

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТРАНСПОРТА**

---

---

Кафедра «ПРИКЛАДНАЯ МАТЕМАТИКА»

**И. В. МАКСИМЕЙ, В. С. СЕРЁГИНА**

# **ЗАДАЧИ И МОДЕЛИ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ**

**Часть 2**

**Методы нелинейного и стохастического  
программирования**

Гомель 1999

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТРАНСПОРТА**

---

---

Кафедра «ПРИКЛАДНАЯ МАТЕМАТИКА»

**И. В. МАКСИМЕЙ, В. С. СЕРЁГИНА**

# **ЗАДАЧИ И МОДЕЛИ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ**

**Часть 2**

**Методы нелинейного и стохастического  
программирования**

*Допущено Министерством образования Республики Беларусь  
в качестве учебного пособия  
для студентов инженерно-технических специальностей  
высших учебных заведений*

Под общей редакцией профессора **И. В. Максимея**

Гомель 1999

УДК 519.85 (075.8)  
М 171

**И. В. Максимей, В. С. Серёгина**

М 171 Задачи и модели исследования операций. Ч. 2. Методы нелинейного и стохастического программирования: Уч. пособие. – Гомель: БелГУТ, 1999. – 103 с.  
ISBN 985-6550-10-6

Содержатся технологии формализации и методики решения производственных задач с помощью пяти классов математических моделей: нелинейного программирования, вероятностных моделей на основе метода статистических испытаний, моделей случайных процессов, моделей теории надёжности систем, аналитических моделей массового обслуживания.

Предназначено для студентов и аспирантов инженерно-технических специальностей высших учебных заведений.

Рецензенты: кафедра «Прикладная математика» Национальной политехнической академии Украины «Киевский политехнический институт»;  
докт. техн. наук, профессор Белорусского государственного университета **О. М. Тихоненко**;  
проректор БелГУТа докт. техн. наук, профессор **В. Я. Негрей**.

Учебное издание

Иван Васильевич Максимей  
Валентина Серафимовна Серёгина

## **ЗАДАЧИ И МОДЕЛИ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ Часть 2. Методы нелинейного и стохастического программирования**

Учебное пособие

Редактор Н. А. Дашкевич  
Технический редактор Т. А. Пугач  
Корректор И. И. Эвентов

Подписано в печать  
Формат бумаги  $60 \times 84 \frac{1}{16}$ . Бумага офсетная № 1. Гарнитура Таймс.  
Печать офсетная. Усл. печ. л. Уч.-изд. л. 6,17. Тираж 200 экз.  
Зак. № . Изд. № 3083

Редакционно-издательский отдел БелГУТа,  
246653, г. Гомель, ул. Кирова, 34. Лицензия ЛВ № 57 от 22.10.97.

Ротапринт типографии БелГУТа,  
246022, г. Гомель, ул. Кирова, 34. Лицензия ЛП № 75 от 12.03.93.

## ВВЕДЕНИЕ

Построение математической модели (ММ) для описания сложных технических систем (СТС), имеющих вероятностную природу, является достаточно трудной задачей. Класс методов решения подобных задач очень разнообразен. Во второй части пособия предложены наиболее известные подходы к формализации вероятностных процессов в СТС, которые широко распространены и обладают необходимыми технологическими возможностями для решения задач на транспорте или в строительстве. Ниже рассматриваются пять классов ММ: модели нелинейного программирования, вероятностные модели на основе метода статистических испытаний, модели для описания случайных процессов, модели надёжности СТС, аналитические модели массового обслуживания.

Первый класс ММ позволяет решать задачи нелинейной оптимизации СТС. Формулируются особенности постановки задач нелинейного программирования. Излагаются наиболее известные алгоритмы одномерной оптимизации нелинейных ММ СТС. Формулируются алгоритмы численных методов безусловной оптимизации СТС. Излагается алгоритм широко известного метода неопределённых множителей Лагранжа для решения нелинейных задач. Завершается рассмотрение класса нелинейных ММ постановкой задач и основных алгоритмов решения задач дискретной оптимизации СТС.

Второй и третий классы ММ используют в качестве аппарата формализации метод статистических испытаний (метод Монте-Карло). Приводятся примеры использования метода Монте-Карло при исследовании СТС. Излагаются наиболее известные алгоритмы моделирования псевдослучайных величин. На их основе излагаются примеры: моделирование случайных процессов в СТС, представления случайных процессов временными рядами, моделирование поведения СТС марковскими процессами.

Четвёртый класс ММ составляют модели надёжности СТС. Даются необходимый минимум сведений из теории надёжности. На этой основе определяются наиболее известные меры повышения надёжности транспортных систем и типовые модели расчёта параметров надёжности СТС. Рассматриваются модели для расчёта надёжности невосстанавливаемых резервированных элементов СТС. Предлагаются наиболее известные подходы к моделированию надёжности систем с монотонной структурой.

Пятый класс ММ составляют аналитические модели массового обслуживания. Формулируются предпосылки описания СТС аппаратом теории массового обслуживания. Приведены алгоритмы построения аналитических моделей СТС в виде одноканальных и многоканальных систем массового обслуживания (СМО). Рассматривается технология оптимизации СМО с ожиданием и даются примеры использования СМО при решении эксплуатационных задач транспорта.

Первый раздел пособия написан кандидатом физико-математических наук, доцентом В. С. Серёгиной, остальные разделы и научное редактирование пособия выполнены доктором технических наук, профессором И. В. Максимеем.

Авторы выражают благодарность рецензентам за ценные советы и замечания, способствовавшие улучшению содержания пособия.

Авторы надеются, что пособие поможет студентам освоить методы исследований операций при изучении вероятностных процессов на аналитических ММ и тем самым сэкономит их время на подготовку и приобретение навыков системного анализа сложных технических систем.

# 1. МОДЕЛИ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

## 1.1. Основные понятия нелинейного программирования

В общем виде задача нелинейного программирования формулируется следующим образом:

минимизировать  $f(x_1, x_2, \dots, x_n)$  при ограничениях

$$\begin{aligned} g_j(x_1, x_2, \dots, x_n) &\leq 0, \quad j = \overline{1, m}; \\ h_j(x_1, x_2, \dots, x_n) &= 0, \quad j = \overline{1, k}. \end{aligned} \quad (1.1)$$

Задачи такого рода возникают как в теории управления, так и в естественных науках, и их систематическое исследование, начатое в конце 40-х годов, привело к возникновению самостоятельной научной дисциплины – нелинейного программирования.

Введём некоторые обозначения и дадим ряд необходимых для дальнейшего изложения определений.

Для простоты совокупность неизвестных  $(x_1, x_2, \dots, x_n)$  обозначим через  $X$  и будем считать  $X^*$  точкой в пространстве  $R^n$ .

Тогда соотношение (1.1) можно записать в более компактном виде:

$$\begin{aligned} \text{минимизировать } f(X) \text{ при ограничениях} \\ g(X) \leq 0, \quad h(X) = 0. \end{aligned} \quad (1.1')$$

Точка  $X^*$  в  $R^n$ , удовлетворяющая условиям  $g(X^*) \leq 0$  и  $h(X^*) = 0$ , называется **решением или глобальным решением** задачи (1.1), если  $f(X^*) \leq f(X)$  для всех  $X$ , удовлетворяющих условиям  $g(X) \leq 0$  и  $h(X) = 0$ .

Говорят, что точка  $X^*$  в  $R^n$  является **локальным решением** задачи (1.1), если для всех  $X \neq X^*$ , удовлетворяющих условиям  $g(X) \leq 0$  и  $h(X) = 0$ , и

$$\|X - X^*\| \leq \delta, \quad \text{где } \|X - X^*\| = \sqrt{\sum_{j=1}^n (x_j - x_j^*)^2} \text{ для некоторого числа } \delta > 0.$$

Говорят, что точка  $X^*$  в  $R^n$  является **единственным локальным решением** задачи (1.1), если для всех  $X \neq X^*$ , удовлетворяющих условиям  $g(X) \leq 0$  и  $h(X) = 0$ , имеет место строгое неравенство  $f(X^*) < f(X)$ , а для некоторого числа  $\delta > 0$  выполняется неравенство  $\|X - X^*\| \leq \delta$ .

В частном (и довольно редком) случае ограничения  $g(X) \leq 0$  и  $h(X) = 0$  могут отсутствовать, т. е. задача (1.1) формулируется так:

$$\text{минимизировать } f(X). \quad (1.1'')$$

В этих случаях мы говорим, что задача (1.1'') имеет **безусловное глобальное решение**  $X^*$ , если  $f(X^*) < f(X)$  для любых  $X$ .

Если неравенство  $f(X^*) < f(X)$  выполняется не для всех  $X$ , а только для точек, лежащих в  $\varepsilon$ -окрестности точки  $X^*$  ( $\varepsilon > 0$ ), то говорят о **безусловном локальном решении** задачи (1.1'').

Мы будем также пользоваться понятиями выпуклой и вогнутой функций.

Функция  $f$ , определённая на множестве  $\Omega \subseteq R^n$ , называется **выпуклой в точке  $X^*$**  (по отношению к  $\Omega$ ), если из условий

$$X^* \in \Omega, \quad 0 < \lambda < 1, \quad (1-\lambda)X + \lambda X^* \in \Omega$$

следует

$$(1-\lambda)f(X^*) + \lambda f(X) \geq f((1-\lambda)X + \lambda X^*).$$

Если последнее неравенство является строгим для  $X \neq X^*$ , то функция  $f$  называется строго выпуклой в  $X^*$ . В случае, когда знак неравенства изменён на обратный ( вместо  $\geq$  имеет место  $\leq$ ), функция  $f$  называется **вогнутой** в точке  $X^*$ .

Функция  $f$  **выпукла** на множестве  $\Omega$ , если она выпукла в каждой точке  $X^* \in \Omega$ .

## 1.2. Особенности задач нелинейного программирования и их геометрическая интерпретация

Решение задачи (1.1) сопряжено с определёнными трудностями, порождаемыми нелинейностью функций  $f$ ,  $g$  и  $h$ .

