k\mu)p_k = \lambda p_{k-1} + (k+1)p_{k+1},\\ .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\\ n\mu p_n = \lambda p_{n-1},\\ p_0 + p_1 + \ldots + p_k + \ldots + p_n = 1.\end{array}\right\}$$

Для решения этой системы из первого уравнения выразим p_1 через p_0:

$$p_1 = \frac{\lambda}{\mu} p_0 .$$

Из второго уравнения, с учетом первого уравнения и уже найденного выражения, получим:

$$p_2 = \frac{\lambda}{2\mu} p_1 = \frac{\lambda^2}{2\mu^2} p_0 .$$

Таким образом, для любого k (от 1 до n)

$$p_k = \frac{\lambda^k}{k!\,\mu^k} p_0 .$$

Поэтому все вероятности $p_0, p_1, \ldots, p_n$ выражаются через одну из них (p_0). Подставим эти вероятности в нормировочное условие и вынесем за скобку p_0 :

$$p_0\left(1+\frac{\lambda}{\mu}+\ldots+\frac{\lambda^k}{k!\,\mu^k}+\ldots+\frac{\lambda^n}{n!\,\mu^n}\right)=1,$$

откуда получим выражение для p_0 :

$$p_0 = \left(1+\frac{\lambda}{\mu}+\ldots+\frac{\lambda^k}{k!\,\mu^k}+\ldots+\frac{\lambda^n}{n!\,\mu^n}\right)^{-1}.$$

Заметим, что в выражениях для финальных вероятностей интенсивности λ и μ входят не по отдельности, а только в виде отношения λ/μ. Обозначим

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

и будем называть величину ρ «**траффик-интенсивностью** потока заявок». Ее смысл – среднее число заявок, приходящих за среднее время обслуживания одной заявки. Пользуясь этим обозначением, получим финальные вероятности состояний в виде:

$$p_0 = \left(1 + \rho + \ldots + \frac{\rho^k}{k!} + \ldots + \frac{\rho^n}{n!}\right)^{-1},$$
$$p_k = \frac{\rho^k}{k!} p_0, \quad k = \overline{1,n} \quad . \tag{3.1}$$

Полученные выражения называются **формулами Эрланга** – в честь основателя теории массового обслуживания [13]. Большинство других формул этой теории (сегодня их больше, чем грибов в лесу) не носят никаких специальных имен.

Займемся теперь характеристиками эффективности системы. Сначала найдем $P_{\text{отк}}$ – вероятность того, что пришедшая заявка получит отказ и покинет систему необслуженной. Для этого нужно, чтобы все каналы были заняты, значит,

$$P_{\text{отк}} = p_n = \frac{\rho^n}{n!} p_0. \tag{3.2}$$

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

$$q = 1 - P_{\text{отк}} = 1 - \frac{\rho^n}{n!} p_0.$$

Абсолютную пропускную способность получим, умножая интенсивность потока заявок λ на q

$$A = \lambda q = \lambda\left(1 - \frac{\rho^n}{n!} p_0\right).$$

Для расчета среднего числа занятых каналов заметим, что абсолютную пропускную способность можно интерпретировать как интенсивность потока **обслуженных системой** заявок. Каждый занятый канал в единицу времени обслуживает в среднем μ заявок. Значит, среднее число занятых каналов равно

$$\bar{k} = \frac{A}{\mu} = \rho\left(1 - \frac{\rho^n}{n!} p_0\right). \tag{3.3}$$

▪ *Пример* 1. Диспетчерский пункт приема срочных вызовов по телефону 911 в Нью-Йорке обслуживает поток заявок с интенсивность 840 вызовов в час, причем каждый вызов обслуживается в среднем 80 секунд. Необходимо определить количество каналов обслуживания, при которых вероятность отказа меньше 5 %.

Уровень суммарных потерь связан с простоем среднего числа операторов и средним количеством необслуженных вызовов в час. Определить оптимальное число операторов, при которых суммарные потери будут минимальными.

Решение. Дано: λ = 840 выз/час; $\overline{t_{\text{об}}}$ = 0,0(2) час/выз. Необходимое количество каналов обслуживания, при которых вероятность отказа меньше 5 %, является натуральным решением неравенства:

$$P_{\text{отк}} < 0{,}05,$$

в котором вероятность отказа определяется с помощью формул (3.1), (3.2), а величина траффик-интенсивности вычисляется по формуле

$$\rho = \lambda \overline{t_{\text{об}}} = 18{,}(6).$$

В таблице 3.1 приведены результаты расчета вероятности отказа в зависимости от числа каналов обслуживания.

Таблица 3.1

Число n	10	20	23	24	30	31	32
$P_{\text{отк}}$	0,5092	0,1255	0,0599	0,0445	0,004	0,0024	0,0014
W	428,53	109,08	55,74	43,54	14,78	14,41	14,54

Если число каналов n больше 23-х, то вероятность отказа меньше 5 %. Уровень суммарных потерь $W = n_{\text{св}} + v_{\text{необс}}$, где $n_{\text{св}}$ – среднее число простаивающих операторов, а $v_{\text{необс}}$ – среднее число необслуженных вызовов в час. Среднее число простаивающих операторов рассчитывается по формуле: $n_{\text{св}} = n - \overline{k}$, где среднее число занятых каналов $\overline{k}$ определяется согласно (3.3). Среднее число необслуженных вызовов в час рассчитывается по формуле: $v_{\text{необс}} = \lambda\, P_{\text{отк}}$. Таким образом, уровень суммарных потерь имеет вид:

$$W = n - \rho + (\lambda + \rho)\frac{\rho^n}{n!}\, p_0.$$

Результаты расчета уровня суммарных потерь (см. таблицу 3.1) показывают, что оптимальное число операторов равно 31.

3.5 Одноканальная СМО с ожиданием и неограниченной очередью

На практике часто встречаются системы с ожиданием и неограниченной очередью. В этих системах, как правило, может находиться большое число заявок. Например, врач, обслуживающий пациентов или преподаватель, принимающий экзамен у студентов. Эти системы считаются наиболее изученными. Поэтому этим системам уделяется особое внимание.

В большинстве СМО имеется несколько обслуживающих каналов, а дисциплина очереди, как правило, оказывается весьма сложной. Так, например, прежде чем направиться к какому-нибудь контрольно-расчетному прилавку в супермаркете, покупатель вначале посмотрит, сколько человек стоит в каждой очереди и какое количество различных продуктов находится в руках у стоящих. Кроме того, он попытается сообразить, какой из контрольно-расчетных прилавков находится ближе всего к нужному ему выходу, и мысленно отметит, какая из кассирш работает быстрее других. Но иногда действительно имеет место строго соблюдаемая дисциплина «первым пришел – первым обслуживаешься». Такого характера очереди возникают, например, на бензозаправочных станциях, возле касс кинотеатров, в мастерских, где производится срочный ремонт обуви, и т. п.

Следует отметить, что построение операционной модели для описания реальной ситуации всегда сопряжено с необходимостью принятия ряда упрощающих предположений. Часто удается получить приближенное представление об операционных характеристиках сложной системы путем анализа некоторых»экстремальных» или предельных случаев. Один из таких приближенных методов заключается в следующем: СМО, насчитывающая n обслуживающих каналов, рассматривается как «механическое» объединение n одноканальных систем, функционирующих независимо одна от другой. Пусть, например, обслуживающая система состоит из пяти каналов, а интенсивность входного потока равняется 20 человек/час. Тогда приближенно данную систему можно рассматривать как совокупность пяти автономных систем с одним обслуживающим каналом, каждая из которых характеризуется входным потоком с интенсивностью 4 человека/час.

Этот метод является *приближенным* по двум причинам: во-первых, предполагается, что заявка может попасть на вход любой из упомянутых одноканальных «подсистем» с одинаковой вероятностью (независимо от того, какая длина соответствующей очереди), и, во-вторых, постулируется, что, попав в одну из очередей, заявка *должна* оставаться именно в этой первоначально выбранной очереди. Ожидаемое количество заявок, находящихся во всех автономных подсистемах такого рода гипотетической системы, и среднее время пребывания в ней заявки, как правило, превышают значения соответствующих операционных характеристик реальных многоканальных систем. Это объясняется тем, что в нашей гипотетической системе часто возникает ситуация, когда один из обслуживающих каналов находился в незанятом состоянии, в то время как заявки простаивали бы в очередях в других каналах.

- <u>*Постановка задачи*</u>. Пусть имеется одноканальная система с очередью, на которую не наложено никаких ограничений ни по длине очереди, ни по времени ожидания. На эту СМО поступает поток заявок с интенсивностью λ; поток обслуживания имеет интенсивность μ, обратную среднему времени обслуживания заявки $\overline{t_{\text{об}}}$. Требуется найти финальные вероятности состояний СМО, а также характеристики ее эффективности:

$\bar{k}$ – среднее число заявок в системе,

$\overline{t_{\text{пр}}}$ – среднее время пребывания заявки в системе,

$\overline{l_{\text{оч}}}$ – среднюю длину очереди,

$\overline{t_{\text{ож}}}$ – среднее время пребывания заявки в очереди,

$P_{\text{зан}}$ – вероятность того, что канал занят (степень загрузки канала).

Что касается абсолютной и относительной пропускных способностей, то вычислять их нет надобности: в силу того, что очередь неограничена, каждая заявка рано или поздно будет обслужена. Поэтому $A = \lambda$, по той же причине $q = 1$.

Решение. Состояния системы S_0, S_1, … , S_n, … , как и ранее, нумеруются по числу заявок, находящихся в СМО. Размеченный граф системы приведен на рисунке 3.5.

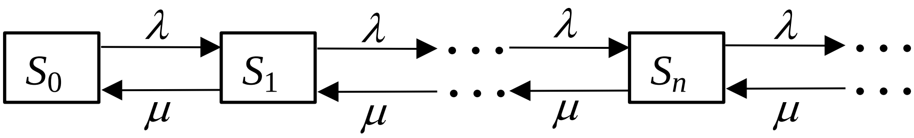

Рисунок 3.5 – Размеченный граф одноканальной системы с неограниченной очередью

Возникает вопрос: существуют ли в этом случае финальные вероятности состояний? Ведь число состояний бесконечно и, в принципе, при $t \to \infty$ очередь может неограниченно возрастать. Действительно, можно доказать, что если $\rho < 1$, то финальные вероятности существуют, а при $\rho \geq 1$ очередь растет неограниченно.

Следует отметить, что граф на рисунке 3.5 отличается от графа на рисунке 3.4 интенсивностями потоков обслуживания заявок и бесконечным числом состояний. Поэтому можно воспользоваться полученными ранее выражениями (3.1) для финальных вероятностей состояний, изменив эти интенсивности:

$$p_0 = (1 + \rho + ... + \rho^n + ...)^{-1},$$
$$p_k = \rho^k p_0, \quad k = \overline{1, \infty} \quad .$$

Полученные вероятности образуют бесконечную сходящуюся геометрическую прогрессию. Поэтому можно использовать формулу суммы членов этой прогрессии: $p_0 = 1 - \rho$, значит

$$p_k = \rho^k (1 - \rho), \quad k = \overline{0, \infty}.$$

Рассмотрим характеристики эффективности системы. Найдем среднее число заявок в СМО, которое определяется как математическое ожидание случайного числа заявок $\overline{k} = \sum_{k=0}^{\infty} k p_k$. Подставим в эту формулу выражения для p_k:

$$\overline{k} = \sum_{k=0}^{\infty} k \rho^k (1-\rho).$$

Теперь вынесем за знак суммы $\rho(1-\rho)$:

$$\overline{k} = \rho(1-\rho)\sum_{k=0}^{\infty} k \rho^{k-1}.$$

Применим свойство степенных рядов, которые можно дифференцировать почленно по переменной ρ в интервале сходимости. Значит

$$\overline{k} = \rho(1-\rho)\frac{d}{d\rho}\sum_{k=0}^{\infty} \rho^k.$$

Снова воспользуемся формулой суммы членов бесконечной геометрической прогрессии и вычислим производную, значит:

$$\overline{k} = \rho(1-\rho)\frac{d}{d\rho}(1-\rho)^{-1} = \rho(1-\rho)(1-\rho)^{-2} = \frac{\rho}{1-\rho}.$$

Найдем теперь среднюю длину очереди, которая определяется как математическое ожидание случайного числа заявок, стоящих в очереди $\overline{l_{\text{оч}}} = \sum_{k=1}^{\infty} (k-1) p_k$. Однако поступим проще: число заявок, стоящих в очереди, равно числу заявок в системе минус число заявок, находящихся под обслуживанием. По правилу вычитания математических ожиданий среднее число заявок, стоящих в очереди, равно среднему числу заявок в системе минус среднее число заявок под обслуживанием. Число заявок под обслуживанием может быть либо нулем (если канал свободен), либо единицей (если он занят). Математическое ожидание такой величины равно вероятности того, что канал занят. Обозначим ее $P_{\text{зан}}$ и вычислим как единица минус вероятность того, что канал свободен:

$$P_{\text{зан}} = 1 - p_0 = \rho.$$

Следовательно, среднее число заявок под обслуживанием равно ρ, а средняя длина очереди

$$\overline{l_{\text{оч}}} = \bar{k} - \rho = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho} .$$

Рассмотрим **формулы Литтла.** Первая формула Литтла связывает среднее число заявок $\bar{k}$, находящихся в СМО, и среднее время пребывания заявки в СМО $\overline{t_{\text{пр}}}$. Рассматривается любая СМО (одноканальная, многоканальная, марковская, немарковская, с неограниченной или с ограниченной очередью) и связанный с ней поток заявок, прибывающих в СМО с интенсивностью λ в предельном стационарном режиме. Пусть некоторая заявка прибывает в СМО в нулевой момент времени, который считается началом наблюдения. Так как поток заявок рассматривается в предельном режиме, то эта заявка проведет в СМО $\overline{t_{\text{пр}}}$ единиц времени и покинет СМО вместе с теми заявками, которые прибыли в СМО ранее. Если за единицу времени в СМО в предельном режиме прибывает λ заявок, то за время $\overline{t_{\text{пр}}}$ в СМО прибудет $\lambda\overline{t_{\text{пр}}}$ заявок. Все эти заявки будут оставаться в СМО в период наблюдения (0, $\overline{t_{\text{пр}}}$). Поэтому среднее число заявок в СМО имеет вид

$$\bar{k} = \lambda\overline{t_{\text{пр}}} .$$

Аналогично рассуждая, получим вторую формулу Литтла

$$\overline{l_{\text{оч}}} = \lambda\overline{t_{\text{ож}}} .$$

Используя формулы Литтла, найдем среднее время пребывания заявки в системе и в очереди

$$\overline{t_{\text{пр}}} = \frac{\rho}{\lambda(1-\rho)}, \quad \overline{t_{\text{ож}}} = \frac{\rho^2}{\lambda(1-\rho)}.$$

Таким образом, все характеристики эффективности СМО найдены.