Чтобы уяснить, в чём заключаются эти трудности, сопоставим особенности задач линейного и нелинейного программирования.

1. Ограничения в задачах линейного программирования представляют собой линейные неравенства (и равенства), чем определяется тот факт, что область допустимых решений – это выпуклое множество с конечным циклом угловых точек. В задачах нелинейного программирования множество допустимых решений может быть невыпуклым, несвязным и иметь бесконечное число крайних точек. Геометрически это иллюстрируется на рис. 1.1 и 1.2. На рис. 1.1 штриховкой изображён пример области допустимых решений задачи линейного программирования, а на рис. 1.2 – пример области допустимых решений для задачи нелинейного программирования.

2. Минимальное значение линейной целевой функции достигается в одной из угловых точек области допустимых значений, в то время как при нелинейной целевой функции минимум может достигаться и внутри области допустимых решений.

3. В задачах линейного программирования локальный минимум целевой функции всегда совпадает с глобальным решением задачи. В задачах же не-

линейного программирования нелинейная целевая функция может иметь несколько локальных решений.

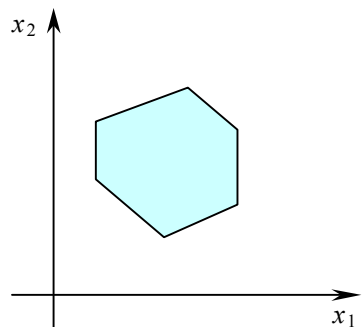


Рис. 1.1. Вид области допустимых решений в задачах линейного программирования

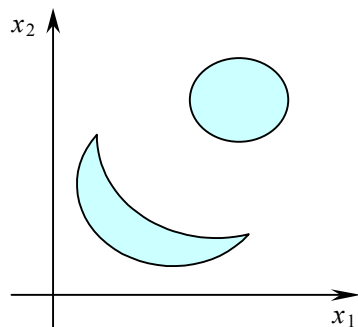


Рис. 1.2. Вид области допустимых решений в задачах нелинейного программирования

Рассмотрим пример задачи нелинейного программирования.

**Пример 1.1.** Найти минимум  $f = x_1 - x_2^2$  при ограничениях

$$x_1^2 - x_2^2 \geq 4; \quad x_1^2 - x_2^2 \leq 9; \quad x_1 - x_2 \leq 2,5; \quad x_1 > 0; \quad x_2 \geq 0.$$

Здесь область допустимых решений распадается на две части. На рис. 1.3 она заштрихована. Линии уровня целевой функции изображены пунктиром. Очевидно, что минимум  $f$  достигается в точке А (0; 2,5) и равен  $\min f = -6,25$ . Максимальное значение функция  $f$  принимает в точке В(2,5; 0) и  $\max f = 2,5$ .

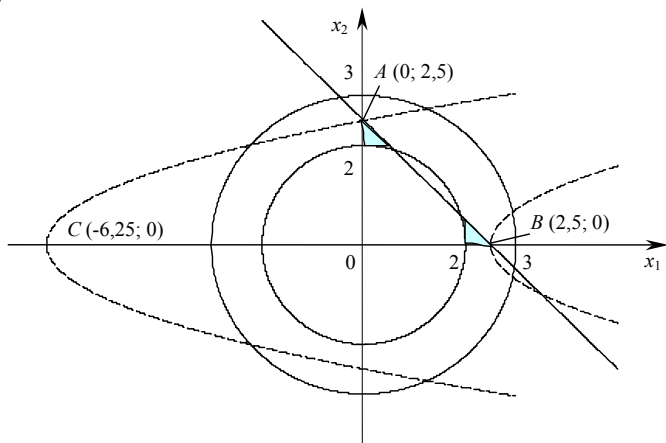
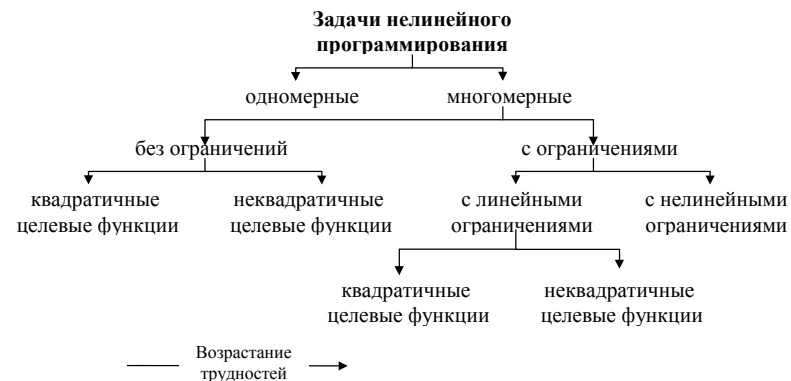


Рис. 1.3. Геометрическое представление задачи минимизации функции

### 1.3. Классификация задач и методов нелинейной оптимизации

Трудности, связанные с решением задач нелинейной оптимизации, зависят от вида функций  $f$ ,  $g$  и  $h$ , которые могут быть линейными, квадратичными или более сложной структуры и зависеть от одной, двух и более переменных. Условно можно выделить следующие классы задач нелинейного программирования:



Степень сложности задач в приведённой классификации очень разная. Самыми простыми для исследователя являются одномерные задачи нелинейной оптимизации и наиболее сложными – многомерные задачи с нелинейными ограничениями. Такое разнообразие классов задач породило и разнообразие методов решения, которые можно классифицировать, например, так.

1. Аналитические методы, использующие возможности дифференциального и вариационного исчисления.
2. Численные методы.
3. Графические методы, базирующиеся на графическом представлении целевых функций и областей допустимых решений.
4. Методы анализа допустимых решений задачи с целью выбора лучшего из них.

Последующие пункты этого раздела посвящены изложению некоторых методов решения задач нелинейного программирования.

### 1.4. Типовые задачи нелинейного программирования, в которых целевая функция зависит от одной переменной

Рассмотрим следующую задачу. Найти минимум  $f(x)$  при отсутствии ограничений, где  $f$  – нелинейная функция.

Здесь и далее мы будем рассматривать задачи нахождения минимума целевой функции, поскольку задачу нахождения максимума функции всегда можно свести к задаче вычисления минимума функции  $-f$ .

**Задача 1.** Пусть  $f(x)$  имеет вид

$$f = ax^2 + bx + c, \quad a \neq 0, \quad x \in R^1.$$

Известно, что для квадратного трёхчлена  $\min f(x) = c - b^2/4a$ , если  $a > 0$  и  $x^* = -b/2a$ . Очевидно, что при  $a < 0$  функция  $f(x)$  достигает максимума в точке  $x^* = -b/2a$ , причём  $\max f(x) = c - b^2/4a$  (рис. 1.4 и 1.5).

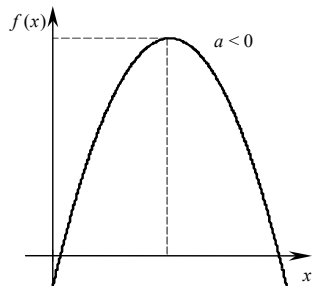


Рис. 1.4. Вид функции с единственным максимумом

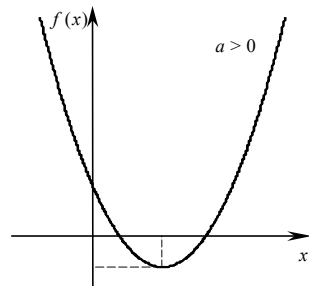


Рис. 1.5. Вид функции с единственным минимумом

**Задача 2.** Пусть целевая функция имеет вид  $f = ax^3 + bx^2 + cx + d$ , где  $a, b, c, d$  – действительные числа и  $a \neq 0$ .

Пусть ограничения, накладываемые на переменную  $x$ , отсутствуют.

При различных значениях коэффициентов  $a, b, c$  и  $d$  могут получиться графики  $f(x)$ , изображённые на рис. 1.6.

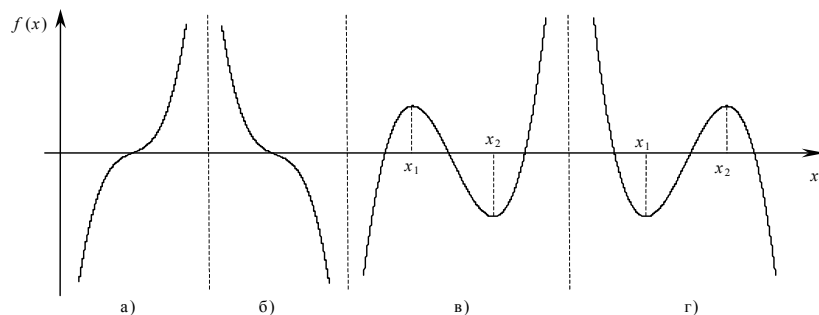


Рис. 1.6. Графики кубической целевой функции

Очевидно, что в случаях а) и б) при отсутствии ограничений экстремумы целевой функции не достигаются и для нас представляют интерес только случаи в) и г).

Вид функции  $f$  ещё позволяет использовать простейший аналитический метод нахождения точки  $x^*$ , в которой достигается экстремум  $f(x)$ . Здесь мы дадим окончательные формулы для вычисления абсцисс  $x_1$  и  $x_2$  (см. рис 1.6, случаи в) и г)), поскольку вывод их очень прост, и интересующийся читатель может найти все выкладки, например, в [1.1].

Итак, если функции третьей степени имеют локальные экстремумы, то их обязательно два (минимум и максимум), и абсциссы  $x_1$  и  $x_2$  вычисляются по формуле

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 3ac}}{3a}.$$

Если  $a > 0$ ,  $\max f(x)$  достигается при  $x^*$ , равном меньшему из чисел  $x_1$  и  $x_2$ , а  $\min f(x)$  – при большем из этих чисел. Если  $a < 0$ , то наоборот,  $\max f(x)$  достигается при  $x^*$ , равном большему из чисел  $x_1$  и  $x_2$ , а  $\min f(x)$  – при меньшем из  $x_1$  и  $x_2$ .