▪ *Пример* 1. В стоматологическую поликлинику прибывает простейший поток больных с интенсивностью λ = 5 $^{\text{чел}}/_{\text{час}}$. Средняя продолжительность сеанса лечения составляет $\overline{t_{\text{об}}}$ = 0,5 $^{\text{час}}/_{\text{чел}}$. Уровень суммарных потерь связан с простоем среднего числа свободных стоматологов $n_{св}$ и пребывания среднего числа больных в очереди $L_{оч}$. Построить график зависимости $f(n) = n_{св} + L_{оч}$, где n – количество стоматологов, и определить оптимальное количество стоматологов, при котором уровень суммарных потерь минимален.

Решение. Используется модель СМО с неограниченной очередью. В этой модели среднее число больных, стоящих в очереди ко всем врачам имеет вид

$$L_{оч} = n\overline{l_{оч}} = \frac{n\rho^2}{1-\rho},$$

где траффик-интенсивность потока больных $\rho = \lambda\,\overline{t_{об}}/n$. Число свободных стоматологов

$$n_{св} = np_0 = n\,(1 - \rho).$$

Тогда функция суммарных потерь сводится к виду

$$f(n) = n - \lambda\overline{t_{об}} + \frac{(\lambda\overline{t_{об}})^2}{n - \lambda\overline{t_{об}}},$$

а график этой функции изображен на рисунке 3.6. Согласно этому рисунку функция $f(n)$ достигает минимума при $n = 5$. Поэтому в случае пяти работающих стоматологов реализуется оптимальная ситуация: уровень суммарных потерь минимален.

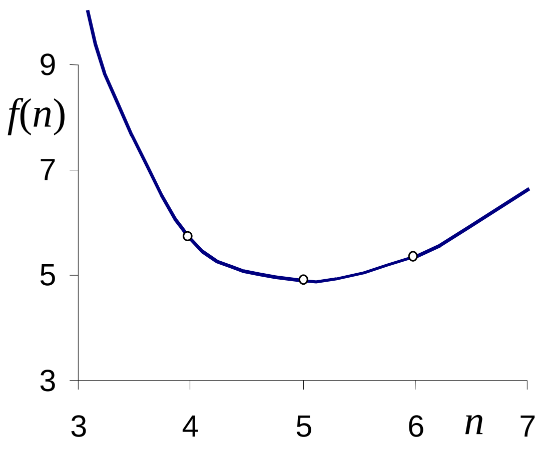

Рисунок 3.6 – Уровень суммарных потерь

- *Пример 2.* В стоматологическую поликлинику прибывает простейший поток больных с интенсивностью $\lambda = 5$ $^{чел}/_{час}$ в обычное время дня, в часы «пик» она возрастает до $\lambda_+ = 12$ $^{чел}/_{час}$, а в часы «спада» достигает величины $\lambda_- = 3$ $^{чел}/_{час}$. Средняя продолжительность сеанса лечения составляет $\overline{t_{об}}$ = 0,5 $^{час}/_{чел}$. Определить вероятность образования очереди $P_{оч}$ и среднюю длину очереди $\overline{l_{оч}}$ в течение дня, а также необходимое число стоматологов в часы «пик» n_+ и «спада» n_-, обеспечивающих такую же длину очереди и вероятность ее образования, как в обычное время дня.

<u>*Решение.*</u> Вероятность образования очереди определяется как единица минус вероятность отсутствия очереди

$$P_{оч} = 1 - (p_0 + p_1) = \rho^2,$$

а средняя длина очереди

$$\overline{l_{оч}} = \rho^2 / (1 - \rho),$$

где $\rho = \lambda\, \overline{t_{об}} / n$. Поэтому в течение дня согласно заданным значениям $\lambda_\pm$ и формулам получаются значения $P_\pm$, $l_\pm$. Для сохранения прежних значений $P_{оч}$, $\overline{t_{об}}$ необходимо сохранить значение ρ. Поэтому необходимое число стоматологов $n_\pm = n_0\,(\lambda_\pm / \lambda)$

$$\begin{pmatrix} P_{оч} \\ P_+ \\ P_- \end{pmatrix} = \begin{pmatrix} 0{,}25 \\ 1 \\ 0{,}09 \end{pmatrix}; \quad \begin{pmatrix} \overline{l_{оч}} \\ l_+ \\ l_- \end{pmatrix} = \begin{pmatrix} 0{,}5 \\ \infty \\ 0{,}13 \end{pmatrix}; \quad n_\pm = \begin{pmatrix} 12 \\ 3 \end{pmatrix}.$$

3.6 Одноканальная СМО с ожиданием и ограниченной очередью

Системы с ожиданием и ограниченной очередью встречаются в случаях, когда обслуживаются громоздкие конструкции (грузовая автомашина, пароход, железнодорожный состав), которые занимают довольно значительное место. Поэтому возникает естественное ограничение на суммарное количество заявок, которое может оказаться в обслуживающей системе в тот или иной момент времени.

Предположим, что в СМО может быть не более *M* заявок и, следовательно, в любой момент времени в очереди разрешается находиться не более чем (*M* – 1) заявкам (например, автозаправочная станция

(АЗС) с одной бензоколонкой, расположенная в узком переулке, выходящем на одну из центральных улиц города). Если заявка поступает в момент, когда в обслуживающей системе уже находится M объектов, то этой заявке отказывается в обслуживании и, таким образом, обслуживающая система ее теряет. По этой причине такого вида модели иногда называют моделями массового обслуживания с **вынужденными отказами**. Существенная разница между предыдущей моделью и моделью с конечной длиной очереди заключается в том, что в последнем случае статистическое равновесие достигается при любом значении траффик-интенсивности.

▪ *Постановка задачи*. Пусть имеется одноканальная система с очередью, на которую наложено ограничение: в очереди разрешается находиться не более чем (M – 1) заявкам. На эту СМО поступает простейший поток заявок с интенсивностью λ; поток обслуживания имеет интенсивность μ, обратную среднему времени обслуживания заявки $\overline{t_{об}}$. Требуется найти финальные вероятности состояний СМО, а также характеристики ее эффективности:

A – абсолютную пропускную способность;

q – относительную пропускную способность;

$P_{отк}$ – вероятность отказа, то есть того, что заявка покинет систему необслуженной;

$\bar{k}$ – среднее число заявок в системе;

$P_{зан}$ – вероятность того, что канал занят (степень загрузки канала);

$\overline{l_{оч}}$ – среднюю длину очереди;

$\overline{t_{пр}}$ – среднее время пребывания заявки в системе;

$\overline{t_{ож}}$ – среднее время ожидания заявки обслуживания.

Решение. Состояния системы S_0, S_1, … , S_k , … , S_M , как и ранее, нумеруются по числу заявок, находящихся в СМО. Размеченный граф системы приведен на рисунке 3.6.

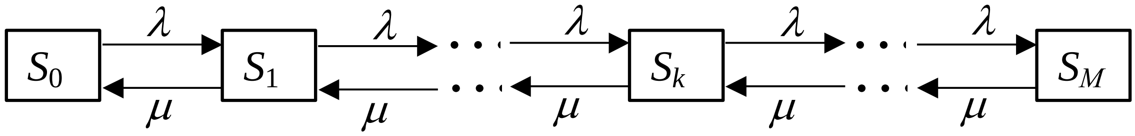

Рисунок 3.6 – Размеченный граф одноканальной системы с ограниченной очередью

Возникает вопрос: существуют ли в этом случае финальные вероятности состояний? Действительно, можно доказать, что при любом ρ финальные вероятности существуют.

Следует отметить, что граф на рисунке 3.6 отличается от графа на рисунке 3.5 конечным числом состояний. Поэтому можно воспользоваться полученными ранее выражениями (3.1) для финальных вероятностей состояний, в которых учитывается конечное число слагаемых

$$p_0 = (1 + \rho + \ldots + \rho^M)^{-1},$$
$$p_k = \rho^k p_0, \quad k = \overline{1, M} \quad .$$

Полученные вероятности образуют конечную геометрическую прогрессию. Поэтому можно использовать формулу суммы членов этой прогрессии

$$p_0 = \frac{\rho - 1}{\rho^{M+1} - 1} \text{ при } \rho \neq 1; \quad p_0 = \frac{1}{M+1} \text{ при } \rho = 1. \tag{3.4}$$

Значит

$$p_k = \rho^k p_0, \quad k = \overline{0, M}. \tag{3.5}$$

Рассмотрим характеристики эффективности системы. Начнем с вероятности отказа:

$$P_{\text{отк}} = p_M = \frac{\rho^M (\rho - 1)}{\rho^{M+1} - 1} \text{ при } \rho \neq 1; \quad P_{\text{отк}} = \frac{1}{M+1} \text{ при } \rho = 1. \tag{3.6}$$

Отсюда находим относительную и абсолютную пропускные способности: $q = 1 - P_{\text{отк}}$, $A = \lambda q$.

Найдем среднее число заявок в СМО, которое определяется как математическое ожидание случайного числа заявок $\bar{k} = \sum_{k=0}^{M} k p_k$. Подставим в эту формулу выражения (3.4), (3.5) для p_k при $\rho \neq 1$ и вынесем за знак суммы $\rho(\rho-1)$:

$$\bar{k}=\sum_{k=0}^{M}k\rho^{k}p_{0}=\frac{\rho(\rho-1)}{\rho^{M+1}-1}\sum_{k=0}^{M}k\rho^{k-1}.$$

Применим свойство степенных рядов, которые можно дифференцировать почленно по переменной ρ в интервале сходимости. Значит среднее число заявок в СМО

$$\bar{k}=\frac{\rho(\rho-1)}{\rho^{M+1}-1}\frac{d}{d\rho}\sum_{k=0}^{M}\rho^{k}.$$

Снова воспользуемся формулой суммы членов конечной геометрической прогрессии и вычислим производную, значит

$$\bar{k}=\frac{\rho(\rho-1)}{\rho^{M+1}-1}\frac{d}{d\rho}[(\rho-1)^{-1}(\rho^{M+1}-1)]=\frac{(M+1)\rho^{M+1}}{\rho^{M+1}-1}-\frac{\rho}{\rho-1}.$$

При $\rho = 1$ случайная величина k имеет равномерное распределение, математическое ожидание которой $\bar{k}=M/2$. В общем случае среднее число заявок в СМО имеет вид

$$\bar{k}=\begin{cases}\dfrac{(M+1)\rho^{M+1}}{\rho^{M+1}-1}-\dfrac{\rho}{\rho-1} & \text{при} \quad \rho\neq 1;\\ \dfrac{M}{2} \quad \text{при} \quad \rho=1.\end{cases} \tag{3.7}$$

Найдем теперь среднюю длину очереди, которая определяется как математическое ожидание случайного числа заявок, стоящих в очереди $\overline{l_{\text{оч}}}=\sum_{k=1}^{\infty}(k-1)p_{k}$. Однако, поступим так же, как в модели с неограниченной очередью: число заявок, стоящих в очереди, равно числу заявок в системе минус число заявок, находящихся под обслуживанием. По правилу вычитания математических ожиданий среднее число заявок, стоящих в очереди, равно среднему числу заявок в системе минус среднее число заявок под обслуживанием. Число заявок под обслуживанием может быть либо нулем (если канал свободен), либо единицей (если он занят). Математическое ожидание такой величины равно вероятности того, что канал занят. Обозначим ее $P_{\text{зан}}$ и вычислим как единица минус вероятность того, что канал свободен, т. е. p_0

согласно (3.4)

$$P_{\text{зан}} = 1 - p_0 = \begin{cases} \dfrac{\rho^{M+1} - \rho}{\rho^{M+1} - 1} & \text{при} \quad \rho \neq 1; \\ \dfrac{M}{M+1} & \text{при} \quad \rho = 1. \end{cases} \tag{3.8}$$

Следовательно, среднее число заявок под обслуживанием равно $P_{\text{зан}}$, а средняя длина очереди имеет вид

$$\overline{l_{\text{оч}}} = \bar{k} - \begin{cases} \dfrac{\rho^{M+1} - \rho}{\rho^{M+1} - 1} \text{ при } \rho \neq 1; \\ \dfrac{M}{M+1} \text{ при } \rho = 1. \end{cases} \tag{3.9}$$

Некоторые тонкости необходимо учитывать как при определении времени пребывания заявки в СМО, так и продолжительности ожидания в очереди. Если заявка поступает в тот момент, когда в СМО уже находится M других заявок, то она не может присоединиться к очереди, и, следовательно, время ее пребывания в СМО в буквальном смысле равняется нулю. Поэтому среднее время пребывания в СМО можно определить, либо учитывая все заявки независимо от того, имелась для них возможность присоединиться к очереди или нет, либо принимая в расчет только те заявки, которым вход в СМО был действительно разрешен. Мы примем второй вариант, поскольку в подавляющем большинстве случаев нас интересуют задержки в обслуживании только тех заявок, которые имели возможность войти в СМО. Тогда при условии, что заявка, поступающая в некоторый произвольно выбранный момент времени, получает право на ожидание в очереди, и что очередь характеризуется дисциплиной «первым пришел – первым обслуживаешься», для средней продолжительности пребывания данной заявки в системе получаем следующие выражения:

$$\overline{t_{\text{пр}}} = \begin{cases} \dfrac{M\rho^M}{\mu(\rho^M - 1)} - \dfrac{1}{\mu(\rho - 1)} & \text{при} \quad \rho \neq 1; \\ \dfrac{M+1}{2\mu} & \text{при} \quad \rho = 1. \end{cases} \tag{3.10}$$

Среднее время ожидания заявки обслуживания определяется как разность среднего времени пребывания заявки в системе и среднего времени обслуживания. Поэтому последняя характеристика эффективности СМО имеет вид:

$$\overline{t_{\text{ож}}} = \overline{t_{\text{пр}}} - \overline{t_{\text{об}}} = \overline{t_{\text{пр}}} - \frac{1}{\mu}. \qquad (3.11)$$

Таким образом, все характеристики эффективности СМО найдены.

▪ <u>*Пример* 1.</u> Рассмотрим АЗС, имеющую две (n = 2) колонки обслуживания. Площадка при каждой колонке допускает пребывание не более трех ($M = 3$) машин одновременно. Поток прибывающих машин – простейший с интенсивностью λ = 1 $^{\text{маш}}/_{\text{мин}}$. Среднее время обслуживания одной машины составляет $\overline{t_{\text{об}}}$ = 4 $^{\text{мин}}/_{\text{маш}}$. Необходимо определить характеристики эффективности АЗС.

<u>*Решение.*</u> Рассматривается модель M/M/1 с ограниченной очередью. Первая буква M аббревиатуры означает, что поток прибывающих машин – простейший (марковский). Вторая буква M означает, что поток обслуживания – простейший. Последняя цифра 1 означает, что имеется один канал обслуживания. Если на АЗС имеется 2 колонки обслуживания, то предполагается, что обе колонки находятся в одинаковом режиме. Поэтому весь поток машин делится на две равных части, прибывающие на каждую колонку, а модель становится одноканальной.