**Задача 3.** Рассмотрим теперь случай, когда целевая функция  $f$  – гладкая и непрерывная функция. Можно воспользоваться классическим методом нахождения экстремумов функций: решить уравнение  $f'(x) = 0$ . В результате мы получим стационарные точки, и, чтобы выделить среди них точки экстремума, необходимо для всех стационарных точек вычислить  $f''(x)$ . Если есть точка  $x^*$ , для которой  $f'(x^*) = 0$  и  $f''(x^*) < 0$ , то в точке  $x^*$  функция  $f(x)$  имеет максимум, если же  $f''(x^*) > 0$ , то функция  $f(x)$  имеет в точке  $x^*$  минимум. Если получают, что  $f''(x^*) = 0$ , то для решения вопроса о существовании экстремума в точке  $x^*$  необходимо рассматривать чётные производные более высоких порядков.

**Задача 4.** В более общем случае, когда целевая функция  $f(x)$  – произвольная нелинейная, следует воспользоваться численными методами нахождения

минимума (максимума) функции. Из многочисленных численных методов рассмотрим методы дихотомии и золотого сечения. Эти методы предполагают, что функция  $f(x)$  – унимодальная. Если на самом деле это требование не выполняется, то после построения графика целевой функции  $f$  можно выделить отрезки (рис. 1.7), на каждом из которых существует только один минимум (максимум)  $f(x)$  и затем применить численный метод для нахождения экстремума  $f(x)$  на выбранном отрезке.

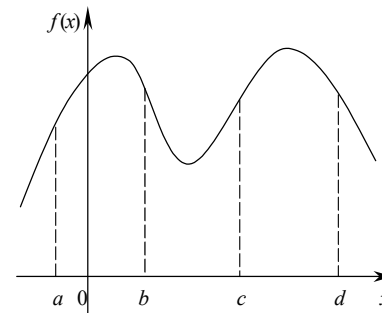


Рис. 1.7. График нелинейной целевой функции с локальными экстремумами

**Метод дихотомии.** Пусть целевая функция  $f(x)$  имеет на отрезке  $[a, b]$  минимум в точке  $x^*$  (рис. 1.8). Вычисление абсциссы  $x^*$  точки минимума  $f(x)$  фактически сводится к вычислению отрезка  $[a_n, b_n]$ , длина которого  $|b_n - a_n| < \varepsilon$ ,  $\varepsilon$  – заданная точность вычисления и  $x \in [a_n, b_n]$ .

Сущность метода дихотомии заключается в следующем:

1) полагаем  $x_1 = a$  и  $x_5 = b$ . Делим отрезок  $[a, b]$  точками  $x_2, x_3$  и  $x_4$  на четыре равные части;

2) вычисляем значения функции  $f(x)$ :  $f(x_1), f(x_2), f(x_3), f(x_4), f(x_5)$ ;

3) среди чисел  $f(x_i)$  находим минимальные. Пусть, например,  $f(x_k) = \min \{f(x_i), i = \overline{1, 5}\}$ .

В силу унимодальности  $f(x)$  на отрезке  $[a, b]$  точка  $x^*$  находится на отрезке  $[x_{k-1}, x_{k+1}]$ , длина которого по построению вдвое меньше длины отрезка  $[a, b]$ .

Если всё-таки  $|x_{k+1} - x_{k-1}| > \varepsilon$ , то полагаем  $a = x_{k-1}$ ,  $b = x_{k+1}$  и вновь выполняем все действия, начиная с 1). Если на  $n$ -й итерации мы получим, что локализованный отрезок  $|x_{k+1} - x_{k-1}| \leq \varepsilon$ , то в качестве значения  $x^*$  можно, например, принять  $x^* = \frac{x_{k-1} + x_{k+1}}{2}$ .

**Пример 1.2.** Найти методом дихотомии минимум функции

$$f(x) = \sin x + \sqrt{x}$$

на отрезке  $[3, 6]$ . Точность  $\varepsilon$  принять равной 1.

График функции  $f(x)$  имеет вид, представленный на рис. 1.9.

В соответствии с вышеизложенным алгоритмом выполняем указанные действия.

Первая итерация:

1)  $x_1 = a = 3$ ;  $x_5 = b = 6$ ;  $x_2 = 3,75$ ;  $x_3 = 4,5$ ;  $x_4 = 5,25$ ;

2) вычисляем значения  $f(x_i), i = \overline{1, 5}$ :  
 $f(x_1) = 1,87$ ;  $f(x_2) = 1,36$ ;  $f(x_3) = 1,14$ ;  
 $f(x_4) = 1,43$ ;  $f(x_5) = 2,17$ ;

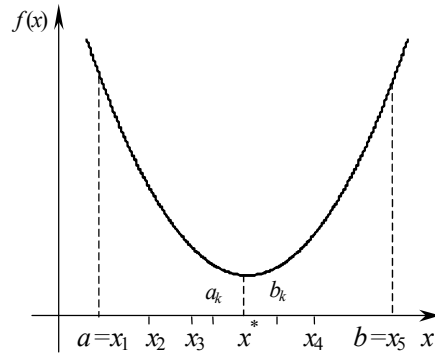


Рис. 1.8. Пример функции с одним минимумом

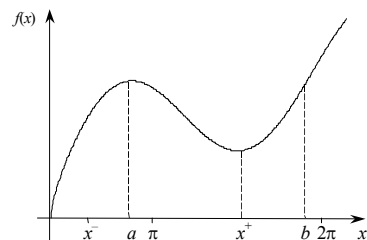


Рис. 1.9. Пример целевой функции с двумя экстремумами

3) выбираем  $\min \{f(x_i), i = \overline{1, 5}\} = f(x_3) = 1,14$ . Следовательно,  $x^* \in [x_2, x_4] = [3,75; 5,25]$ . Длина этого отрезка равна  $5,25 - 3,75 = 1,5$ , что больше  $\varepsilon$ , и мы вынуждены повторить все действия, полагая теперь, что новый отрезок  $[a, b]$ , на котором находится  $x^*$ , равен  $[3,75; 5,25]$ .

Вторая итерация:

1)  $x_1 = a = 3,75$ ,  $x_5 = b = 5,25$ ,  $x_2 = 4,12$ ,  $x_3 = 4,5$ ,  $x_4 = 4,88$ ;

2)  $f(x_1) = 1,36$ ,  $f(x_5) = 1,43$ ,  $f(x_2) = 1,20$ ,  $f(x_3) = 1,14$ ,  $f(x_4) = 1,22$ ;

3)  $\min \{f(x_i), i = \overline{1, 5}\} = f(x_3) = 1,14$ . Выбираем отрезок  $[x_2, x_4] = [4,12; 4,88]$ .

Длина этого отрезка равна  $0,76$ , т. е.  $|x_4 - x_2| < \varepsilon$ . Следовательно, можно сказать, что  $x^* = \frac{4,88 + 4,12}{2} = 4,5$ , а  $f(x^*) \approx 1,1438$ . Безусловно,  $\varepsilon = 1$  обеспечивает нам очень грубое решение задачи. На практике, как правило, задают  $\varepsilon$  малым, например, равным  $0,000001$ . Но наша цель – продемонстрировать практическое применение метода дихотомии для нахождения минимума целевой функции – достигнута.

**Метод золотого сечения.** Прежде всего напомним, что если  $d$  – длина отрезка, то золотое сечение отрезка определяется пропорцией  $\frac{d}{x} = \frac{x}{d-x}$ , т. е.  $x$

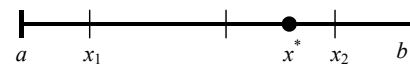


Рис. 1.10. Пример выбора точки деления отрезка по методу золотого сечения

находится из квадратного уравнения  $x^2 + dx - d^2 = 0$ , и следовательно,  $x_1 \approx 0,38d$  и  $x_2 \approx 0,62d$  (рис. 1.10).

Пусть по-прежнему целевая функция  $f(x)$  имеет на отрезке  $[a, b]$  единственный минимум в точке  $x^*$ . Абсциссу точки  $x^*$  мы и стремимся найти с заданной точностью  $\varepsilon = 1$ . Метод золотого сечения предписывает выполнение следующих действий:

1) вычисляем  $x_1 = a + 0,38(b-a)$  и  $x_2 = a + 0,62(b-a)$ .

2) вычисляем  $f(x_1)$  и  $f(x_2)$ . При этом могут иметь место следующие случаи:

а)  $x_2 > x_1$  и  $f(x_1) > f(x_2)$ . В этом случае  $x^* \in [x_1, b]$ ;

б)  $x_2 > x_1$  и  $f(x_1) < f(x_2)$ . В этом случае  $x^* \in [a, x_2]$ ;

в)  $x_2 > x_1$  и  $f(x_1) = f(x_2)$ . В этом случае  $x^* \in [x_1, x_2]$ ;

3) выбираем отрезок, которому принадлежит точка  $x^*$ . Назовём его  $[a_1, b_1]$ . Длина этого отрезка меньше длины исходного отрезка  $[a, b]$ , но если только  $|b_1 - a_1| \leq \varepsilon$ , мы можем считать что с точностью  $\varepsilon$   $x^* = \frac{b_1 + a_1}{2}$ .

В противном случае мы повторяем действия 1)–3) до тех пор, пока на  $n$ -м шаге не будет локализован отрезок  $[a_n, b_n]$ , для которого будет выполне-

но требование  $|b_n - a_n| \leq \varepsilon$ .

Проиллюстрируем применение метода золотого сечения для нахождения минимума целевой функции из примера 1.2.

**Пример 1.3.** Найти методом золотого сечения минимум функции  $f(x) = \sin x + \sqrt{x}$  на отрезке  $[3; 6]$ . Точность  $\varepsilon$  принять равной 1.