Ключевой характеристикой АЗС является траффик-интенсивность потока прибывающих машин $\rho = \lambda \cdot \overline{t_{\text{об}}} / n$ = 2 ≠ 1. Поэтому вероятность того, что площадка уже занята, рассчитывается по формуле

$$P_{\text{отк}} = \frac{\rho^M(\rho - 1)}{\rho^{M+1} - 1} = 0{,}53,$$

а относительная пропускная способность АЗС $q = 1 - P_{\text{отк}} = 0.47$, абсолютная пропускная способность $A = \lambda \cdot q$ = 0,47 $^{\text{маш}}/_{\text{мин}}$. Среднее количество машин около каждой колонки имеет вид

$$\bar{k} = \frac{(M+1)\rho^{M+1}}{\rho^{M+1} - 1} - \frac{\rho}{\rho - 1} = 2{,}27 \text{маш}.$$

Вероятность того, что колонка занята (степень загрузки канала)

$$P_{\text{зан}} = \frac{\rho^{M+1} - \rho}{\rho^{M+1} - 1} = 0{,}875,$$

а среднее количество машин, ожидающих заправки определяется по формуле

$$\overline{l_{\text{оч}}} = \bar{k} - \frac{\rho^{\text{M}+1} - \rho}{\rho^{\text{M}+1} - 1} = 1{,}33 \text{ маш}.$$

Среднее время пребывания машины на АЗС имеет вид

$$\overline{t_{\text{пр}}} = \frac{M\rho^M}{\mu(\rho^M - 1)} - \frac{1}{\mu(\rho - 1)} = 9{,}7 \text{мин},$$

а среднее время ожидания обслуживания

$$\overline{t_{\text{ож}}} = \overline{t_{\text{пр}}} - \overline{t_{\text{об}}} = 5{,}7 \text{мин}.$$

- *<u>Пример</u>* <u>2.</u> Рассмотрим АЗС, имеющую пять (n = 5) колонок обслуживания. Поток прибывающих машин – простейший с интенсивностью λ = 1 $^{\text{маш}}/_{\text{мин}}$. Среднее время обслуживания одной машины составляет $\overline{t_{\text{об}}}$ = 4 $^{\text{мин}}/_{\text{маш}}$. Каково должно быть допустимое число машин возле каждой колонки, чтобы относительная пропускная способность АЗС была бы больше 95 % ?

<u>Решение.</u> Рассматривается модель М/М/1 с ограниченной очередью. Если вероятность обслуживания q > 0,95, то $P_{\text{отк}}$ < 0,05. Используя формулу для $P_{\text{отк}}$, получаем неравенство

$$M \ge M_1 = \left[-\frac{lg\left(\rho_1 + \dfrac{1-\rho_1}{1-0.95}\right)}{lg\,\rho_1} \right] + 1,$$

где $\rho_1 = \dfrac{\lambda \overline{t_{\text{об}}}}{n} = 0{,}8$, а квадратные скобки означают целую часть числа. Согласно данным задачи и неравенства получим $M \ge 8$.

3.7 Работа по теме «Модели СМО»

1. На оптовую базу прибывают автомашины с промышленными товарами. Поток простейший и поступает с интенсивностью λ автомобилей в час. На территории базы могут одновременно находиться не более M машин. Имеющиеся на базе n бригад грузчиков разгружают одновременно все только одну машину. Среднее время разгрузки одной машины составляет $t_{обс.}$ Необходимо определить основные показатели системы массового обслуживания оптовой базы: относительную и абсолютную пропускную способность, среднее число машин на базе, среднюю длину очереди, среднее время пребывания машины на базе и среднее время ожидания обслуживания при следующих значениях исходных данных:

$$n = 1,\ M = 3,\ \lambda = 2\ \text{авт/ч},\ t_{обс} = 1{,}5\ \text{ч}.$$

2. На оптовую базу прибывают автомашины с промышленными товарами. Поток простейший и поступает с интенсивностью λ автомобилей в час. На территории базы могут одновременно находиться не более M машин. Имеющиеся на базе n бригад грузчиков разгружают одновременно все только одну машину. Среднее время разгрузки одной машины составляет $t_{обс.}$ Необходимо определить основные показатели системы массового обслуживания оптовой базы: относительную и абсолютную пропускную способность, среднее число машин на базе, среднюю длину очереди, среднее время пребывания машины на базе и среднее время ожидания обслуживания при следующих значениях исходных данных:

$$n = 2,\ M = 3,\ \lambda = 2\ \text{авт/ч},\ t_{обс} = 1\ \text{ч}.$$

3. Универсам получает ранние овощи и зелень из теплиц пригородного совхоза. Машины прибывают в универсам в неопределенное время. В среднем прибывает λ автомашин в день. Подсобное помещение и оборудование для подготовки овощей к продаже позволяют обработать и хранить товар объемом не более M машин одновременно. В универсаме работают n фасовщиков, каждый из которых в среднем может обработать товар с одной машины в течении $t_{обс}$ дня. Определить вероятность обслуживания приходящей машины $P_{обс}$. Какова должна быть емкость подсобных помещений M_1, чтобы вероятность

обслуживания была бы не меньше заданной величины, т. е. $P_{обс} \geq$ $\geq P^{*}_{обс}$?

$$\lambda = 3 \text{ авт/день}, t_{обс} = 0{,}5, n = 2, M = 2, P^{*}_{обс} = 0{,}92.$$

4. Универсам получает ранние овощи и зелень из теплиц пригородного совхоза. Машины прибывают в универсам в неопределенное время. В среднем прибывает λ автомашин в день. Подсобное помещение и оборудование для подготовки овощей к продаже позволяют обработать и хранить товар объемом не более M машин одновременно. В универсаме работают n фасовщиков, каждый из которых в среднем может обработать товар с одной машины в течении $t_{обс}$ дня. Определить вероятность обслуживания приходящей машины $P_{обс}$. Какова должна быть емкость подсобных помещений M_1, чтобы вероятность обслуживания была бы не меньше заданной величины, т. е. $P_{обс} \geq$ $\geq P^{*}_{обс}$?

$$\lambda = 3 \text{ авт/день}, t_{обс} = 0{,}3, n = 2, M = 2, P^{*}_{обс} = 0{,}97.$$

5. Универсам получает ранние овощи и зелень из теплиц пригородного совхоза. Машины прибывают в универсам в неопределенное время. В среднем прибывает λ автомашин в день. Подсобное помещение и оборудование для подготовки овощей к продаже позволяют обработать и хранить товар объемом не более M машин одновременно. В универсаме работают n фасовщиков, каждый из которых в среднем может обработать товар с одной машины в течении $t_{обс}$ дня. Определить вероятность обслуживания приходящей машины $P_{обс}$. Какова должна быть емкость подсобных помещений M_1, чтобы вероятность обслуживания была бы не меньше заданной величины, т. е. $P_{обс} \geq$ $\geq P^{*}_{обс}$?

$$\lambda = 6 \text{ авт/день}, t_{обс} = 0{,}25, n = 4, M = 2, P^{*}_{обс} = 0{,}93.$$

6. В магазин самообслуживания поступает пуассоновский поток покупателей с интенсивностью λ человек в минуту. Средняя продолжительность обслуживания на расчетном узле составляет $t_{обс}$ мин. Уровень суммарных потерь связан с простоем среднего числа свободных контролеров-кассиров $n_{св.}$ и пребыванием среднего числа покупа-

телей в очереди $l_{оч}$. Построить график зависимости суммы среднего числа свободных контролеров-кассиров $n_{св}$ и среднего числа покупателей в очереди $l_{оч}$ от числа контролеров-кассиров n, $(l_{оч} + n_{св.}) = f(n)$. Определить по нему оптимальное число контролеров-кассиров n_0, при котором суммарные потери будут минимальными:

$$\lambda = 1 \text{ пок/мин}, \ t_{обс} = 2 \text{ мин.}$$

7. В магазин самообслуживания поступает пуассоновский поток покупателей с интенсивностью λ человек в минуту. Средняя продолжительность обслуживания на расчетном узле составляет $t_{обс}$ мин. Уровень суммарных потерь связан с простоем среднего числа свободных контролеров-кассиров $n_{св.}$ и пребыванием среднего числа покупателей в очереди $l_{оч}$. Построить график зависимости суммы среднего числа свободных контролеров-кассиров $n_{св}$ и среднего числа покупателей в очереди $l_{оч}$ от числа контролеров-кассиров n, $(l_{оч} + n_{св.}) = f(n)$. Определить по нему оптимальное число контролеров-кассиров n_0, при котором суммарные потери будут минимальными:

$$\lambda = 2 \text{ пок/мин}, \ t_{обс} = 1{,}4 \text{ мин.}$$

8. В магазин самообслуживания поступает пуассоновский поток покупателей с интенсивностью λ человек в минуту. Средняя продолжительность обслуживания на расчетном узле составляет $t_{обс}$ мин. Уровень суммарных потерь связан с простоем среднего числа свободных контролеров-кассиров $n_{св.}$ и пребыванием среднего числа покупателей в очереди $l_{оч}$. Построить график зависимости суммы среднего числа свободных контролеров-кассиров $n_{св}$ и среднего числа покупателей в очереди $l_{оч}$ от числа контролеров-кассиров n, $(l_{оч} + n_{св.}) = f(n)$. Определить по нему оптимальное число контролеров-кассиров n_0, при котором суммарные потери будут минимальными:

$$\lambda = 0{,}5 \text{ пок/мин}, \ t_{обс} = 1{,}8 \text{мин.}$$

9. В магазин самообслуживания поступает пуассоновский поток покупателей с интенсивностью λ человек в час. В течение дня их обслуживают n контролеров-кассиров с интенсивностью μ покупателей в час. Интенсивность входного потока покупателей в часы «пик» возрас-

тает до величины λ_{max}, а в часы «спада» достигает величины λ_{min}. Определить вероятность образования очереди в магазине $P_{оч}$ и среднюю длину очереди $l_{оч}$ в течение дня, а затем необходимое число контролеров-кассиров в часы «пик» n_{max} и часы «спада» n_{min}, обеспечивающих такую же длину очереди $l_{оч}$ и вероятность ее образования $P_{оч}$:

$$\lambda = 200 \text{ пок/час}, \mu = 90 \text{ пок/час}, n = 3,$$

$$\lambda_{max} = 400 \text{ пок/час}, \lambda_{min} = 100 \text{ пок/час}.$$

10. В магазин самообслуживания поступает пуассоновский поток покупателей с интенсивностью λ человек в час. В течение дня их обслуживают n контролеров-кассиров с интенсивностью μ покупателей в час. Интенсивность входного потока покупателей в часы «пик» возрастает до величины λ_{max}, а в часы «спада» достигает величины λ_{min}. Определить вероятность образования очереди в магазине $P_{оч}$ и среднюю длину очереди $l_{оч}$ в течение дня, а затем необходимое число контролеров-кассиров в часы «пик» n_{max} и часы «спада» n_{min}, обеспечивающих такую же длину очереди $l_{оч}$ и вероятность ее образования $P_{оч}$:

$$\lambda = 300 \text{ пок/час}, \mu = 120 \text{ пок/час}, n = 3.$$

$$\lambda_{max} = 500 \text{ пок/час}, \lambda_{min} = 140 \text{ пок/час}.$$

11. Диспетчерский пункт приема неотложных медицинских вызовов по телефону 02 в г. Новокузнецке обслуживает поток заявок с интенсивностью 100 вызовов в час, причем каждый вызов обслуживается в среднем 30 секунд. Необходимо определить количество каналов обслуживания, при котором вероятность отказа меньше 5 %. Уровень потерь связан с простоем среднего числа операторов и средним количеством необслуженных вызовов в час. Определить оптимальное число операторов, при которых суммарные потери будут минимальными.

12. Диспетчерский пункт приема неотложных медицинских вызовов по телефону 02 в г. Кемерово обслуживает поток заявок с интенсивностью 90 вызовов в час, причем каждый вызов обслуживается в среднем 50 секунд. Необходимо определить количество каналов обслуживания, при котором вероятность отказа меньше 5 %. Уровень по-

терь связан с простоем среднего числа операторов и средним количеством необслуженных вызовов в час. Определить оптимальное число операторов, при которых суммарные потери будут минимальными.

4 Модели управления запасами

4.1 Исследование операций в экономике

Среди многочисленных проблем, возникновение которых обусловлено бурно развивающейся научно-технической революцией, особое место занимает проблема управления во всех звеньях промышленности. В настоящее время разработан и интенсивно развивается раздел математики «Исследование операций», целью которого является количественное обоснование принимаемых решений по организации управления. Термин «исследование операций» возник во время Второй мировой войны. Однако зачатки научного мышления, характерного для операционных исследований, появились гораздо раньше. Так, например, некоторые, хотя и весьма примитивные, модели математического программирования были предложены еще в 1759 г. экономистом Куисни и в 1874 г. экономистом Вальрасом. Более сложные экономические модели были разработаны в 1937 г. фон Нейманом и в 1939 г. Леонидом Витальевичем Канторовичем, академиком из Ленинградского госуниверситета. Занявшись планированием работы агрегатов фанерной фабрики, он решил несколько задач: о наилучшей загрузке оборудования, о раскрое материалов с наименьшими потерями, о распределении грузов по нескольким видам транспорта и др. Леонид Витальевич сформулировал новый класс условно-экстремальных задач и предложил универсальный метод их решения, положив начало новому направлению прикладной математики – линейному программированию.

Несмотря на то, что результаты этих первых поисковых работ получили признание и высокую оценку, математические модели как элемент анализа управляющих решений стали использоваться лишь недавно. Руководители крупных корпораций поняли, что совершенствование старых сложившихся способов накопления и обработки информации приводит к существенным выгодам. Этот процесс стимулировало широкое распространение быстродействующих ЭВМ.

Квинтэссенцией операционного подхода к решению организационных задач является построение математических моделей, которые играют роль, аналогичную лабораторному эксперименту в естественных науках. Модель задачи помогает привести сложные и подчас неопределенные факторы в логически стройную схему, доступную для детального анализа. Такая модель позволяет выявить альтернативы решения задачи и оценить результаты, к которым они приводят, а также дает возможность определить, какие данные необходимы для оценки имеющихся альтернатив. В итоге это обеспечивает получение обоснованных выводов.

По своей содержательной постановке множество задач исследования операций может быть разбито на ряд классов [3].

Задачи сетевого планирования и управления рассматривают соотношения между сроками окончания крупного комплекса операций (работ) и моментами начала всех операций комплекса. Эти задачи состоят в нахождении минимальных продолжительностей комплекса операций, оптимального соотношения величин стоимости и сроков их выполнения.

Задачи массового обслуживания посвящены изучению и анализу систем обслуживания с очередями заявок или требований и состоят в определении показателей эффективности работы систем, их оптимальных характеристик, например, времени ожидания обслуживания, длины очереди и т. п.

Задачи управления запасами состоят в отыскании оптимальных значений уровня запасов и размера объема выпуска. Особенность таких задач заключается в том, что с увеличением уровня запасов, с одной стороны, увеличиваются затраты на их хранение, но с другой стороны, уменьшаются потери вследствие необходимой переналадки оборудования.

Задачи распределения ресурсов возникают при определенном наборе операций (работ), которые необходимо выполнить при ограниченных наличных ресурсах, и требуется найти оптимальные распределения ресурсов между операциями или состав операций.