Первая итерация:

1) вычислим абсциссы точек золотого сечения  $x_1 = 3 - 0,38(6 - 3) = 4,15$  и  $x_2 = 6 - 0,38(6 - 3) = 4,86$ ;

2) вычислим  $f(x_1) = 1,19$ ,  $f(x_2) = 1,21$ . Очевидно, мы имеем случай б), т.е. при  $x_2 > x_1$ ,  $f(x_1) < f(x_2)$ . Следовательно,  $x^* \in [3; 4,86]$ .

3) длина отрезка, содержащего  $x^*$ , равна 1,86, что больше  $\varepsilon$ , поэтому мы вынуждены выполнить ещё одну итерацию, принимая отрезок  $[a, b_1] = [3; 4,86]$ .

Вторая итерация:

1)  $x_1 = 3 + 0,38 \cdot 1,86 = 3,7$ ,  $x_2 = 4,86 - 0,38 \cdot 1,86 = 4,15$ ;

2)  $f(x_1) = 1,39$ ,  $f(x_2) = 1,19$ . Поскольку  $f(x_1) > f(x_2)$ , то  $x^* \in [3,7; 4,86]$ . Длина этого отрезка равна  $4,86 - 3,7 > \varepsilon$ , поэтому выполним третью итерацию.

Третья итерация:

1)  $x_1 = 3,7 + 0,38 \cdot 1,16 = 4,15$ ,  $x_2 = 4,86 - 0,38 \cdot 1,16 = 4,42$ ;

2)  $f(x_1) = 1,19$ ,  $f(x_2) = 1,44$ ,  $f(x_1) < f(x_2)$ , следовательно,  $x^* \in [4,15; 4,86]$ . Так как длина последнего отрезка равна  $0,71 < \varepsilon$ , мы можем считать, что с точностью  $\varepsilon$   $x^* = \frac{4,15 + 4,86}{2} = 4,505$ , а значит,  $f(x^*) = 1,1439$ .

## 1.5. Решение задач нелинейного программирования методом неопределённых множителей Лагранжа

Пусть задача нелинейного программирования формулируется следующим образом:

требуется найти минимум  $f(X)$  (1.3)

при ограничениях в виде равенств  $h_j(X) = b_j$ ,  $j = \overline{1, m}$ . (1.4)

Будем считать, что функции  $f$  и  $h_j$ ,  $j = \overline{1, m}$  обладают непрерывными первыми частными производными по аргументам  $x_1, x_2, \dots, x_n$ . При высказанных предположениях для решения задачи (1.3), (1.4) можно использовать метод *множителей Лагранжа*.

Составим вспомогательную функцию:

$$L(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m) = f(x_1, x_2, \dots, x_n) + \sum_{j=1}^m \lambda_j (b_j - h_j(x_1, x_2, \dots, x_n)),$$

которая называется *функцией Лагранжа*, а множители  $\lambda_j$ ,  $j = \overline{1, m}$  называются *неординарными* множителями Лагранжа. Обозначим  $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ , тогда можно записать функцию Лагранжа в более компактном виде:

$$L(X, \Lambda) = f(X) + \sum_{j=1}^m \lambda_j (b_j - h_j(X)). \quad (1.5)$$

Доказано, что всякое решение  $X^*$  задачи (1.3), (1.4) необходимо удовлетворяет системе уравнений

$$\begin{cases} \frac{\partial L(X, \Lambda)}{\partial x_i} = \frac{\partial f(X)}{\partial x_i} + \sum_{j=1}^m \lambda_j \frac{\partial h_j(X)}{\partial x_i} = 0, & i = \overline{1, n}; \end{cases} \quad (1.6)$$

$$\begin{cases} \frac{\partial L(X, \Lambda)}{\partial \lambda_j} = b_j - h_j(X) = 0, & j = \overline{1, m}. \end{cases} \quad (1.7)$$

Таким образом, идея метода множителей Лагранжа заключается в том, что исходная задача (1.3), (1.4) отыскания условного экстремума целевой функции  $f$  сводится к задаче отыскания безусловного экстремума функции Лагранжа  $L(X, \Lambda)$ , которая решается классическим методом приравнивания к нулю частных производных по  $x_i$ ,  $i = \overline{1, n}$  и  $\lambda_j$ ,  $j = \overline{1, m}$ .

Проблемы при применении метода Лагранжа существуют и вызваны тем, что, во-первых, для отыскания точек экстремума функции  $L(X, \Lambda)$  требуется

проводить анализ производных  $\frac{\partial^2 L(X, \Lambda)}{\partial x_i^2}$  и  $\frac{\partial^2 L(X, \Lambda)}{\partial \lambda_j^2}$ , и, во-вторых,

условия (1.6), (1.7) не являются достаточными для существования решений  $X^*$  задачи (1.3), (1.4), т.е. необязательно, чтобы любое решение системы (1.6), (1.7) было и решением задачи (1.3), (1.4). Однако простота составления функции Лагранжа, возможность обобщения задачи на случай ограничений-неравенств обусловила фундаментальную роль метода множителей Лагранжа в оптимизационной науке.

Итак, если мы хотим найти решение задачи (1.3), (1.4) методом Лагранжа, мы должны выполнить следующие действия:

1) составить функцию Лагранжа (1.5);

2) вычислить частные производные  $\frac{\partial L(X, \Lambda)}{\partial x_i}$ ,  $i = \overline{1, n}$  и  $\frac{\partial L(X, \Lambda)}{\partial \lambda_j}$ ,

$j = \overline{1, m}$  и приравнять их нулю;



3) решить систему уравнений (1.6), (1.7) любым известным нам методом, получив тем самым стационарные точки  $(X, \Lambda)$ ;

4) исследовать функцию (1.3) в полученных стационарных точках.

**Пример 1.4.** Найти прямоугольник наименьшего периметра, ограничивающий заданную площадь.

Математическая модель задачи имеет следующий вид.

Найти  $\min f(x_1, x_2) = 2x_1 + 2x_2$  при ограничениях  $x_1 x_2 = s$ ;  $x_1 \geq 0$ ;  $x_2 \geq 0$ .

Запишем функцию Лагранжа  $L(x_1, x_2, \lambda) = 2x_1 + 2x_2 + \lambda(s - x_1 x_2)$  и составим систему уравнений

$$\begin{cases} \frac{\partial L}{\partial x_1} = 2 - \lambda x_2 = 0; \\ \frac{\partial L}{\partial x_2} = 2 - \lambda x_1 = 0; \\ \frac{\partial L}{\partial \lambda} = s - x_1 x_2 = 0; \end{cases} \Rightarrow \begin{cases} x_1 = \frac{2}{\lambda}; \\ x_2 = \frac{2}{\lambda}; \\ \lambda^2 s = 4. \end{cases}$$

Решение этой системы (рис 1.11):

$$x_1^* = \sqrt{s}; \quad x_2^* = \sqrt{s}; \quad \lambda = \frac{2}{\sqrt{s}}.$$

На рис. 1.11 пунктирной линией изображены уровни функции цели  $f(x_1, x_2) = 2x_1 + 2x_2$ . На линии  $x_1 x_2 = s$  лежат точки, удовлетворяющие ограничениям задачи, и точка  $M(x_1^* = \sqrt{s}; x_2^* = \sqrt{s})$  является решением задачи.

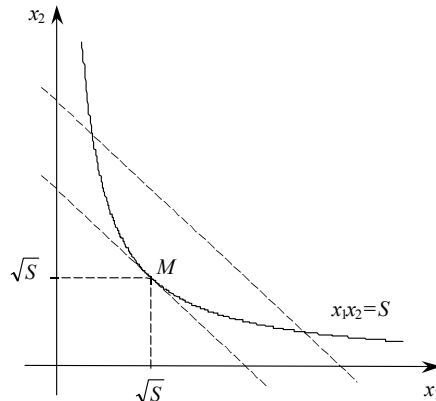


Рис. 1.11. Пример нахождения решения системы уравнений

### 1.6. Использование градиентных методов для решения задач нелинейного программирования

Чтобы понять идею градиентных методов решений задач нелинейного программирования, следует вспомнить точное определение градиента и его свойства.

В математике **градиент** – это вектор, определяющий направление и скорость наискорейшего роста функции  $f$  от нескольких переменных. В декартовых прямоугольных координатах:

$$\nabla f(X) = \text{grad } f(X) = \left( \frac{\partial f(X)}{\partial x_1}, \frac{\partial f(X)}{\partial x_2}, \dots, \frac{\partial f(X)}{\partial x_n} \right).$$

Вектор  $[-\text{grad } f(X)]$  называется **антиградиентом** функции  $f(X)$ . Он ука-

зывает направление наискорейшего убывания функции  $f(X)$ .

Для простоты рассмотрим функцию двух переменных  $f(x_1, x_2)$ . На рис. 1.12 изображены две линии уровня  $f(x_1, x_2) = C$  и  $f(x_1, x_2) = C + \Delta C$ . (Напомним, что линия уровня функции  $f$  – это множество точек, в которых данная функция  $f(x_1, x_2)$  принимает одно и то же фиксированное значение.)