Задачи ремонта и замены оборудования актуальны в связи с износом и старением оборудования и необходимостью его замены с течением времени. Задачи сводятся к определению оптимальных сроков, числа профилактических ремонтов и проверок, а также моментов замены старого оборудования новым.

Задачи составления расписания (календарного планирования) состоят в определении оптимальной очередности выполнения операций на различных видах оборудования с целью минимизации простоя оборудования.

Задачи назначения состоят в отыскании оптимальных кандидатов для занятия имеющихся вакантных должностей. Эти задачи широко применяются, например, в спорте, когда требуется подобрать оптимальный состав олимпийской сборной по разным видам спорта. Каждый кандидат в команду имеет ряд показателей, а состав называется оптимальным, если суммарный показатель команды с учетом всех показателей ее членов и коэффициентов важности показателя на каждой вакансии является максимальным.

Задачи выбора маршрута состоят в отыскании таких объемов перевозок от производителей некоторой однородной продукции к ее потребителям, при которых достигаются минимальные транспортные затраты с учетом спроса потребителей и запаса производителей.

В настоящее время самым хорошо изученным разделом исследования операций является линейное программирование, подразделом которого является динамическое программирование. В этом подразделе рассматриваются наиболее интересные задачи планирования, носящие мультивременной характер. Характерной в этом смысле является задача управления запасами, в которой условия принятия решения меняются во времени. При этом она не сводится полностью к задаче оптимизации для последовательных периодов времени, рассматриваемых изолированно друг от друга. Текущее управляющее решение «проявляется» как в период, относящийся непосредственно к моменту принятия решения, так и в последующие периоды. Следовательно, наиболее важные экономические последствия проявляются в разные периоды, а не только в течение одного периода. Такого рода экономические последствия, как правило, оказываются существенными в тех случаях, когда речь идет об управляющих решениях, связанных с возможностью новых капиталовложений, увеличения производственных мощностей или обучения персонала с целью создания предпосылок для повышения прибыльности предприятия или сокращения издержек в последующие периоды.

При решении конкретных задач управления методами исследования операций предполагается [2]:

- построение экономических и математических моделей для задач принятия решения в сложных ситуациях;
- изучение взаимосвязей, определяющих впоследствии принятие решений, и установление критериев эффективности, позволяющих оценить преимущество того или иного варианта действия.

В настоящей работе рассматриваются две задачи [2]: задача о дилижансах и задача управления запасами. Первая задача носит аллегорический характер и помогает понять некоторые важные идеи динамического программирования, а также построить систему условных обозначений динамических моделей. Вторая задача демонстрирует возможности динамического программирования не только решать конкретную практически важную задачу, но и оперативно реагировать на изменения условий задачи.

Задаче управления запасами посвящено много работ, выполненных с начала 60-х годов прошлого века. Вот только некоторые фундаментальные работы, выпущенные в этот период: [4, 16, 19, 20, 23, 26], а количество статей в научных журналах просто невозможно перечислить.

4.2 Задача о дилижансах

Жил некогда мистер М., который решил отправиться искать счастья в Сан-Франциско. В те дни дилижансы были единственным видом общественного транспорта для поездки из восточных штатов, где проживал мистер М., на Запад. В бюро путешествий ему показали карту Соединенных Штатов (рисунок 4.1) с нанесенными на ней дилижансовыми маршрутами, которые обслуживались в то время. Каждый квадрат на карте изображает один из штатов (состояний); для удобства штаты пронумерованы. Заметим, что какой бы из вариантов пути от штата 1 (Восток) до штата 10 (Запад) мы ни выбрали, он включает 4 дилижансовых маршрута – или 4 «шага».

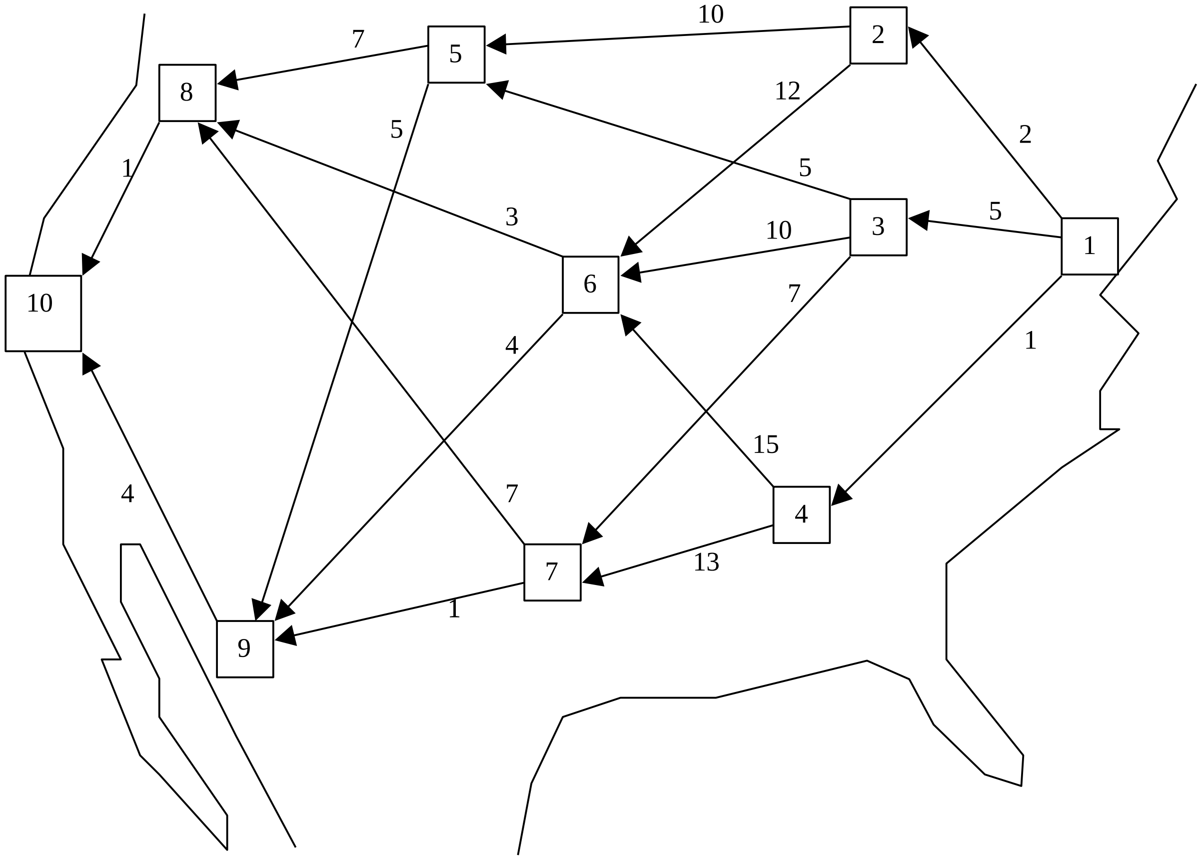

Рисунок 4.1 – Сеть дилижансовых маршрутов

Поскольку мистеру М. было известно, что путешествие связано с серьезными опасностями для здоровья и жизни, перед отъездом он решил застраховаться. Ставка страхового платежа (иными словами, стоимость, отвечающая принятой стратегии выбора пути) зависела от избираемых дилижансовых маршрутов, и она была тем выше, чем опаснее маршрут. Обозначим через C_{ij} стоимость страхового полиса для переезда из штата i в штат j (в денежных единицах). Условные числовые значения C_{ij} проставлены на рисунке 4.1. Цель мистера М. – выбрать такой путь от штата 1 до штата 10, для которого общая стоимость страхования является минимальной.

Мистер М. начал анализ проблемы с того, что признал существенным следующий принцип.

Принцип оптимальности. Оптимальная стратегия обладает тем свойством, что, каков бы ни был путь достижения некоторого штата, последующие решения должны принадлежать оптимальной стратегии для части пути, начинающейся с этого штата. Иначе говоря, каждая часть оптимального пути должна быть оптимальной сама по себе.

Таким образом, мистер М. понял, что, например, оптимальный путь из штата 6 не зависит от того, каким маршрутом он прибудет в него. Развивая эту логическую идею далее, мистер М. пришел к выводу, что если бы он знал оптимальные пути из штатов 5, 6 и 7, то смог бы достаточно легко определить и оптимальный путь из штата 3 – конечно, если бы он решил ехать через этот штат. В самом деле, потребовалось бы лишь *суммировать* стоимость проезда из штата 3 (будь то C_{35}, C_{36} или C_{37}) с ранее вычисленной стоимостью оптимального пути (соответственно из штатов 5, 6 или 7), а затем сравнить полученные суммы и выбрать тот штат, для которого эта сумма минимальна. Аналогичным образом, найдя оптимальные пути из штатов 2, 3 и 4, можно определить оптимальный путь из штата 1.

Для того, чтобы учесть сформулированный принцип оптимальности, удобно использовать следующее обозначение: Y_s – сстоимость проезда, отвечающая оптимальному пути из штата s в штат 10. Тогда, возвращаясь к стоящей перед мистером М. задаче, получим

$$Y_{10} = 0\,, \tag{4.1}$$

поскольку для штата 10 число оставшихся шагов равно нулю, т. е. в этом штате путешествие заканчивается. Затем он видит, что очень легко вычисляются Y_8 и Y_9: к Y_{10} попросту надо прибавить соответственно $C_{8,10}$ или $C_{9,10}$. Ободренный достигнутым успехами, мистер М. решил вычислить Y_6 – стоимость стратегии минимальных затрат, отвечающей его нахождению в штате 6, т. е. за два шага до конечного пункта. Он заметил, что если он попадет в штат 6, то далее можно следовать только по двум путям. Первый из них – через штат 8; стоимость соответствующей стратегии определится, если прибавить C_{68} к вычисленному ранее значению Y_8. Второй путь – через штат 9; для вычисления его стоимости прибавим C_{69} к значению Y_9, которое также уже вычислено. Мистер М. понял, что значение Y_6 должно равняться меньшей из этих двух сумм.

Здесь мистер М. начал подозревать, что в его подходе можно обнаружить известную методичность, и, разумеется, он был прав. Кратко описываемый метод можно представить в виде так называемого динамического рекуррентного соотношения:

$$Y_i = \min_{j \in сети} [C_{ij} + Y_j], \; i = \overline{1,9}, \; Y_{10} = 0. \qquad (4.2)$$

Вычислительный процесс. Рассмотрим применение рекуррентного соотношения (4.2), выполнив необходимые вычисления для численных значений C_{ij}, проставленных на рисунке 4.1. $Y_9 = 4$; $Y_8 = 1$, так как эти штаты связаны единственным путем со штатом 10;

$Y_7 = \min[C_{78} + Y_8, C_{79} + Y_9] = \min[7 + 1, \underline{1 + 4}] = 5;$

$Y_6 = \min[C_{68} + Y_8, C_{69} + Y_9] = \min[\underline{3 + 1}, 4 + 4] = 4;$

$Y_5 = \min[C_{58} + Y_8, C_{59} + Y_9] = \min[\underline{7 + 1}, 5 + 4] = 8;$

$Y_4 = \min[C_{46} + Y_6, C_{47} + Y_7] = \min[15 + 4, \underline{13 + 5}] = 18;$

$Y_3 = \min[C_{35} + Y_5, C_{36} + Y_6, C_{37} + Y_7] = \min[5 + 8, 10 + 4, \underline{7 + 5}] = 12;$

$Y_2 = \min[C_{25} + Y_5, C_{26} + Y_6] = \min[10 + 8, \underline{12 + 4}] = 16;$

$Y_1 = \min[C_{12} + Y_2, C_{13} + Y_3, C_{14} + Y_4] = \min[2 + 16, \underline{5 + 12}, 1 + 18] = 17.$

Оптимальная страховая сумма составляет 17 денежных единиц. Оптимальный путь: $1 \to 3 \to 7 \to 9 \to 10$ (определяется подчеркнутыми слагаемыми). Первый его участок $1 \to 3$ определяется в выражении для Y_1. Следующий участок $3 \to 7$ в выражении для Y_3 и так далее.

4.3 Анализ чувствительности решения к условиям задачи

1. Проделанные вычисления позволяют не только получить оптимальное решение, но также принять правильное решение при изменившихся условиях задачи. Рассмотрим, например, ситуацию, когда сестра мистера М. миссис Н. проживает в штате 5. Найдем оптималь-

ный путь из штата 1 в штат 10, проходящий через штат 5.

Для решения этой задачи надо определить два оптимальных пути: из штата 1 в штат 5 и из штата 5 в штат 10. Первый путь определяется из двух возможных: путь 1→2→5 приводит к страховой сумме в 12 д. е., а путь 1→3→5 приводит к 10 д. е. и поэтому является оптимальным. Из штата 5 в штат 10 оптимальный путь уже найден при вычислении Y_5 : 5→8→10, который приводит к 8 д. е. Окончательное решение задачи: путь 1→3→5→8→10, который приводит к 18 д. е.

2. В качестве второго примера предположим, что между штатами 7, 9 нет дилижансового сообщения. Каким теперь будет оптимальный путь из штата 1 в штат 10 ?

Для решения этой задачи необходимо выяснить, для каких штатов изменится оптимальная сумма страхования. Это, прежде всего, штат 7 и штаты 1, 3, 4, из которых можно попасть в штат 7.

Вычислительный процесс: $Y_7 = \min[C_{78} + Y_8] = \min[7 + 1] = 8;$

$Y_4 = \min[C_{46} + Y_6, C_{47} + Y_7] = \min[\underline{15 + 4}, 13 + 8] = 19;$

$Y_3 = \min[C_{35} + Y_5, C_{36} + Y_6, C_{37} + Y_7] = \min[\underline{5 + 8}, 10 + 4, 7 + 8] = 13;$

$Y_1 = \min[C_{12} + Y_2, C_{13} + Y_3, C_{14} + Y_4] = \min[\underline{2 + 16}, \underline{5 + 13}, 1 + 19] = 18.$

Таким образом, минимальная страховая сумма составляет 18 д. е. и достигается двумя путями: 1→2→6→8→10 и 1→3→5→8→10.

Аналогичным образом исследуются другие возможные изменения условий задачи:

3. Пусть введен дополнительный маршрут из штата 3 в штат 8. Какой должна быть наименьшая ставка страхового платежа на этом маршруте, чтобы мистер М. по-прежнему считал предпочтительным выбранный раннее путь ?

4. Определить диапазон ставок страхового платежа для переезда из штата 1 в штат 3, в рамках которого раннее выбранный путь остается оптимальным. Найти аналогичный диапазон для переезда из штата 3 в штат 7, а также из штата 2 в штат 6.

4.4 Задача управления запасами с вогнутой функцией затрат

Рассматривается упрощенная модель управления запасами [2], позволяющая сосредоточить внимание на небольшом числе важнейших особенностей динамических процессов управления запасами. Предполагается, что затраты, связанные с хранением и пополнением запасов, являются вогнутыми функциями. Такая ситуация часто встречается в случае, когда выпуск продукции связан с необходимостью затрат на подготовительные операции или переналадку (единовременные затраты), после чего выпуску каждой дополнительной единицы продукции соответствуют не меняющиеся пропорциональные затраты.