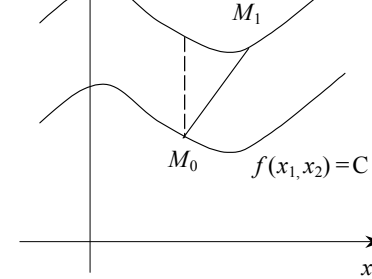


Рис. 1.12. Линии уровня функции двух переменных

Рассмотрим какую-либо точку  $M_0(x_1^0, x_2^0)$  на линии  $f(x_1, x_2) = C$  и точку  $M_1(x_1^1, x_2^1)$  на линии  $f(x_1, x_2) = C + \Delta C$ . Приращение функции  $\Delta f$  при переходе из точки  $M_0$  в точку  $M_1$  равно  $\Delta f = \Delta C$ , и это приращение не зависит от положения точки  $M_1$  на линии  $f(x_1, x_2) = C + \Delta C$ . Отношение приращения функции к длине отрезка  $M_0M_1$  называется средней скоростью изменения функции  $f(x_1, x_2)$  на отрезке  $M_0M_1$ . Очевидно, что эта средняя скорость  $\Delta f / |M_0M_1|$  зависит от длины  $M_0M_1$  или,

иначе говоря, от положения точки  $M_1$  на линии  $f(x_1, x_2) = C + \Delta C$ . И также очевидно, что скорость  $\Delta f / |M_0M_1|$  достигает наибольшего значения там, где длина отрезка  $M_0M_1$  будет наименьшей. Известно, что длина отрезка  $M_0M_1$  будет наименьшей в том случае, когда направление  $M_0M_1$  перпендикулярно касательной к линии  $f(x_1, x_2) = C$  в точке  $M_0$  при  $\Delta C \rightarrow 0$ . Теперь, если мы рассмотрим  $\lim_{|M_0M_1| \rightarrow 0} \frac{\Delta f}{|M_0M_1|}$  и вспомним определение производной и её

физический смысл, то неизбежно придём к утверждению, что частные производные  $\frac{\partial f(M_0)}{\partial x_1}$  и  $\frac{\partial f(M_0)}{\partial x_2}$  определяют наибольшие скорости изменения  $f(x_1, x_2) = C$  в точке  $M_0$  по координатным осям.

Теперь перейдём к рассмотрению идеи градиентных методов.

Пусть, например, требуется найти экстремум функции  $f(X)$ .

1. Выберем начальную точку  $M_0$ , принадлежащую области определения функции (надо отметить, что это непростая задача, не имеющая формальных методов решения).

2. Найдём направления градиента функции в точке  $M_0$ :

$$\left( \frac{\partial f(M_0)}{\partial x_1}, \frac{\partial f(M_0)}{\partial x_2}, \dots, \frac{\partial f(M_0)}{\partial x_n} \right).$$

3. Выполним перемещение из точки  $M_0$  на величину, равную некоторому шагу  $\alpha$ , в точку  $M_1$ . Известно, что параметрическое уравнение прямой, проходящей через заданную точку  $M_0$  в заданном направлении  $\text{grad} f(M_0)$ , имеет вид

$$x_j(M_1) = x_j(M_0) + \alpha \frac{\partial f(M_0)}{\partial x_j}, \quad j = \overline{1, n}.$$

Если мы ищем минимум функции, то должны перемещаться в направлении антиградиента, и в этом случае мы принимаем  $\alpha < 0$ ; если ищется максимум функции, то надо перемещаться в направлении градиента, т. е. берётся  $\alpha > 0$ .

Признаком достижения минимума (максимума) функции служит обращение в нуль градиента  $\text{grad} f = 0$ . Но на практике, как правило, ограничиваются приближённым решением и критерием достижения требуемого результата является выполнение условия

$$\sum_{j=1}^n \left( \frac{\partial f(X)}{\partial x_j} \right)^2 \leq \varepsilon, \quad \text{где } \varepsilon > 0 \text{ – наперёд}$$

заданное число.

Существуют и другие критерии. Например, после каждой итерации можно сравнивать достигнутое значение целевой функции с её значением в предыдущей точке. Если выполняется  $|f(M_n) - f(M_{n-1})| \leq \varepsilon$ , где  $\varepsilon > 0$  – наперёд заданное число, то поиск экстремума функции прекращается.

Разновидности градиентных методов отличаются лишь выбором шага  $\alpha$  вдоль градиентного направления. Выбор длины шага  $\alpha$  – достаточно сложная и важная задача. Если длина шага большая, то можно перескочить экстремум и появляется опасность заикливания. Если длина шага малая, то процесс поиска экстремума может оказаться очень медленным. Эти трудности с выбором длины шага преодолеваются в методе наискорейшего спуска, когда в каждой итерации вычисляется оптимальное значение  $\alpha$ .

Градиентные методы можно применять при нахождении экстремумов нелинейной целевой функции без ограничений и с ограничениями.

Теперь рассмотрим пример.

**Пример 1.5.** Фабрика закупает сырьё двух видов  $x_1$  и  $x_2$  тонн. Расходы на хранение сырья на складе определяются по функции

$$f(x_1, x_2) = x_1^2 - 7x_1 + x_2^2 - 4x_2 - x_1x_2 + 35 \text{ (тыс. руб.)}.$$

Определить объёмы сырья, расходы на хранение которого будут мини-

мальны. Принять точность вычислений  $\varepsilon = 1$ . В качестве начальной точки  $M_0$  возьмём точку с координатами  $x_1 = 1, x_2 = 1$ , шаг  $\alpha$  примем равным 0,5.

Расходы на хранение сырья при  $x_1 = 1$  и  $x_2 = 1$   $f(M_0) = 25$ .

Вычислим

$$\frac{\partial f}{\partial x_1} = 2x_1 - x_2 - 2; \quad \frac{\partial f}{\partial x_2} = 2x_2 - x_1 - 4; \quad \frac{\partial f(M_0)}{\partial x_1} = -6; \quad \frac{\partial f(M_0)}{\partial x_2} = -3.$$

Переходим к точке  $M_1(x_1^1, x_2^1)$ :

$$x_1^1 = 1 - 0,5 \cdot (-0,6) = 4; \quad x_2^1 = 1 - 0,5 \cdot (-0,3) = 2,5;$$

$$\frac{\partial f(M_1)}{\partial x_1} = -1,5; \quad \frac{\partial f(M_1)}{\partial x_2} = -3; \quad f(M_1) = 9,25.$$

Разность  $|f(M_1) - f(M_0)| = 15,75 > \varepsilon$ , следовательно, выполняем вторую итерацию и переходим к точке  $M_1(x_1^2, x_2^2)$ :

$$x_1^2 = 4 - 0,5 \cdot (-1,5) = 4,75; \quad x_2^2 = 2,5 - 0,5 \cdot (-0,3) = 4;$$

$$\frac{\partial f(M_2)}{\partial x_1} = -1,5; \quad \frac{\partial f(M_2)}{\partial x_2} = -0,75; \quad f(M_2) = 5,31;$$

$$|f(M_2) - f(M_1)| = 3,94 > \varepsilon.$$

Мы должны выполнить третью итерацию:

$$x_1^3 = 4,75 - 0,5 \cdot (-1,5) = 5,5; \quad x_2^3 = 4 - 0,5 \cdot (-0,75) = 4,4;$$

$$\frac{\partial f(M_3)}{\partial x_1} = -0,38; \quad \frac{\partial f(M_3)}{\partial x_2} = -4,74; \quad f(M_3) = 4,31, \quad |f(M_3) - f(M_2)| = 1 = \varepsilon.$$

Учитывая, что требуемая точность результата достигнута, процесс вычисления можно прекратить. Минимальные издержки при объёмах сырья  $x_1 = 5,5$ ; и  $x_2 = 4,4$  составят 4,31 ден. ед.

## 2. ВЕРОЯТНОСТНЫЕ МОДЕЛИ СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМ

### 2.1. Особенности использования вероятностных моделей

Когда параметры элементов СТС имеют стохастическую природу и когда многие ограничения, вводимые при формализации СТС, не выполнимы, оказывается, что аналитические модели трудно применять для исследования динамики функционирования СТС. В этих случаях система рассматривается как «чёрный ящик», на вход которого задаётся вектор входных воздействий  $X=(x_1, \dots, x_k)$ , на выходе которого фиксируется вектор откликов  $Y=(y_1, \dots, y_n)$ . И обычно ставится задача найти вид зависимости  $Y=F(X, \alpha_j)$ , где  $\alpha_j$  – набор случайно изменяющихся значений неуправляемых (но контролируемых) параметров;  $j = \overline{1, J}$ , где  $J$  – число неуправляемых параметров СТС. Предметом исследований является изучение случайных процессов, возникающих и протекающих в «чёрном ящике». Как правило, о поведении СТС известен (в лучшем случае) только алгоритм. Проще всего решаются подобные задачи, когда алгоритм поведения СТС представлен в виде графа, где вершинами будут некоторые алгоритмические действия, изменяющие значения компонентов вектора откликов  $y_i$ , а дугами оказываются возможные направления выбора будущих алгоритмических действий. Причём выбор этот может осуществляться либо по случайному, либо по детерминированному, либо по смешанному закону распределения вероятностей. Как правило, этот граф представляет собой дерево вариантов реализации случайного процесса. В ряде случаев узлами этого дерева могут быть процедуры, определяющие (измеряющие) значения неуправляемых параметров СТС ( $\alpha_j$ ). В зависимости от значений  $\alpha_j$  путь вычислений в ходе математического моделирования поведения СТС будет проходить по разным ветвям, и отсюда значения вектора откликов  $Y_j$  как суперпозиции могут вычисляться по разным ветвям этого дерева алгоритмов реализации случайных процессов в СТС.

Второй особенностью вероятностных моделей СТС является зависимость результата моделирования от исходных данных. В любой данный момент времени предсказательные возможности ММ вероятностных процессов гораздо ниже, чем у аналитических детерминированных ММ. Любое моделирование случайного процесса в СТС позволяет дать ответ на вопрос, что же произойдёт при конкретных исходных начальных условиях моделирования поведения СТС. И чтобы узнать больше, необходим повторный эксперимент. В ходе таких экспериментов исследователь накапливает статистику

поведения системы.

Третьей предпосылкой и особенностью построения, использования и анализа результатов моделирования является необходимость накопления статистики результатов моделирования и последующего статистического анализа данных моделирования. Необходимость использования методов прикладного статистического анализа данных требует дополнительного расхода ресурсов на планирование экспериментов, постановку серии экспериментов и обработку статистики вариантов моделирования для получения оценок вектора откликов  $Y_i, i = \overline{1, J}$  и неуправляемых параметров  $\alpha_{j=\overline{0, J}}$ .

Четвёртой особенностью применения вероятностных моделей является необходимость использования в расчётах механизма получения случайных и «псевдослучайных» чисел. Для этой цели в вероятностных моделях реализуется стандартный механизм, позволяющий моделировать по вектору вероятностей ситуации выбора направления реализации случайного процесса.

### 2.2. Использование метода статистических испытаний при исследовании СТС

Когда реальные СТС сложны и имеют вероятностную природу, аналитическими методами исследовать динамику их поведения зачастую настолько сложно и результаты вариантов моделирования зачастую противоречат друг другу, что исследователю проще использовать для этой цели универсальный метод статистических испытаний (метод «Монте-Карло»). Вместо того чтобы описывать СТС с помощью семейства аналитических моделей СТС, производится  $N$  «розыгрышей» этого случайного явления (заранее заданным числом раз  $N$ ). При этом фиксируются исходная информация (значения компонентов  $X$ ) и результаты моделирования (значения компонентов  $Y$ ). На вход «чёрного ящика», алгоритм поведения которого должен быть известен заранее, подаётся  $N$  раз одно и то же входное воздействие  $X$ , и в ходе моделирования фиксируется вектор откликов  $Y$ . Так же, как в реальности, конкретная  $j$ -я реализация случайного процесса (из-за случайного характера алгоритма реализации случайного процесса) в СТС всякий раз будет складываться по-разному. Таким образом, проведя  $N$  экспериментов с ММ СТС при одних и тех же значениях вектора  $X$  из-за случайного характера алгоритма и неуправляемых параметров  $\{\alpha_j\}$  получают выборку значений  $\{Y_j\}, j = \overline{1, N}$ . Усредняя выборку значений  $\{Y_j\}$ , находят вектор математических ожиданий компонент вектора  $\bar{Y} = (\bar{Y}_{x_1}, \dots, \bar{Y}_{x_j})$  и дисперсию  $D = S^2_y$  из выборки  $\{Y_j\}$ . Число экспериментов  $N$  определяется из необходимой точности оценки  $Y$ . Задавшись точностью вычислений  $\epsilon$ , по таблицам нормального распределения для доверительной вероятности  $\beta = 1 - \epsilon$  находят требуемое

число экспериментов  $N$ , обеспечивающее оценку среднего  $\bar{Y}$  с вероятностью ошибки  $\varepsilon$ .

При оценивании среднего значения компонентов вектора  $\bar{Y}$  наиболее часто встречаются следующие случаи:

а)  $Y_i$  имеют нормальное распределение, но объём выборки мал ( $N \leq 30$ ). Тогда составляется  $t$ -статистика [2]

$$t = \sqrt{N-1}(\bar{Y} - \mu)S_Y, \quad (2.1)$$

где  $S_Y^2$  и  $\mu$  – оценка дисперсии и истинное математическое ожидание компонент вектора откликов  $Y$ ;  $\bar{Y}$  – среднее значение компонент вектора откликов. Как известно,  $t$ -статистика имеет распределение Стьюдента [1]. Поэтому по соответствующим таблицам при  $N-1$  степенях свободы и заданном уровне значимости  $\alpha$  оценивают доверительные интервалы для математических ожиданий вектора откликов:

$$\bar{Y} \pm d = \bar{Y} \pm t_{kp} S_Y / \sqrt{N-1}, \quad (2.2)$$

откуда требуемый объём выборки для оценки отклика с точностью  $d$  будет

$$N \geq \left( \frac{t_{kp} - S_Y}{d} \right)^2 + 1; \quad (2.3)$$

б)  $Y_i$  имеют нормальное распределение; истинное математическое ожидание  $\mu$  известно, объём выборки велик ( $N \geq 30$ ). В этом случае используется двусторонняя нормальная статистика  $Z$  при заданном уровне значимости  $\alpha$ . Доверительный интервал можно определить в виде соотношения

$$\bar{Y} \pm d = \bar{Y} \pm Z_{\alpha/2} \frac{S_X}{\sqrt{N}}, \quad (2.4)$$

и требуемый объём выборки

$$N \geq \left( \frac{S_Y - Z_{\alpha/2}}{d} \right)^2, \quad (2.5)$$

где  $Z_{\alpha/2}$  – значение нормированного нормального распределения при уровне значимости  $\alpha$ ;  $d$  – допустимая величина ошибки для оценки средних значений откликов;

в) нормальность распределений компонентов вектора  $Y$  предположить нельзя, но известны истинные значения математического ожидания  $\mu$  и дис-

персии  $\sigma^2$  при большом объёме выборки ( $N \geq 30$ ). В этом случае лучше всего использовать неравенство Чебышёва:

$$P\left\{|\bar{Y} - \mu| \geq K \frac{\sigma}{\sqrt{N}}\right\} \leq \frac{1}{K^2}. \quad (2.6)$$

Из этого неравенства доверительный интервал для оценки среднего значения можно записать в виде

$$\bar{Y} \pm d = \bar{Y} \pm \frac{K\sigma}{\sqrt{N}} = \bar{Y} \pm \frac{\sigma}{\sqrt{N(1-\alpha)}}, \quad (2.7)$$

а требуемый объём выборки отклика

$$N \geq \frac{\sigma}{d^2(1-\alpha)}. \quad (2.8)$$

Обычно задаются  $d \approx \sigma/4$ ,  $\alpha = 0.05$ . Тогда  $N \geq 320$ .

При оценивании дисперсии  $\sigma^2$  задача отыскания её оценки  $S_Y^2$  с достоверностью  $1 - \alpha$  записывается в виде

$$P\{(1-d)\sigma^2 \leq S_Y^2 \leq (1+d)\sigma^2\} = 1 - \alpha. \quad (2.9)$$

Эту задачу можно решать только при больших объёмах выборки  $N$ .

В этих случаях удобнее всего использовать  $\chi^2$ -статистику:

$$\chi^2 = \frac{(N-1)S^2}{\sigma^2} \text{ с } (N-1) \text{ степенями свободы.} \quad (2.10)$$

Из-за большого  $N$  эту статистику можно аппроксимировать нормальным распределением. Поэтому формула для расчёта требуемого объёма выборки отклика имеет вид

$$N \geq 1 + 2 \left( \frac{Z_{\alpha/2}}{d} \right)^2, \quad (2.11)$$

где  $Z_{\alpha/2}$  – значение нормально распределённой величины при уровне значимости  $\alpha$ ;  $d$  – доверительный интервал оценки, чаще всего  $d = \frac{\sigma_3}{4}$ ;  $\sigma_3^2$  – эмпирическая оценка дисперсии отклика.

Итак, установив число опытов с «чёрным ящиком», приступают к серии из  $N$  экспериментов. Иногда такой «силовой приём» оказывается проще построения сложной аналитической модели и исследования на ней зависимости  $Y = F(X, \alpha)$ . Поэтому в реальных СТС метод Монте-Карло применяется в

следующих случаях:

1. Моделируются сложные операции, в которых присутствует большое число сложно взаимодействующих случайных неуправляемых воздействий  $\alpha_i$  на СТС.

2. Требуется проверить применимость более простых аналитических методов в конкретных условиях.

3. При нахождении конкретных значений параметров эмпирических формул или при их конструировании по данным моделирования.

Перечисление возможных случаев применения метода Монте-Карло можно было бы продолжить. Основной операцией, из совокупности которых складывается процесс моделирования по методу Монте-Карло, является некая  $j$ -я реализация случайного процесса  $F(X, \alpha_i)$ . Она представляет собой как бы один экземпляр реализации алгоритмов развития СТС в дереве вариантов его функционирования. Как правило, она складывается из последовательности, состоящей из двух типов составляющих. Первый тип составляющих представляет собой вычислительные процедуры  $F$ , а вторая составляющая означает розыгрыш значений  $\alpha_i$  с помощью специального разработанного алгоритма, носящего название «бросание жребия». Каждый раз, когда в ход моделируемого явления вмешивается случайность, её влияние учитывается не расчётом, а «жребием». Далее расчётами  $F$  уже учитывается влияние исхода «жребия» на дальнейшее течение процесса. Учитываются при этом и условия розыгрыша следующего «жребия».

### 2.3. Виды и алгоритмы реализации единичных жребиев

Условимся называть «единичным жребием» любой опыт со случайным исходом, в котором разыгрывается:

1) произошло или не произошло событие  $A$ , имеющее вероятность появления  $P$ ;

2) какое из возможных событий  $A_1, A_2, \dots, A_m$  произошло, если заданы вероятности их появления  $P_1, P_2, \dots, P_m$ ;

3) какое значение  $x$  приняла случайная величина  $X$ ;

4) какую совокупность значений  $(x_1, x_2, \dots, x_k)$  приняла система случайных чисел  $(X_1, X_2, \dots, X_k)$ .

Существует стандартный механизм, с помощью которого можно осуществить любой из перечисленных видов жребия. Он состоит в формировании случайного числа  $\xi$ , значения которого равновероятны на интервале  $[0, 1]$ , и последующем формульном преобразовании этого случайного числа  $\xi$ .

Рассмотрим алгоритмы моделирования всех четырёх видов жребия с помощью генератора случайных чисел  $\xi$ , равномерно распределённых на интервале  $[0, 1]$ .

1<sup>0</sup>. Моделирование события  $A$ . Моделируется очередное значение случайного числа  $\xi$  путём обращения к генератору случайных чисел ГСЧ. Если  $p > \xi$ , то  $A :=$  «истина» (событие произошло), иначе  $A :=$  «ложь» (событие не произошло).

2<sup>0</sup>. При заданном векторе вероятностей появления  $(P_1, P_2, \dots, P_m)$  моделирование одного из возможных событий  $(A_1, A_2, \dots, A_m)$ . Делят интервал  $[0, 1]$  на  $m$  участков, длины которых соответствуют вероятностям появления событий  $P_i, i = \overline{1, m}$ . Разыгрывается очередное значение величины  $\xi$  путём обращения к ГСЧ. Если  $\xi$  накрывает  $i$ -й участок, то считают, что появилось  $i$ -е событие ( $A_i :=$  «истина», а остальные  $A_j :=$  «ложь»,  $i \neq j, j = \overline{1, m-1}$ ). При этом просмотр участков ведётся от начала, поскольку все  $A_i$  уже упорядочены. Считается также, что все события  $A_i$  независимы.