Вогнутая функция затрат часто возникает и в других случаях: когда запасы пополняются у внешнего поставщика, который делает скидки при крупных заказах.

Эта модель играет такую же роль в исследовании операций, как закон Ньютона в физике. Хотя рассматриваемая ситуация является идеальной, в модели все же учитывается множество важных факторов, влияющих на выбор правил управления запасами.

Пример фирмы «Надежный компьютер». Фирма должна разработать календарную программу выпуска компьютеров на плановый период, состоящий из N отрезков. Предполагается, что для каждого из этих отрезков имеется точный прогноз спроса на выпускаемую продукцию.

Время изготовления партии компьютеров настолько мало, что им можно пренебречь; соответственно продукция, изготовленная в течение отрезка t, может быть использована для полного или частичного покрытия спроса в течение этого отрезка. Для разных отрезков спрос неодинаков; кроме того, на экономические показатели производства влияют размеры изготовляемых партий. Это связано с *дополнительными затратами на переналадку производства*, необходимую всякий раз, как только возобновляется производство компьютеров. Поэтому фирме нередко бывает выгодно изготовлять в течение некоторого месяца продукцию в объеме, превышающем спрос в пределах этого отрезка, и хранить излишки, используя их для удовлетворения последующего спроса. Вместе с тем хранение возникающих при этом запасов связано с определенными затратами. В зависимости от обстоятельств затраты обусловлены такими факторами, как проценты на капитал, взятый взаймы для создания запасов, арендная плата за складские помещения, страховые взносы и расходы на содержание запасов. Эти затраты необходимо учитывать при установлении программы выпуска компьютеров.

Цель фирмы «Надежный компьютер» – разработать такую программу, при которой общая сумма затрат на производство и содержание запасов минимизируется при условии полного и своевременного удовлетворения спроса на продукцию. Анализ этой задачи начнем с преобразования ее в разновидность задачи о дилижансах.

Построение модели. Введем обозначения:

x_t – объем выпуска продукции в течение отрезка t;

i_t – уровень запасов на конец отрезка t;

D_t – спрос на продукцию для отрезка t;

$C_t(x_t, i_t)$ – затраты на отрезке t, связанные с выпуском продукции и хранением запасов.

Целью задачи является минимизация суммарных затрат за весь плановый период

$$\sum_{t=1}^{N} C_t(x_t, i_t) \rightarrow min.$$

На значения переменных x_t, i_t наложено несколько ограничений. Во-первых, предполагается, что эти переменные могут принимать только целочисленные значения 0, 1, 2, … и т. д. Во-вторых, предполагается, что для администрации фирмы желателен нулевой уровень запасов на конец отрезка N

$$i_N = 0 .$$

Наконец, в-третьих, ставится условие полного и своевременного удовлетворения спроса в пределах каждого месяца. Выполнение этого условия можно обеспечить, введя два ограничения. Первое из них представляется уравнением вида

$$i_t = i_{t-1} + x_t - D_t \ (t = 1, 2, \dots, N-1),$$

а второе – целочисленностью и неотрицательностью уровней запасов

$$i_t = 0, 1, 2, \dots \; (t = 1, 2, \dots, N-1)\,.$$

4.5 Алгоритм оптимизации модели. Числовой пример

В теории рассматриваемой задачи управления запасами, когда затраты $C_t\,(x_t,\; i_t)$ зависят от необходимой переналадки производства (случай вогнутой функции затрат), важную роль играет следующий факт: ***оптимальным является такой план выпуска продукции, при котором на любом отрезке объем выпуска равен нулю, если остаток на начало этого отрезка строго положителен***.

Введем следующие обозначения:

C_{ij} – общие затраты, связанные с хранением и производством в конце отрезка i объема выпуска, необходимого для удовлетворения спроса на отрезках $i + 1, \dots j$, т. е. с конца отрезка i до конца отрезка j $(i < j)$;

h_t – затраты на хранение единицы продукции в течение отрезка t;

Y_i – минимальные общие затраты с конца отрезка i до конца планового периода ($i = 0, 1, \dots, N$).

Рассмотрим следующий числовой пример задачи управления запасами, в котором плановый период состоит из 4 месяцев.

Спрос на продукцию каждый месяц составляет $D_t = 1$ при всех $t = 1$, 2, 3, 4. Стоимость хранения $h_t(i_t) = h_t \cdot i_t$, где $h_t = h = 0{,}1$ при всех $t = 1$, 2, 3, 4.

Затраты $C_t(x_t,\; i_t) = C_t(x_t) + h_t(i_t)$, где $C_t(x_t)$ – стоимость выпуска объема x_t определена с учетом затрат на переналадку производства в виде следующей таблицы 4.1.

Таблица 4.1

Месяц t	Объем выпуска компьютеров x_t					
	0	1	2	3	4	5
1	0	5	10	12	14	16
2	0	4	8	10	12	14

3	0	3	6	8	10	12
4	0	6	8	10	12	14
5	0	2	3	4	5	6

С помощью заданной функции затрат вычислим общие затраты C_{ij}

$$C_{ij} = C_{i+1}(j - i) + h \cdot (j - i) \cdot (j - i - 1)/2 \,,$$

которые представлены в виде следующей таблицы 4.2.

Таблица 4.2

i	j				
	1	2	3	4	5
0	5	10,1	12,3	14,6	17
1		4	8,1	10,3	12,6
2			3	6,1	9,3
3				6	8,1
4					2

Затраты C_{ij} напоминают стоимости страхования при переезде из штата i в штат j , а Y_i – минимальные стоимости проезда из штата i в конечный штат N . Поэтому рассматриваемая задача об управлении запасов сводится к частному случаю задачи о дилижансах с сетью дилижансовых маршрутов, приведенных на рисунке 4.2 а.

Нулевой штат соответствует концу нулевого или началу первого планового месяца, а штат 4 – концу планового периода, состоящего из 4-х месяцев.

Оптимальное решение отыскивается с помощью рекуррентного соотношения, аналогичного (4.2)

$$Y_i = \min_{j \in сети} [C_{ij} + Y_j] \,,\ i = \overline{0,3} \,,\ Y_4 = 0 \,.$$

Вычислительный процесс: $Y_3 = 6$;

$Y_2 = \min[C_{23} + Y_3, C_{24} + Y_4] = \min[3 + 6; \underline{6{,}1}] = 6{,}1$;

$Y_1 = \min[C_{12} + Y_2, C_{13} + Y_3, C_{14} + Y_4] = \min[4 + 6{,}1; 8{,}1 + 6; \underline{10{,}3}] =$
$= 10{,}3$;

$Y_0 = \min[C_{01} + Y_1, C_{02} + Y_2, C_{03} + Y_3, C_{04} + Y_4] = \min[5 + 10{,}3; 10{,}1 + + 6{,}1; 12{,}3 + 6; \underline{14{,}6}] = 14{,}6$.

а б

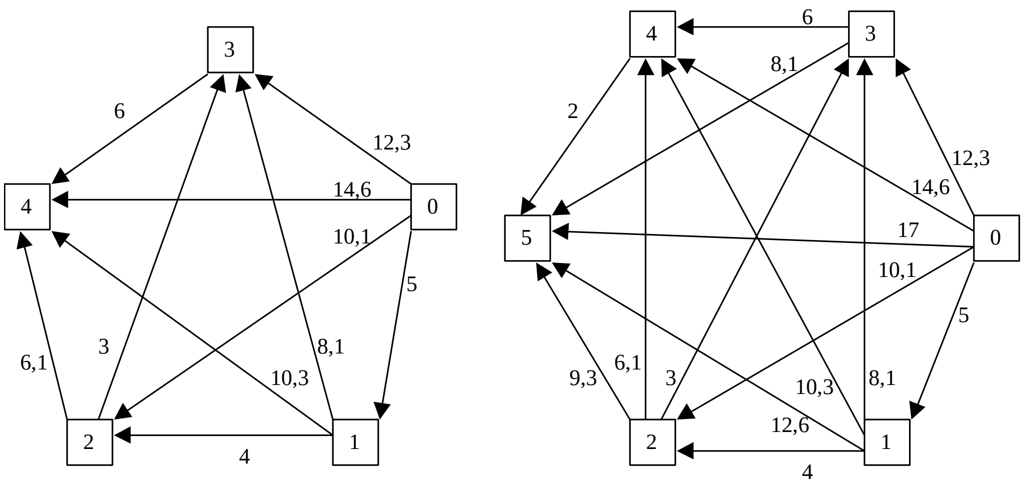

Рисунок 4.2 – Модель управления запасами фирмы «Надежный компьютер»

Минимальные затраты составляют 14,6 д. е. и достигаются при следующем плане выпуска компьютеров: $x_1 = 4; x_2 = x_3 = x_4 = 0$, который определяется с помощью подчеркнутых слагаемых подобно задаче о дилижансах.

4.6 Анализ чувствительности решения к длительности планового периода

1. Полученное решение позволяет без труда исследовать оптимальное решение в случае, когда плановый период составляет 5 месяцев. Для этого составляется сеть из рисунка 4.2 б.

Вычислительный процесс: $Y_5 = 0$;

$Y_4 = 2$; $Y_3 = \min[C_{34} + Y_4, C_{35} + Y_5] = \min[\underline{6 + 2}; 8{,}1] = 8$;

$Y_2 = \min[C_{23} + Y_3, C_{24} + Y_4, C_{25} + Y_5] = \min[3 + 8; \underline{6{,}1 + 2}; 9{,}3] = 8{,}1$;

$Y_1 = \min[C_{12} + Y_2, C_{13} + Y_3, C_{14} + Y_4, C_{15} + Y_5] = \min[\underline{4 + 8{,}1}; 8{,}1 + 8; 10{,}3 + 2; 12{,}6] = 12{,}1$;

$Y_0 = \min[C_{01} + Y_1, C_{02} + Y_2, C_{03} + Y_3, C_{04} + Y_4, C_{05} + Y_5] = \min[5 +$

+ 12,1; 10,1 + 8,1; 12,3 + 8; <u>14,6 + 2</u>; 17] = 16,6 .

Минимальные затраты составляют 16,6 д. е. и достигаются при следующем плане выпуска компьютеров: $x_1 = 4$; $x_2 = x_3 = x_4 = 0$; $x_5 =$ $= 1$.

2. Аналогично исследуется решение задачи в случае, когда плановый период составляет 3 месяца. В этом случае минимальные затраты составляют 12 д. е. и достигаются при следующем плане выпуска компьютеров: $x_1 = x_2 = x_3 = 1$.

Проведенный анализ показывает, что решение задачи существенно зависит от длительности планового периода.

4.7 Статическая модель

Статической называется модель, в которой решение относительно пополнения запасов принимается только раз, а плановый период состоит из одного отрезка. Эта модель хорошо описывает реальную ситуацию, например, в следующих случаях:

1) издательство массовой газеты;
2) выпуск серии нового автомобиля;
3) производство хлебных, мясных и молочных изделий.

Математическая модель. Примем следующие обозначения: $\boldsymbol{i}$ – уровень запасов вначале планового периода; $\boldsymbol{x}$ – объем заказа; $\boldsymbol{y = i + x}$ – суммарный объем запасов; $\boldsymbol{q}$ – уровень спроса; $\boldsymbol{p(q)}$ – распределение вероятностей уровней спроса $\boldsymbol{q = 0, \dots, Q}$.

Поскольку уровень спроса – случаен, то величина $\boldsymbol{y - q}$ может оказаться отрицательной, то есть спрос неудовлетворен. *Штраф за неудовлетворенный спрос* описывается функцией, в которой $\boldsymbol{\pi}$ – розничная цена изделия

$$\pi(y - q) = \begin{cases} \pi \cdot (q - y) \text{ при } q - y > 0; \\ 0 \text{ при } q - y \le 0. \end{cases}$$

Если величина $\boldsymbol{y - q}$ окажется положительной, то нереализованная продукция хранится на складе (автомобиль) или ликвидируется (торт, газета) из-за истечения срока годности. *Затраты на хранение* описываются функцией, в которой $\boldsymbol{h}$ – стоимость хранения единицы продукции

$$h(y-q)=\begin{cases} h\cdot(y-q) \text{ при } y-q>0; \\ 0 \text{ при } y-q\le 0. \end{cases}$$

Сумма рассмотренных двух видов затрат является случайной функцией $\xi(y)$ с параметром y и следующим (таблица 4.3) законом распределения.

Таблица 4.3

$\boldsymbol{\xi(y)}$	$\boldsymbol{\pi(y-q_1)}$ при всех $\boldsymbol{q_1 \ge y}$	$\boldsymbol{h(y-q_2)}$ при всех $\boldsymbol{q_2 < y}$
p	$p(q_1)$	$p(q_2)$

Ожидаемые затраты определяются как математическое ожидание случайной функции $\xi(y)$ и являются функцией от объема суммарного запаса $L(y) = M[\xi(y)]$.

$$L(y)=\sum_{q<y} h(y-q)\cdot p(q)+\sum_{q>y}\pi(y-q)\cdot p(q).$$

Затраты, связанные с пополнением запасов, определяются функцией, в которой $\boldsymbol{c}$ – оптовая цена единицы продукции, а $\boldsymbol{K}$ – накладные расходы, связанные с оформлением и транспортировкой заказа

$$C(x)=\begin{cases} c\cdot x+K \quad \text{при } x>0; \\ 0 \quad \text{при } x=0. \end{cases}$$

В этих условиях оптимальной является ($\boldsymbol{s}$, $\boldsymbol{S}$) стратегия пополнения запасов, согласно которой суммарный объем запасов определяется начальным уровнем запасов в виде следующей функции:

$$y(i)=\begin{cases} S \quad \text{при } i<s; \\ i \quad \text{при } i\ge s, \end{cases}$$

где $\boldsymbol{s}$ – пороговое значение запаса, при котором не проводится пополнение запасов, а $\boldsymbol{S}$ – суммарный объем пополнения запасов.

Расчет ($\boldsymbol{s}$, $\boldsymbol{S}$) стратегии проводится по следующей схеме.

1) Рассчитывается параметр $\boldsymbol{S}$ как минимальное целое решение неравенства $\boldsymbol{F(S) \ge R}$, в котором $\boldsymbol{F(x)}$ – функция распределения значений случайного спроса $\boldsymbol{q}$, а $\boldsymbol{R}$ – критическое отношение

$$R = \frac{\pi - c}{\pi + h}.$$

2) Параметр $\boldsymbol{s}$ является минимальным целым решением неравенства $\boldsymbol{L(s) \le K + c(S - s) + L(S)}$ **.**

Правая часть неравенства равна суммарным затратам в случае пополнения запасов с начального уровня $\boldsymbol{s}$ до объема $\boldsymbol{S}$, а левая часть – в случае, когда рассматриваемый уровень запасов не пополняется.