3<sup>0</sup>. Какое значение  $x$  приняла случайная величина  $X$ . Если  $X$  дискретная величина и известно распределение вероятностей и её значений:

$\{(x_1, P_1), (x_2, P_2), \dots, (x_k, P_k)\}$ , где  $\sum_{i=1}^k P_i = 1$ , формируют эмпирическую функцию

дискретного распределения  $\{(x_1, P_1), (x_2, P_1 + P_2), \dots, (x_k, \sum_{i=1}^k P_i)\}$ . Разыгрывается очередное значение  $\xi$  и проверяется выполнение неравенства

$\sum_{i=1}^l P_i < \xi \leq \sum_{i=1}^{l+1} P_i$ . Если это неравенство выполняется для  $l$ -го дискретного интервала, то считают, что  $x := x_l$ .

Если же  $X$  – непрерывная величина и известна функция плотности вероятностей её распределения  $f(x)$ , формируют функцию распределения вероятностей появления величин  $x$ :

$F(x) = \int_{-\infty}^{+\infty} f(x) dx$ . Разыгрывается очередное

значение величины  $\xi$  путём обращения к ГСЧ. Далее вычисляется обратная функция  $\Psi$  от  $F(x)$  при заданном значении  $x' = \xi$ ; тогда  $x := \Psi(\xi)$  и будет искомым значением случайной величины  $X$ . Предполагается при этом, что кривая  $F(x)$  аналитически задана, поэтому и возможна процедура получения обратной функции  $\Psi$  в точке  $x' = \xi$ . Если же  $F(x)$  и  $\Psi(x)$  трудно представимы аналитически, то её заменяют (аппроксимируют) линейно-кусочной (полигоном частот) или ступенчатой линией (гистограммой). А затем уже поступают аналогично предыдущему случаю для нахождения значений обратной функции в точке  $x' = \xi$ .

4<sup>0</sup>. Какую совокупность значений  $(x_1, x_2, \dots, x_k)$  приняла система случайных чисел  $(X_1, X_2, \dots, X_k)$ . Возможны два случая:  $X_i$  независимы друг от дру-

га;  $X_i$  представляют систему зависимых величин. В первом случае розыгрыш  $(x_1, x_2, \dots, x_k)$  сводится к  $k$  раз повторенному моделированию величин  $X$  по жребью  $3^0$ . Гораздо хуже обстоит дело, когда  $x_1, x_2, \dots, x_k$  зависимы. В этом случае, разыграв одну из переменных, например  $x_1 = \tilde{x}_1$  необходимо разыгрывать вторую величину  $x_2 = \tilde{x}_2$ , но уже согласно условному распределению двух величин и с учётом того, что  $x_1 = \tilde{x}_1$ . Третью величину, например  $x_3$  разыгрывают по условным распределениям уже трёх переменных, а именно:  $x_3 = \tilde{x}_3$  при условии, что  $x_1 = \tilde{x}_1$ ,  $x_2 = \tilde{x}_2$  и т.д.

#### 2.4. Моделирование псевдослучайных величин

В основе реализации четырёх видов единичных жребиев лежит последовательное генерирование равномерно распределённых величин  $\xi_i$ ,  $i=0, 1, 2, \dots$ . Во всех ГСЧ используется процедура формирования равномерно распределённых величин  $\xi_i$  на интервале  $[0, 1]$ . Затем уже производится функциональное преобразование  $\xi_i$  в распределение  $X$  с заданным законом распределения  $F(x)$ .

Исторически возникло два подхода к моделированию величин  $\xi$ : с помощью специальных физических генераторов-приставок к ЭВМ и с помощью специально подобранного алгоритма, реализованного на ЭВМ в виде специальной процедуры. Первый подход позволяет получить действительно случайные величины  $\xi$  на любом интервале их значений. Никогда предыдущее значение  $\xi'$  не повторится вновь. Моделирование  $\xi'$  основано на известных физических явлениях (гамма-излучение, космические лучи, белый шум в электрических цепях и т. д.). с заданными законами распределений. Основными достоинствами таких генераторов являются: действительно случайный характер их появления, неповторяемость и возможность получать одно число за одну специальную команду обращения к генератору-приставке к ЭВМ. При этом они обладают и существенными недостатками: нестабильностью работы, зависимостью от температуры, влажности, погоды, давления. Кроме того, неповторяемость  $\xi'$  затрудняет организацию многовариантных исследований процессов, требуя для этого специальных запоминаний последовательностей  $\{\xi'_j\}$ . Эти недостатки и побудили появление ГСЧ второго типа – программные ГСЧ. Это специальные процедуры, реализованные на ЭВМ, позволяющие моделировать последовательность  $\{\xi''_j\}$ . При этом  $\xi''_j$  случайны на некотором интервале значений  $j$ , называемом периодом аперидичности ( $N_{АП}$ ). Далее, начиная с некоторого  $N_{АП}$ , все  $\xi''_j$  начинают повторяться. Поэтому числа  $\xi''_j$  и назвали псевдослучайными. Достоинства ГСЧ второго типа в том, что они стабильно работают и не боятся перепадов температур, влажности и давления; достигается полная воспроизводимость всей после-

довательности чисел. Главным недостатком, кроме наличия периода повторяемости, является их ресурсоёмкость. На получение одного числа  $\xi''_j$  необходимо выполнение целой программы (процедуры) генератора. Однако несмотря на указанные недостатки, эти процедуры практически вытеснили из употребления физические датчики случайных чисел.

Базовый программный ГСЧ формирует последовательность псевдослучайных величин, равномерно распределённых на интервале  $[0, 1]$ ;  $\xi_1, \xi_2, \dots, \xi_n$ . Практически в силу дискретности представления чисел в ЭВМ можно генерировать  $M=2^k$  различных чисел (здесь  $k$  – разрядность кода ЭВМ, а  $M$  – максимальное число, представленное в ЭВМ). Вероятность появления очередного соседнего числа составляет  $P(\xi_i = x_i) = \frac{1}{M}$ . Обычно моделирование в ГСЧ второго типа ведётся по итеративной формуле, т. е. очередное значение  $x_i^*$  является функцией всех предыдущих значений:

$$x_i^* = \varphi(x_{i-1}^*, \dots, x_0^*). \quad (2.12)$$

Нормировка по модулю  $M$  необходима для перевода значений  $x_i^*$  в интервал  $[0, 1]$ . Реализации  $x_i^*$  получаются внутри алгоритма путём подобранного преобразования  $\varphi$  из ранее имитированных чисел  $x_0^*, x_1^*, \dots, x_{i-1}^*$ . Существует множество процедур, реализующих в ГСЧ различные алгоритмы преобразований  $\varphi$ . Наиболее распространён конгруэнтный метод генерации псевдослучайных чисел, сводящийся к формульному преобразованию:

$$x_i^* = (A x_{i-1}^* + C) \bmod M. \quad (2.13)$$

Работу этого алгоритма представим для случая  $A=2$  и  $C=0$ . т. е.  $x_i^* = 2x_{i-1}^* \bmod M$ .

Графически моделирование по этому алгоритму представлено на рис. 2.1 и 2.2. В первом случае начальное значение  $x_0$  подобрано очень неудачно, что приводит к очень короткому периоду аперидичности ГСЧ ( $N_{АП}=2$ ). Этот ГСЧ позволил получить всего 2 различных значения, и далее началось повторение этой пары чисел. Здесь диагональ играет роль подстановки предыдущего значения  $x$  в качестве  $x_{i-1}$  для расчёта  $x_i$ . Во втором случае  $x'_0$  подобрано гораздо лучше, что позволило существенно увеличить период аперидичности. И, как видно из рис. 2.2, все шесть  $x_i$  действительно отличаются друг от друга и распределены относительно равномерно по всему интервалу значений (от 0 до 1).

Именно переходный режим в ГСЧ и позволяет получать псевдослучайные последовательности чисел  $\xi_i$ .

Кроме рассмотренного, существует ещё один способ, называемый «способом срединных квадратов». Суть этого способа заключается в следующем. Берут произвольное  $m$ -разрядное число  $x_0$ , возводят его в квадрат, и

результат уже имеет  $2m$  разрядов. Из него выбирают средние  $m$  цифр, получая таким способом новое число  $x_1$ . Далее процедура повторяется, где в качестве начального значения используют  $x_1$ . Его возводят в квадрат. Из него выбирают средние  $m$  цифр и получают новое число  $x_2$ .