Числовой пример. Пусть случайный спрос является равномерно распределенной величиной, которая принимает значения $\boldsymbol{q} = 1, \ldots, 10$, а $\boldsymbol{p(q)} = 0{,}1$ при всех значениях $\boldsymbol{q}$. Функция распределения значений спроса $\boldsymbol{F(x)}$ равна вероятности того, что величина спроса не превосходит аргумента $\boldsymbol{x}$. График этой функции изображен на рисунке 4.3.

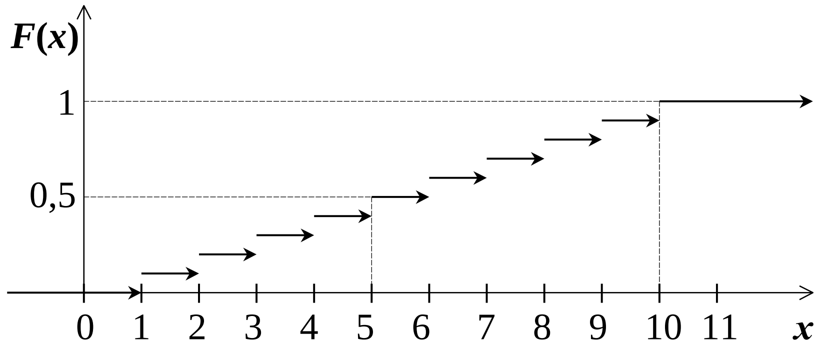

Рисунок 4.3 – График функции распределения спроса

Для каждого целого значения аргумента $\boldsymbol{x} = 0,\ldots,10$ справедливо равенство $\boldsymbol{F(x)} = \boldsymbol{x}/10$.

Поэтому $\boldsymbol{S}$ является минимальным целым решением неравенства

$$S \ge 10\frac{\pi - c}{\pi + h}.$$

Слагаемые сумм, входящих в выражение функции ожидаемых затрат $\boldsymbol{L(y)}$, образуют арифметическую прогрессию. Поэтому используется формула суммы членов арифметической прогрессии, согласно которой

$$L(y) = \left(\frac{h+\pi}{20}\right)y^2 - \left(\frac{h+21\pi}{20}\right)y + \frac{11\pi}{2}.$$

В этом случае $\boldsymbol{s}$ является минимальным целым решением неравенства

$$\frac{h+\pi}{20}(S^2-s^2)-\left(\frac{h+21\pi}{20}-c\right)(S-s)+K\geq 0.$$

Отметим некоторые свойства оптимальной стратегии.

1) При возрастании розничной цены товара $\boldsymbol{\pi}$ объем пополнения запасов $\boldsymbol{S}$ возрастает.

2) При возрастании оптовой цены товара $\boldsymbol{c}$ либо стоимости хранения единицы товара $\boldsymbol{h}$ объем пополнения запасов $\boldsymbol{S}$ убывает.

3) Объем пополнения запасов $\boldsymbol{S}$ не зависит от величины накладных расходов $\boldsymbol{K}$.

4) При $\boldsymbol{K} = 0$ $\boldsymbol{s} = \boldsymbol{S}$, а при возрастании $\boldsymbol{K}$ значение $\boldsymbol{s}$ убывает.

▪ _Пример 1._ Пусть $\boldsymbol{\pi} = 10$; $\boldsymbol{c} = 3$; $\boldsymbol{K} = 8$; $\boldsymbol{h} = 2$. Найти оптимальную ($\boldsymbol{s}$, $\boldsymbol{S}$) стратегию пополнения запасов.

Решение. Сначала рассчитаем $\boldsymbol{S}$ как минимальное целое решение неравенства $S\geq 10\frac{\pi-c}{\pi+h}=\frac{70}{12}\approx 5{,}8$. Поэтому $\boldsymbol{S} = 6$.

Затем рассчитаем $\boldsymbol{s}$ с помощью следующей функции

$$f(s)=\frac{h+\pi}{20}(S^2-s^2)-\left(\frac{h+21\pi}{20}-c\right)(S-s)+K=0{,}6(36-s^2)-7{,}6(6-s)+8.$$

Составим таблицу 4.4 значений $f(s)$ при $\boldsymbol{s} = 6; 5; 4; 3; 2$, то есть, начиная с $\boldsymbol{S}$ до тех пор, когда значение этой функции окажется отрицательным.

Таблица 4.4

$\boldsymbol{s}$	6	5	4	3	2
$\boldsymbol{f(s)}$	8	7	4,8	1,4	–3,2

Минимальное целое решение неравенства $\boldsymbol{f(s)} \geq 0$ имеет вид $\boldsymbol{s} = 3$.

Поэтому оптимальная стратегия пополнения запасов имеет вид

$$y(i)=\begin{cases}6 & \text{при } i<3;\\ i & \text{при } i\geq 3;\end{cases}\qquad x(i)=\begin{cases}6-i & \text{при } i<3;\\ 0 & \text{при } i\geq 3.\end{cases}$$

4.8 Случай единовременного штрафа

Рассмотренная выше ($\boldsymbol{s}$, $\boldsymbol{S}$) стратегия пополнения запасов является оптимальной только в случае выпуклых функций $\pi(y - q)$ и $h(y - q)$. В противном случае оптимальная стратегия выглядит иначе.

Рассмотрим подобный пример, в котором штраф за неудовлетворенный спрос $\pi(y - q)$ не зависит от *величины* неудовлетворенного спроса $y - q$. Такая ситуация характерна, например, для авиакомпании, составляющей расписание авиарейсов между Нью-Йорком и Вашингтоном. На рейсы такого рода билеты не резервируются. Если пассажир прибывает в аэропорт вовремя, то авиакомпания *гарантирует* ему место в самолете. Если при этом запланированных мест не хватает, то назначается дополнительный аэробус, даже если в нем будет всего один пассажир.

Математическая модель. Штраф за неудовлетворенный спрос описывается функцией, в которой π – величина штрафа

$$\pi(y-q)=\begin{cases}\pi & \text{при } y-q<0;\\ 0 & \text{при } y-q\geq 0.\end{cases}$$

Затраты на хранение отсутствуют (вакантные места не надо охранять) $h = 0$.

Затраты, связанные с пополнением запасов $C(y - i)$ определяются аналогично случаю, рассмотренному выше.

Числовой пример. Рассмотрим задачу, в которой спрос принимает всего два возможных значения $q_1 < q_2$ с равными вероятностями, например

$$q=\begin{cases}4 & \text{с вероятностью } p=0{,}5;\\ 8 & \text{с вероятностью } p=0{,}5.\end{cases}$$

Ожидаемые затраты определяются как математическое ожидание штрафа за неудовлетворенный спрос

$$L(y)=M[\pi(y-q)]=\pi(y-q_1)\frac{1}{2}+\pi(y-q_2)\frac{1}{2}$$

и являются функцией от объема суммарного запаса

$$L(y)=\begin{cases}\pi & \text{при } y=\overline{0{,}3};\\ \pi/2 & \text{при } y=\overline{4{,}7};\\ 0 & \text{при } y=8.\end{cases}$$

▪ *Пример* 1. Пусть $\pi = 12$; $c = 1$; $K = 4{,}5$. Построить оптимальную стратегию пополнения запасов.

Решение. Суммарные ожидаемые затраты g зависят от начального уровня запасов i , суммарного объема запасов y и равны сумме ожидаемых затрат и затрат, связанных с пополнением запасов $g(i\ ,\ y) = L(y) + C(y - i)$. Оптимальная стратегия пополнения запасов $y(i)$ определяет минимальные затраты $g(i\ ,y(i)) = \min[g(i\ ,y)]$ при всех $y \geq i$.

Например, при $i = 0;\ 1;\ 2;\ 3$ наиболее вероятные объемы пополнения запасов $y = i$; $y = 4$; $y = 8$. При первом значении запасы не пополняются и затраты на пополнение запасов отсутствуют. При втором значении запасы пополняются до объема 4, а ожидаемые затраты уменьшаются в два раза. При третьем значении запасы пополняются до объема 8, а ожидаемые затраты отсутствуют. Таким образом, сравниваются затраты только для трех значений y, а $y(i)$ равно тому из них (подчеркнуто снизу), при котором эти затраты минимальны
$g(i\ ,y(i)) = \min[g(i\ ,i);\ g(i\ ,4);\ g(i\ ,8)]$.

Если минимум достигается сразу при двух значениях y, то из них выбирается большее, при котором спрос более удовлетворен.

Сначала рассматриваются уровни $i < q_1$ в порядке убывания.

При $i = 3$ имеем $g(3\ ,\ y(3)) = \min[g(3\ ,\ 3);\ g(3\ ,\ 4);\ g(3\ ,\ \underline{8})] = \min[12\ ;\ 6 + 1 + 4{,}5\ ;\ 5 + 4{,}5] = 9{,}5$. Другими словами $y(3) = 8$.

При $i = 2$ имеем $g(2\ ,\ y(2)) = \min[g(2\ ,\ 2);\ g(2\ ,\ 4);\ g(2\ ,\ \underline{8})] = \min[12\ ;\ 6 + 2 + 4{,}5\ ;\ 6 + 4{,}5] = 10{,}5$. Другими словами $y(2) = 8$.

При $i = 1$ имеем $g(1\ ,\ y(1)) = \min[g(1\ ,\ 1);\ g(1\ ,\ 4);\ g(1\ ,\ \underline{8})] = \min[12\ ;\ 6 + 3 + 4{,}5\ ;\ 7 + 4{,}5] = 11{,}5$. Другими словами $y(1) = 8$.

При $i = 0$ имеем $g(0\ ,\ y(0)) = \min[g(0\ ,\ \underline{0});\ g(0\ ,\ 4);\ g(0\ ,\ 8)] = \min[12\ ;\ 6 + 4 + 4{,}5\ ;\ 8 + 4{,}5] = 12$. Другими словами $y(0) = 0$.

Затем – уровни $q_1 \leq i < q_2$ в порядке убывания. В этих случаях вероятные объемы пополнения запасов $y = i$; $y = 8$.

При $i = 7$ имеем $g(7\ ,y(7)) = \min[g(7\ ,7);\ g(7\ ,\underline{8})] = \min[6\ ;\ 1 + 4{,}5] = 5{,}5$. Другими словами, $y(7) = 8$.

При $i = 6$ имеем $g(6\ ,y(6)) = \min[g(6\ ,\underline{6});\ g(6\ ,8)] = \min[6\ ;\ 2 + 4{,}5] = 6$. Другими словами, $y(6) = 6$, а запасы не пополняются.

При $i = 5;\ 4$, очевидно, запасы тоже не пополняются, так как иначе увеличиваются затраты на пополнение запасов по сравнению со случаем $\boldsymbol{i} = 6$.

Таким образом, запасы пополняются только в четырех случаях, в которых $\boldsymbol{y} > \boldsymbol{i}$. Оптимальная стратегия имеет вид

$$y(i) = \begin{cases} 8 & \text{при } i = 1;\ 2;\ 3;\ 7\ ; \\ i & \text{в остальных случаях;} \end{cases} \quad x(i) = \begin{cases} 8 - i & \text{при } i = 1;\ 2;\ 3;\ 7\ ; \\ 0 & \text{в остальных случаях.} \end{cases}$$

Следует отметить, что полученная стратегия пополнения запасов не является ($\boldsymbol{s}$, $\boldsymbol{S}$) стратегией. Действительно, если запасы выгодно пополнить при уровне $\boldsymbol{i}$ = 7, это должно быть выгодно и при меньшем уровне. Однако оказалось, что запасы невыгодно пополнять при $\boldsymbol{i}$ = 6; 5; 4; 0, что противоречит определению ($\boldsymbol{s}$, $\boldsymbol{S}$) стратегии.

3.8 Работа по теме «Модели управления запасами»

Задание 1. Для выпуклой функции затрат найти оптимальную стратегию пополнения запасов в случае, когда спрос $\boldsymbol{Q}$ равномерно распределен: $\boldsymbol{Q} = \overline{1,10}$; $\boldsymbol{p(q_i)}$ = 0,1, а параметры указаны в таблице Г.1.

Таблица Г.1

№	π	c	K	h	№	π	c	K	h
1	13	1	5	5	14	14	1	5	1
2	22	5	5	5	15	15	2	3	3
3	24	7	3	10	16	16	1	5	6
4	28	3	10	10	17	17	2	4	1
5	30	5	8	10	18	18	1	9	3
6	85	10	30	20	19	19	2	8	6
7	36	5	12	5	20	20	2	8	4
8	15	1	6	10	21	21	2	5	2
9	32	5	10	2	22	22	1	9	1
10	32	1	13	2	23	23	2	4	2

11	11	2	3	3	24	24	1	11	3
12	12	1	3	4	25	25	2	6	4
13	13	1	6	2	26	26	3	4	1

Задание 2. Для единовременного штрафа найти оптимальную стратегию пополнения запасов в случае, когда спрос $\boldsymbol{Q}$ принимает два значения: $\boldsymbol{q_1}$, $\boldsymbol{q_2}$ с вероятностями $p_1 = p_2 = 0{,}5$, параметры указаны в таблице Г.2, а затраты хранения равны нулю. В случае не единственности решения выбрать вариант с максимально возможным удовлетворением спроса.

Таблица Г.2

№	π	c	K	q_1	q_2	№	π	c	K	q_1	q_2
1	13	1	5	2	8	14	14	1	5	3	8
2	22	5	5	4	8	15	15	2	3	2	9
3	24	7	3	3	9	16	16	1	5	1	10
4	28	3	10	2	7	17	17	2	4	3	8
5	30	5	8	3	8	18	18	1	9	4	7
6	85	10	30	4	7	19	19	2	8	2	7
7	36	5	12	2	6	20	20	2	8	2	7
8	15	1	6	2	9	21	21	2	5	2	10
9	32	5	10	1	5	22	22	1	9	3	10
10	32	1	13	1	8	23	23	2	4	2	9
11	11	2	3	3	9	24	24	1	11	3	9
12	12	1	3	4	10	25	25	2	6	3	8
13	13	1	6	2	8	26	26	3	4	4	8

5 Имитационное моделирование

5.1 Область применения

Рассмотренные выше модели управления запасами наиболее успешно применяются в случае, когда плановый период не превышает 10 лет. В противном случае соответствующая многошаговая модель, как правило, содержит лишь среднегодовые показатели. При этом влияние результатов оптимизации на показатели, характеризующие текущие операции на отрезках продолжительностью от 1 недели до 1 месяца, в явной форме не учитывается.

Ограниченность анализа, основанного на использовании метода линейного программирования, обусловлена также отсутствием достоверной информации относительно будущего. Неопределенность прогнозов в той или иной степени свойственна всем задачам планирования. Но эта неопределенность во многих случаях либо не относится к **существу** такого рода задач, либо является **частичной**, т. е. отражает отсутствие информации лишь относительно **некоторых** из фигурирующих в модели параметров. В этих случаях для оценки влияния неопределенности на окончательное решение оказывается достаточным проанализировать модель на чувствительность (подраздел 4.6).

Ориентированные на использование математического аппарата вероятностные динамические модели, в частности модели управления запасами и модели массового обслуживания, также обладают аналогичным недостатком. Чтобы найти численное решение для такого рода моделей, приходится не только ограничиваться случаями, когда операционная система обладает небольшой размерностью, но и вводить упрощающие предположения относительно самой схемы функционирования исследуемой системы. Так, например, с помощью характерных для теории массового обслуживания математических методов, которые аналогичны методам, изложенным в третьем разделе, невозможно полностью **адекватным** образом проанализировать «поведение» очередей в ремонтных мастерских. Соответствующие стохастические модели могут служить лишь грубым приближением протекающих в действительности процессов формирования и обслуживания очередей.

Постоянно возрастающее число публикаций по исследованию операций свидетельствует о непрерывном поиске новых путей преодоления упомянутых выше трудностей. Однако можно утверждать, что операционные методы в обозримом будущем не смогут обеспечить исчерпывающего анализа таких задач организационного управления, как:

1) *Формирование инвестиционной политики при перспективном планировании*. Инвестиционная политика в случае крупной фирмы должна учитывать финансирование научно-исследовательских и опытно-конструкторских работ при создании новых видов продукции, оценку степени риска при планировании тех или иных комплексов работ, сравнительную оценку вариантов слияния с другой фирмой и приобретения последней и т. п.

2) *Выбор средств обслуживания при текущем планировании*. Сюда относятся задачи определения количества: прилавков в большом торговом центре, бензоколонок на АЗС, лифтов в строящемся здании.

3) *Разработка планов с обратной информационной связью и операционных предписаний*. Сюда относятся задачи выработки правил составления календарных планов: на предприятиях с мелкосерийным производством, комбинатах по ремонту различных изделий, вычислительных центрах и т. д. Эти предписания должны учитывать гарантийные сроки выполнения заказов, потребности в обслуживании, наличные ресурсы, производственные мощности, темпы повышения квалификации рабочих, уровень снабжения сырьем. По мере поступления информации о новых уже выполненных заказах предприятие сталкивается с задачей уточнения или полного пересмотра своих планов-графиков.

Почему описанные выше классы задач с трудом поддаются анализу? Причина заключается в необходимости одновременного учета факторов неопределенности, динамической взаимной обусловленности текущих решения и последующих событий, в комплексной взаимозависимости между управляемыми переменными исследуемой операционной системы. Такого рода «глобальные» системные задачи обладают слишком большой размерностью и наличием слишком большого количества внутренних взаимосвязей, в силу чего их не удается решить методами математического программирования. Наиболее эффективным из существующих в настоящее время операционных методов, выходящих за рамки математического программирования, является метод **имитационного моделирования на ЭВМ**.

При имитационном подходе, прежде всего, строится экспериментальная модель системы. Затем производится сравнительная оценка конкретных вариантов функционирования системы путем «проигрывания» различных ситуаций на рассматриваемой модели.

Каждый наверняка припомнит случай, когда ему приходилось сталкиваться с имитационными ситуациями. Так, например, у всякого, кто посещал парк «Дисней-лэнд» и совершал лодочную прогулку по его «джунглям», создавалось впечатление, что он находится в настоящих джунглях. Более серьезными примерами могут служить демонстрации, организуемые в планетариях, и стенды-панорамы в зоологических музеях. А тем, кто служил в армии, хорошо известно, что полевые учения и маневры заключаются в имитации боевых действий.

Однако такие имитации представляются слишком неудобными и дорогостоящими, чтобы решать задачи организационного управления. Более предпочтительным является представление сложной функциональной системы с помощью логико-математической модели, «заложенной» в ЭВМ. При этом факторы неопределенности, динамические характеристики и все взаимосвязи между элементами исследуемой системы представляются в виде формул, хранящихся в памяти ЭВМ. Имитирование системы начинается с некоторого вполне определенного исходного состояния. В результате принимаемых решений, а также ряда случайных событий система переходит в последующие моменты времени в другие состояния. Эволюционный процесс, таким образом, продолжается до тех пор, пока не наступит окончание планового периода. Такое отражение в ЭВМ реального процесса длительностью в несколько лет за несколько секунд называют **сжатием времени**.

Отсутствие единой теории имитационного моделирования на ЭВМ есть одновременно «и благо, и зло». Положительным здесь является то, что имеется возможность строить имитационные модели любой степени сложности при огромном количестве динамических взаимосвязей, а также при отсутствии стационарности и наличии взаимно коррелированных стохастических элементов. Отрицательным же элементом является то, что по мере усложнения модели при анализе данных приходится все более полагаться на математическую статистику. Поэтому в этой ситуации оценка адекватности модели оказывается весьма затруднительной.

5.2 Целевая установка

При построении имитационной модели, предназначенной для углубленного анализа проблем организационного управления, преследуют, по крайней мере, одну из следующих целей [2]:

1) *Изучение действующей функциональной системы*. Например, рассмотрим фирму, которая зарегистрировала увеличение числа заказов на свою продукцию и обратила внимание на заметное ухудшение качества обслуживания своих клиентов в части соблюдения сроков выполнения этих заказов. У нее может появиться желание построить имитационную модель, с помощью которой можно было бы изучить, каким образом существующие процедуры определения сроков выполнения заказов, календарного планирования производства и оформления заявок на поставку сырья порождают наблюдаемые задержки.

2) *Анализ гипотетической функциональной системы*. Обратимся к больнице, руководство которой рассматривает вопрос внедрения новой системы управления запасами медицинских препаратов. Для грамотного внедрения потребуется создать имитационную модель больницы, позволяющую с помощью ретроспективных данных определить средний уровень средств, находящихся в запасах, и как часто будут возникать нехватки различных видов препаратов в случае, если будет реализован предлагаемый план.

3) *Проектирование более совершенной функциональной системы*. Рассмотрим мелкосерийное производство, в котором станочные мощ-

ности распределены в соответствии приоритетами выполняемых работ. Для определения эффективной системы приоритетов потребуется создать имитационную модель производства, позволяющую с помощью ретроспективных данных определить, как часто будут возникать задержки выполнения работ и каков при этом будет коэффициент использования оборудования предприятия.

5.3 Этапы построения и использования имитационной модели

В целях практической реализации метода имитационного моделирования необходимо выполнить следующие виды работ [18, 22, 25, 27]:

1. *Построение модели*. Имитационная модель практически не отличается от любой операционной модели другого типа. Единственное отличие состоит в выборе единицы времени. Например, если модель должна помочь в выборе одного из двух вариантов размещения нового складского помещения, то, по-видимому, нет необходимости делить плановый период на часы или дни: вполне достаточно в качестве единицы времени выбрать неделю. Однако, если с помощью модели нужно решить сколько в новом складе должно быть погрузочно-разгрузочных платформ, то, возможно, возникает необходимость рассмотреть отрезки времени продолжительностью от 5 до 15 минут.

2. *Разработка проекта эксперимента*. Для уменьшения вероятности той или иной ошибки необходимо подробно разработать сопровождающие эксперимент процедуры до того, как модель будет «приведена в действие». То есть необходимо тщательно продумать, какие функциональные характеристики имитируемой системы планируется измерять. Кроме того, следует определить, с помощью какого метода математической статистики будут учитываться изменения экспериментальных данных, полученных в результате этих измерений.

3. *Разработка программного обеспечения*. Весь имитационный эксперимент проводится на ЭВМ. Другими словами, все стадии эво-

люционного развития модели, так же как генерирование случайных событий, протекают в ЭВМ. Поэтому для реализации «вычислительного варианта» модели необходимо выбрать язык программирования с подходящим интерфейсом.

Рассмотрим примеры имитационного моделирования систем

5.4 Оценка надежности систем

▪ _Пример_ 1. Система состоит из двух блоков, соединенных последовательно. Поэтому система отказывает при отказе хотя бы одного блока. Первый блок содержит два элемента: *A, B* (они соединены параллельно) и отказывает при одновременном отказе обоих элементов. Второй блок содержит один элемент *C* и отказывает при отказе этого элемента. Найти методом статистических испытаний оценку P^* надежности (вероятности безотказной работы) системы, зная вероятности безотказной работы элементов: $P(A) = 0{,}8$, $P(B) = 0{,}85$, $P(C)$ = $= 0{,}6$. Найти абсолютную погрешность $|P - P^*|$, где P – надежность системы, вычисленная аналитически. Произвести 50 испытаний.

Решение. Рассмотрим граф системы на рисунке 5.1.

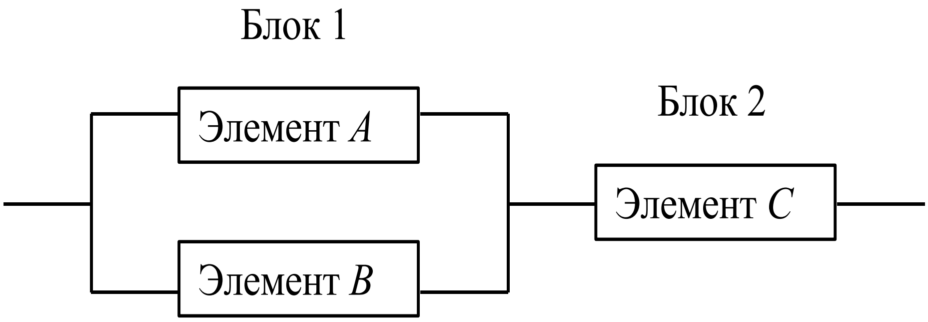

Рисунок 5.1 – Граф системы

В первом испытании разыграем события *A, B, C*, состоящие в безотказной работе элементов *A, B, C*, по правилу пункта 1.3. Выберем для этого три последовательных случайных числа $R(0,\ 1)$, например, 0,10, 0,09, и 0,73. Если случайное число меньше вероятности события, то событие наступило, а если больше – то не наступило.

Результаты испытания будем записывать в расчетную таблицу 5.1.

Таблица 5.1

<table>
<tr><th rowspan="3">Номер испы-тания</th><th rowspan="3">Блок</th><th colspan="3">Случайные числа, моделирующие элементы</th><th colspan="5">Заключение о работе</th></tr>
<tr><th colspan="3"></th><th colspan="3">элементов</th><th rowspan="2">бло-ков</th><th rowspan="2">сис-темы</th></tr>
<tr><th>A</th><th>B</th><th>C</th><th>A</th><th>B</th><th>C</th></tr>
<tr><td rowspan="2">1</td><td>Первый</td><td>0,10</td><td>0,09</td><td></td><td>+</td><td>+</td><td></td><td>+</td><td rowspan="2">–</td></tr>
<tr><td>Второй</td><td></td><td></td><td>0,73</td><td></td><td></td><td>–</td><td>–</td></tr>
<tr><td rowspan="2">2</td><td>Первый</td><td>0,25</td><td>0,33</td><td></td><td>+</td><td>+</td><td></td><td>+</td><td rowspan="2">–</td></tr>
<tr><td>Второй</td><td></td><td></td><td>0,76</td><td></td><td></td><td>–</td><td>–</td></tr>
<tr><td rowspan="2">3</td><td>Первый</td><td>0,52</td><td>0,01</td><td></td><td>+</td><td>+</td><td></td><td>+</td><td rowspan="2">+</td></tr>
<tr><td>Второй</td><td></td><td></td><td>0,35</td><td></td><td></td><td>+</td><td>+</td></tr>
<tr><td rowspan="2">4</td><td>Первый</td><td>0,86</td><td>0,34</td><td></td><td>–</td><td>+</td><td></td><td>+</td><td rowspan="2">–</td></tr>
<tr><td>Второй</td><td></td><td></td><td>0,67</td><td></td><td></td><td>–</td><td>–</td></tr>
</table>

Поскольку $P(A) = 0{,}8$ и $0{,}10 < 0{,}8$, то событие A наступило, т. е. элемент A в этом испытании работает безотказно. Так как $P(B) = 0{,}85$ и $0{,}09 < 0{,}85$, то событие B наступило, т. е. элемент B в этом испытании работает безотказно.

Таким образом, оба элемента первого блока работают. Следовательно, работает и сам первый блок. В соответствующих клетках (таблица 5.1) ставим знак плюс.

Поскольку $P(C) = 0{,}6$ и $0{,}73 > 0{,}6$, то событие C не наступило, т. е. элемент C получает отказ. Другими словами, второй блок, а значит и вся система получают отказ. В соответствующих клетках (таблица 5.1) ставим знак минус.

Аналогично разыгрываются и следующие три испытания (таблица 5.1) и т. д. до пятидесяти испытаний.

Произведя 50 испытаний, получим, например, что в 28 из них система работала безотказно. В качестве оценки искомой надежности P примем относительную частоту $P^* = 28/50 = 0{,}56$.

Найдем надежность системы P аналитически. Вероятности безотказной работы первого и второго блоков соответственно равны:

$$P_1 = 1 - P(\overline{A}) \cdot P(\overline{B}) = 1 - 0{,}2 \cdot 0{,}15 = 0{,}97, \quad P_2 = P(C) = 0{,}6\,.$$

Вероятность безотказной работы системы

$$P = P_1 \cdot P_2 = 0{,}97 \cdot 0{,}6 = 0{,}582\,.$$

Искомая абсолютная погрешность $|P - P^*| = 0{,}582 - 0{,}56 = 0{,}022$.

5.5 Оценка абсолютной пропускной способности СМО

Пример 2. В трехканальную линию телефонной связи с отказами поступает пуассоновский поток заявок. Время между поступлениями двух последовательных вызовов распределено по показательному закону с плотностью $f(\tau) = 5e^{-5\tau}$. Длительность обслуживания каждого вызова равна 0,5 минут. Найти методом статистических испытаний оценку абсолютной A и относительной q пропускной способности СМО в период наблюдения длительностью T = 4 минуты.

Решение. Рассмотрим испытание рассматриваемой системы. Пусть $T_1 = 0$ – момент поступления первый вызов. Он поступит в первый канал и будет им обслужен. Момент окончания обслуживания первого вызова $T_1 + 0{,}5 = 0 + 0{,}5 = 0{,}5$ мин. В счетчик обслуженных вызовов записываем единицу.

Моменты поступления последующих вызовов найдем по формуле

$$T_i = T_{i-1} + \tau_i\,,$$

где τ_i – длительность времени между двумя последовательными вызовами с номерами i – 1 и i.

Возможные значения τ_i разыгрываем по формуле

$$\tau_i = -\frac{\ln r_i}{\lambda}\,.$$

Учитывая, что, по условию, λ = 5, получим $\tau_i = 0{,}2(-\ln r_i)$.

Случайные числа r_i берем из таблицы случайных чисел (*см. приложение Е*), начиная с первой строки сверху. Для нахождения времени между поступлениями первого и второго вызова возьмем случайное число $r = 0{,}10$. Тогда $\tau_2 = 0{,}2 \cdot (-\ln 0{,}10) = 0{,}2 \cdot 2{,}30 = 0{,}460$ мин. Первый вызов поступил в момент $T_1 = 0$. Следовательно, второй вызов поступит в момент времени $T_2 = T_1 + 0{,}460 = 0 + 0{,}460 = 0{,}460$ мин. В этот момент первый канал еще занят обслуживанием первого вызова, поэтому второй вызов поступит во второй канал и будет им обслужен. Момент окончания обслуживания второго вызова $T_2 + 0{,}5 = 0{,}460 + 0{,}5 = 0{,}960$ мин. В счетчик обслуженных заявок записываем единицу.

По очередному случайному числу $r = 0{,}09$ разыграем время τ_3 между поступлениями второго и третьего вызова

$$\tau_3 = 0{,}2 \cdot (-\ln 0{,}09) = 0{,}2 \cdot 2{,}41 = 0{,}482 \text{ мин.}$$

Второй вызов поступил в момент $T_2 = 0{,}460$ мин. Следовательно, третий вызов поступит в момент времени $T_3 = T_2 + 0{,}482 = 0460 + + 0{,}482 = 0{,}942$ мин. В этот момент первый канал уже свободен, и третий вызов поступит в первый канал. Момент окончания обслуживания третьего вызова $T_3 + 0{,}5 = 0{,}942 + 0{,}5 = 1{,}442$ мин. В счетчик обслуженных заявок записываем единицу.

Дальнейший расчет производят аналогично, причем если в момент поступления вызова все каналы заняты (этот момент меньше каждого из моментов окончания обслуживания), то в счетчик отказов добавляют единицу. Результаты расчетов вносим в таблицу 5.2.

Следует отметить, что обслуживание 20-й заявки закончится в момент $4{,}148 > 4$, поэтому этот вызов получает отказ. Испытание прекращают (в таблице 5.2 записывают «стоп»), если момент поступления вызова $T > 4$ мин. К этому моменту поступило 20 вызовов, из которых обслужено $x_1 = 12$ вызовов.

Выполнив аналогично еще пять испытаний системы, получим

$$x_2 = 15,\ x_3 = 14,\ x_4 = 12,\ x_5 = 13,\ x_6 = 15,$$

В качестве оценки A абсолютной пропускной способности системы примем выборочную среднюю

$$\overline{A} = \overline{x} / T = \frac{1}{6}\sum_{i=1}^{6} x_i / 4 = 3{,}375 \quad \text{вызовов/мин}.$$

Оценка $\overline{q}$ относительной пропускной способности

$$\overline{q} = x_1 / 20 = 0{,}6\,.$$

Таблица 5.2

Номер заявки i	Число r_i	$-\ln r_i$	Время τ_i, мин	Момент T_i, мин	Моменты освобождения канала, мин			Счетчик	
					1	2	3	обслуженных вызовов	отказов
1				0	0,500			1	
2	0,10	2,30	0,460	0,460		0,960		1	
3	0,09	2,41	0,482	0,942	1,442			1	
4	0,73	0,32	0,064	1,006		1,506		1	
5	0,25	1,39	0,278	1,284			1,784	1	
6	0,33	1,11	0,222	1,506	2,006			1	
7	0,76	0,27	0,054	1,560		2,060		1	
8	0,52	0,65	0,130	1,690					1
9	0,01	4,60	0,920	2,610	3,110			1	
10	0,35	1,05	0,210	2,820		3,320		1	
11	0,86	0,15	0,030	2,850			3,350	1	
12	0,34	1,08	0,216	3,066					1
13	0,67	0,40	0,080	3,146	3,646			1	
14	0,35	1,05	0,210	3,356		3,856		1	
15	0,48	0,73	0,146	3,502			4,002		1
16	0,76	0,27	0,054	3,556					1
17	0,80	0,22	0,044	3,600					1
18	0,95	0,05	0,010	3,610					1
19	0,90	0,10	0,020	3,630					1
20	0,91	0,09	0,018	3,648	4,148				1
21	0,17	1,77	0,354	4,002					
				(Стоп)					
							Итого	$x_1 = 12$	8

На рисунке 5.2 приведена «имитационная картина» процесса обслуживания трехканальной линии телефонной связи: а – отражает работу первого канала, б – второго, в – третьего, г – периоды, когда вся линия занята.

В соответствии с таблицей 5.2 нижний участок ломаной линии (а – в) отражает период, когда канал свободен, а верхний участок – когда занят. Если вызов поступает в периоды, отмеченные в случае г, то он получает отказ. Общая продолжительность пяти периодов составляет 0,99 мин. и позволяет оценить вероятность отказа $P_{отк}$ = 0,99/4 = = 0,475.

Следует отметить, что для уточнения полученных оценок необходимо рассмотреть более длительный период наблюдения.

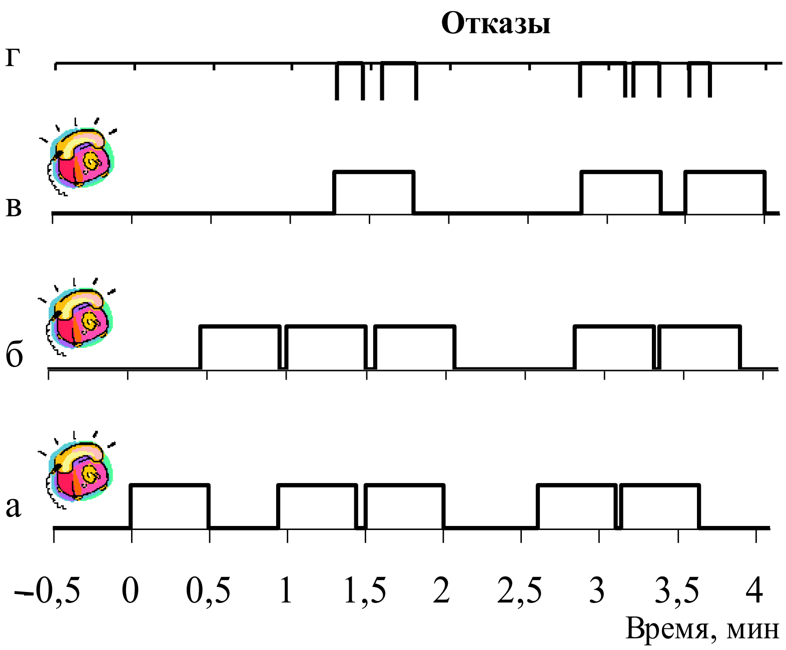

Рисунок 5.2 – Модель трехканальной линии телефонной связи

5.6 Оценка затрат, связанных с управлением запасов

- *Пример 3.* Фирма «Надежный компьютер», которая реализует компьютеры, стремится оптимизировать затраты, связанные с управлением запасов. Пусть суточный спрос является равномерно распределенной величиной, которая принимает значения $\boldsymbol{q} = 1, \ldots, 10$, а вероятности $\boldsymbol{p}(\boldsymbol{q}) = 0{,}1$ при всех значениях $\boldsymbol{q}$. Рассматриваются затраты с такими же параметрами, как в примере 1: $\boldsymbol{\pi} = 10$; $\boldsymbol{c} = 3$; $\boldsymbol{K} = 8$; $\boldsymbol{h} = 2$. Найти методом статистических испытаний оценку суточных суммарных затрат для оптимальной стратегии пополнения запасов и исследовать чувствительность этой оценки к параметрам стратегии.

Решение. Рассмотрим испытание рассматриваемой фирмы. Возможные значения спроса q_n разыгрываем в приложении «Excel» по формуле

$$q_n = \text{округлвверх}(10 \cdot r_n).$$

Случайные числа r_n берем из таблицы случайных чисел в *приложении E*, начиная с первой строки сверху.

Пусть в начале первых суток запас компьютеров равен нулю (остаток в конце нулевых суток $i_0 = 0$). Согласно оптимальной стратегии в статической модели $S = 6$, $s = 3$ (пример 1 в п. 4.7) в начале первых суток проводится пополнение запасов до 6-ти компьютеров. Соответствующие затраты составят $\boldsymbol{K} + \boldsymbol{c} \cdot S = 8 + 3 \cdot 6 = 26$ д. е. Остаточный запас в конце первых суток $i_1 = y_1 - q_1 = 6 - 1 = 5$, а затраты на хранение этих запасов составят $\boldsymbol{h} \cdot i_1 = 2 \cdot 5 = 10$ д. е. Таким образом, суммарные затраты в первые сутки составят 36 д. е.

В начале вторых суток запас компьютеров ($y_2 = i_1 = 5$) больше порогового значения $s = 3$. Поэтому запас не пополняется. Остаток в конце вторых суток $i_2 = y_2 - q_2 = 5 - 1 = 4$, а затраты на хранение этих запасов составят $\boldsymbol{h} \cdot i_2 = 2 \cdot 4 = 8$ д. е. Суммарные затраты в течение вторых суток составят 8 д. е.

В начале третьих суток запас компьютеров ($y_3 = i_2 = 4$) больше порогового значения $s = 3$. Поэтому запас не пополняется. Однако этот запас меньше спроса. Поэтому остаток в конце третьих суток $i_3 = 0$, а затраты, связанные с потерей прибыли из-за неудовлетворенного спроса, составят $\boldsymbol{\pi} \cdot (q_3 - y_3) = 10 \cdot (8 - 4) = 40$ д. е. Суммарные затраты в течение третьих суток составят 40 д. е.

Дальнейший расчет производят аналогично, а результаты расчета всех видов затрат, а также суммарные затраты в течение первых 20-ти суток вносим в таблицу 5.3.

Таблица 5.3

Номер суток n	Число r_n	Спрос q_n	Запас y_n	Остаток i_n	Затраты в д. е. на:			Сумма затрат, д. е.
					пополнение	хранение	неудов. спрос	
0				0				
1	0,10	1	6	5	26	10		36
2	0,09	1	5	4		8		8
3	0,73	8	4	0			40	40
4	0,25	3	6	3	26	6		32
5	0,33	4	3	0			10	10
6	0,76	8	6	0	26		20	46
7	0,52	6	6	0				0
8	0,01	1	6	5	26	10		36
9	0,35	4	5	1		2		2
10	0,86	9	6	0	23		30	53
11	0,34	4	6	2	26	4		30
12	0,67	7	6	0	20		10	30
13	0,35	4	6	2	26	4		30
14	0,48	5	6	1	20	2		22
15	0,76	8	6	0	23		20	43
16	0,80	8	6	0	26		20	46
17	0,95	10	6	0	26		40	66
18	0,90	9	6	0	26		30	56
19	0,91	10	6	0	26		40	66
20	0,17	2	6	4	26	8		34
					Средние затраты			**34,3**

В качестве оценки суточных суммарных затрат примем выборочную среднюю за 20 суток, которая составляет 34,3 д. е.

Если плановый период больше 20-ти суток, то составляется программа расчета затрат на любом языке программирования, которая за доли секунды рассчитает затраты за любое количество суток.

Результаты проведенных таким образом расчетов показывают, что выборочная средняя оценка суточных затрат быстро стабилизируется с увеличением количества суток. На рисунке 5.2 приведена зависимость этой оценки от числа суток планового периода.

Стабилизация оценки на уровне 31,7 д. е. начинается с периода наблюдения в 500 суток.

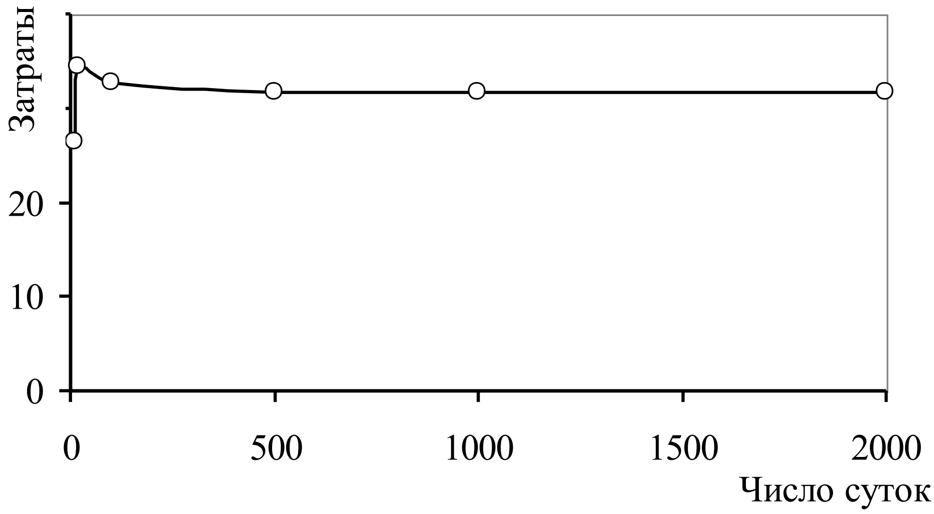

Рисунок 5.2 – Оценка среднесуточных затрат фирмы

Следует отметить, что среднесуточные затраты достигают минимума при $S = 8$, $s = 3$, в отличие от решения статической модели. Этот минимум составляет 29,42 д. е., что на 7,2 % меньше, чем на рисунке 5.2. Это связано с тем, что начальный запас реальной фирмы меняется каждые сутки в отличие от статической модели, которая оценивает затраты только за сутки. В реальности условия модели являются «случайными» и проследить их удается только методом статистических испытаний. Даже в случае стратегии $S = 11$, $s = 3$ среднесуточные затраты составят 30,14 д. е., что на 4,9 % меньше чем в случае решения статической модели. Однако при снижении уровня S пополнения запасов в стратегии $S = 5$, $s = 3$ затраты возрастут до 34,28 д. е. Аналогичный эффект наблюдается при изменении порогового значения s.

5.7 Работа по теме «Имитационное моделирование»

Фирма «Надежный компьютер», которая реализует компьютеры, стремится оптимизировать затраты, связанные с управлением запасов. Пусть суточный спрос является равномерно распределенной величиной, которая принимает значения $\boldsymbol{q} = 1, \ldots, 10$, а вероятности $\boldsymbol{p}(\boldsymbol{q}) = 0{,}1$ при всех значениях $\boldsymbol{q}$. Рассматриваются затраты при оптимальной стратегии управления запасами в случае статической модели с такими же параметрами, как в задании 1 (*см. приложение Г*). Построить имитационную модель и найти методом статистических испытаний оценку суточных суммарных затрат за 20 суток. С помощью программы на любом алгоритмическом языке исследовать чувствительность этой оценки к числу суток $\boldsymbol{n}$ и параметрам стратегии $\boldsymbol{S}$, $\boldsymbol{s}$. Определить оптимальную стратегию пополнения запасов согласно этой модели.

6 Список литературы

1. Линдин, Г. Л. Вероятностные математические модели [Текст]: учебное пособие для студентов специальности 010400.68 – «Прикладная математика и информатика» с квалификацией «магистр» / Г. Л. Линдин. – Новокузнецк : РИО НФИ КемГУ, 2012. – 125 с

2. Свешников, А. А. Прикладные методы теории вероятностей : учебник / А. А. Свешников. — Санкт-Петербург : Лань, 2012. — 480 с. — ISBN 978-5-8114-1219-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/reader/book/3184/#3. — Режим доступа: для авториз. пользователей.

3. Федоткин, М. А. Модели в теории вероятностей [Текст]: учебник / М. А. Федоткин. – Москва : Физматлит ; Нижний Новгород : Нижегородский гос. Университет, 2012. – 608 с. – Гриф УМС «Допущено».