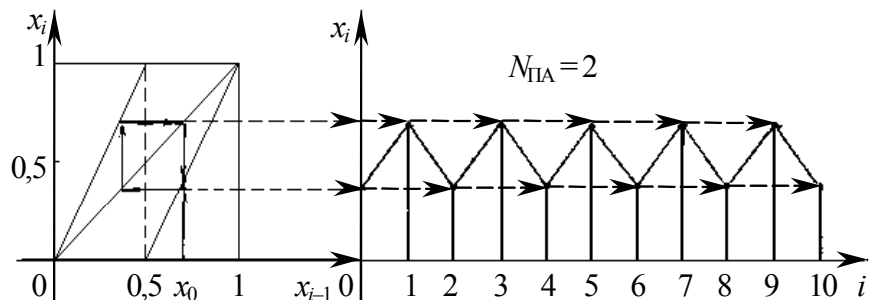


Рис. 2.1. Диаграмма моделирования псевдослучайных последовательностей по конгруэнтному методу когда  $x_i^* = 2x_{i-1}^*$ , ( $C=0$ ), начальное значение  $X_0$

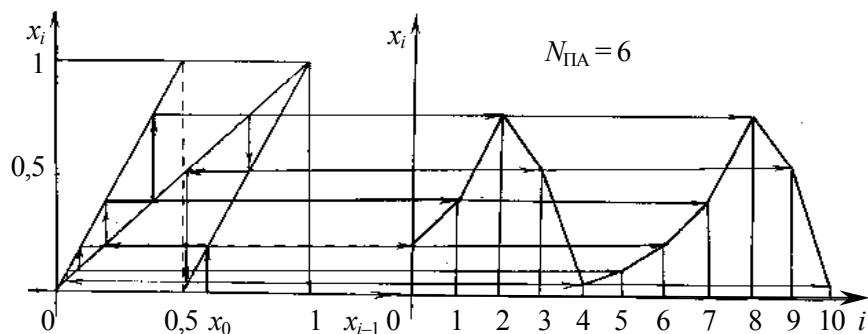


Рис. 2.2. Диаграмма моделирования псевдослучайных последовательностей по конгруэнтному методу ( $x_i^* = 2x_{i-1}^*$ ), начальное значение  $X_0$

Третий способ представляет собой совмещение двух конгруэнтных способов. Он предполагает двойное обращение к ГСЧ, работающему по конгруэнтному способу, а затем формульное преобразование

$$x_{i+1}^* = (x_i - x_{i-1}) \bmod M. \quad (2.14)$$

Эта простейшая процедура увеличивает существенно период периодичности ГСЧ. Наиболее эффективна процедура возмущения основного алгоритма ГСЧ. Суть этой процедуры состоит в следующем. Одновременно ис-

пользуются две процедуры ГСЧ. Основная процедура моделирует последовательность случайных чисел по алгоритму:

$$x_{i1}^* = \varphi_1(x_{i1-1}). \quad (2.15)$$

Для этой процедуры через каждые  $k$  чисел меняется начальное значение  $x_0$ , моделируемое по алгоритму второй процедуры:

$$x_{i2}^* = \varphi_2(x_{i2-1}). \quad (2.16)$$

Комбинацией обращений к двум процедурам ГСЧ удастся увеличить период аperiodичности на несколько порядков.

## 2.5. Примеры построения и использования псевдослучайных величин

Моделирование псевдослучайных величин, имеющих распределение вероятностей, отличающееся от равномерного, осуществляется путём функционального преобразования величин  $\xi$ . Рассмотрим наиболее употребительные алгоритмы такого функционального преобразования.

**Нормальное распределение** можно моделировать множеством способов. Рассмотрим два алгоритма. Первый алгоритм моделирования нормально распределённой величины  $\eta$  с параметрами  $M=0$  и  $\sigma^2=1$  основан на центральной предельной теореме теории вероятностей. Согласно этой теореме при сложении достаточно большого числа независимых случайных величин, сравнимых между собой по своим дисперсиям, получаем новую случайную величину  $\eta$ , распределённую по нормированному нормальному закону. Чем больше складывается значений  $\xi_i$ , тем ближе распределение  $\sum_i \xi_i$  оказыва-

ется к нормальному. Поскольку  $\xi_i$  имеют одну и ту же дисперсию и считаются независимыми друг от друга, то для практики достаточно от 6 до 12 сложений  $\xi_i$ :

$$\eta = \sqrt{\frac{12}{n}} \left[ \sum_{i=1}^n \xi_i - \frac{n}{2} \right], \quad (2.17)$$

где  $n$  – число обращений к ГСЧ.

Второй алгоритм является более экономичным, поскольку позволяет по двум обращениям к базовому ГСЧ ( $\xi_1$  и  $\xi_2$ ) получить пару псевдослучайных чисел ( $\eta_1$  и  $\eta_2$ ), распределённых по нормальному закону. Расчёты ведутся по формулам

$$\eta_1 = \sqrt{-2l_n \xi_1} \cos 2\pi \xi_2; \quad \eta_2 = \sqrt{-2l_n \xi_1} \sin 2\pi \xi_2. \quad (2.18)$$

**Экспоненциальное распределение** широко используется при моделиро-

вании СТС. Наиболее распространены следующие два способа имитации псевдослучайных величин, которые имеют одну и ту же функцию плотности распределения  $f(x) = \lambda e^{-\lambda x}$ , где  $\lambda$  – параметр экспоненциального распределения.

Первый способ основан на методе обратной функции

$$F(x) = 1 - e^{-\lambda x}. \quad (2.19)$$

Обратной к ней будет функция  $e^{-\lambda x}$ . Тогда экспоненциально распределённая величина с параметром  $\lambda$  формируется путём следующего функционального преобразования величины  $\xi$ :

$$\varepsilon = -\frac{1}{\lambda} \ln(1 - \xi). \quad (2.20)$$

Второй способ основан на методе композиции. Опытным путём задаются величины  $\tau$  и  $\theta$  такими, чтобы характеристики величины  $\varepsilon$  были как можно ближе к экспоненте. С помощью ГСЧ моделируется  $\theta$  штук значений  $\xi_1, \xi_2, \dots, \xi_\theta$ . Затем осуществляется пересчёт по формуле

$$\varepsilon = \tau - \max \{ \xi_1, \xi_2, \dots, \xi_\theta \}. \quad (2.21)$$

Для практики достаточно задавать  $\tau=4$  и  $\theta=4$ , что позволяет получать экспоненциально распределённую величину  $\varepsilon$  с параметром  $\lambda=1$ .

**Дискретные распределения** также моделируются из базовых значений  $\xi_i$ . При схеме независимых испытаний часто необходимо моделировать псевдослучайные числа, распределённые по **биномиальному закону**. Так, например, выборка из биномиального распределения получается прямым подсчётом числа единиц в  $n$ -разрядном случайном числе  $\xi$ .

Если необходимо моделировать дискретные случайные величины, имеющие **геометрическое** распределение

$$P(v=y) = P(1-P)^{y-1}, \quad \text{где } y = 1, 2, \dots, \quad (2.22)$$

то находят номер испытания ( $y$ ), которое принесёт первый успех для появления события  $A$  с вероятностью  $P$ . Поэтому последовательно обращаются к ГСЧ и моделируют очередное  $\xi_i$ . Проверяют выполнимость жребия первого типа. Номер обращения ( $y$ ) к ГСЧ, обеспечившего  $A_i$ : «истина», и даёт очередное значение величины  $y$ , распределённой по геометрическому распределению с параметром  $P$ .

В технике широко распространено распределение **Пуассона**:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (2.23)$$

где  $\lambda$  – параметр распределения Пуассона для дискретной переменной  $k$ . Для

моделирования величин  $k$  последовательно обращаются к ГСИ, получают каждый раз очередное значение  $\xi_i$  и проверяют выполнимость неравенства

$$\prod_{i=1}^k \xi_i \geq e^{-\lambda} > \prod_{i=1}^{k+1} \xi_i. \quad (2.24)$$

Как только это неравенство выполнится, то сумма  $\xi_i$  и будет иметь распределение Пуассона с параметром  $\lambda$ . Для более полного ознакомления с алгоритмами моделирования псевдослучайных величин, имеющих различные стандартные распределения, можно рекомендовать книги [1, 2].

В качестве примера использования метода Монте-Карло и генераторов псевдослучайных чисел рассмотрим моделирование поражения ракетми площадной цели со сложной структурой. В распоряжении нападающей стороны имеется  $n$  ракет. Зона разрушения одной ракеты представляет собой круг радиуса  $r$ . В результате  $n$  выстрелов ракеты будет поражена какая-то часть ( $S_n$ ) общей площади  $S_0$ . Чтобы увеличить долю поражения цели  $U = S_n / S_0$  и избежать ненужных перекрытий зон поражения, на площадной цели устанавливают  $n$  точек прицеливания ( $O_1, O_2, \dots, O_n$ ).

Известны характеристики баллистического рассеяния ракет (по осям координат соответственно  $\sigma_x^2$  и  $\sigma_y^2$ ). Предполагается, что систематические ошибки при попадании ракет на цель отсутствуют. Рассеяние ракет подчиняется двумерному нормальному закону распределения. Координаты точек попадания  $(x_j, y_j)$ ,  $j = \overline{1, n}$  независимы друг от друга и от координат других точек попадания.

Требуется: при заданном расположении точек прицеливания  $O_1, O_2, \dots, O_n$  вычислить следующие характеристики обстрела:  $m_u = M[U]$  – среднюю долю поражённой площади цели,  $D_u$  – дисперсию доли поражения площадной цели,  $P(U > \tilde{U})$  – вероятность того, что будет поражено не менее заданной доли площади  $\tilde{U}$ ,  $r_0$  – математическое ожидание числа «промахов» ракет.

Если не делать никаких упрощений о форме цели и зонах поражения, то аналитически эту задачу решить трудно. Проще её решить методом Монте-Карло. Каждая реализация представляет собой «обстрел» цели  $n$  ракетами, при котором все точки попадания  $(x_j, y_j)$  будут разыграны по жребию. Моделирование каждой  $i$ -й реализации будет состоять из единичных жребиев плюс расчёт площади поражённой цели  $S_n$  и производных от неё характеристик поражения цели. На рис. 2.3 представлен пример попадания ракет на площадную цель для  $i$ -й реализации по методу Монте-Карло.

Координаты попадания ракеты  $j$  распределены по двумерному нормальному закону с параметрами  $(m_{xj}, m_{yj}, \sigma_x, \sigma_y)$ , где  $m_{xj}$  и  $m_{yj}$  – координаты точек прицеливания  $O_j$ ,  $j = \overline{1, n}$ . По условию независимости друг от друга  $x_j$  и  $y_j$



полагаем, что коэффициент корреляции  $Z_{xy} = 0$ .

Тогда

$$x_j = \sigma_x \sqrt{2} \left( \sum_{k=1}^6 \xi_k - 3 \right) + m_{x_j}; \quad (2.25)$$

$$y_j = \sigma_y \sqrt{2} \left( \sum_{k=7}^{12} \xi_k - 3 \right) + m_{y_j}, \quad (2.26)$$

где  $\xi_k$  – 12 экземпляров случайных равномерно распределённых величин.

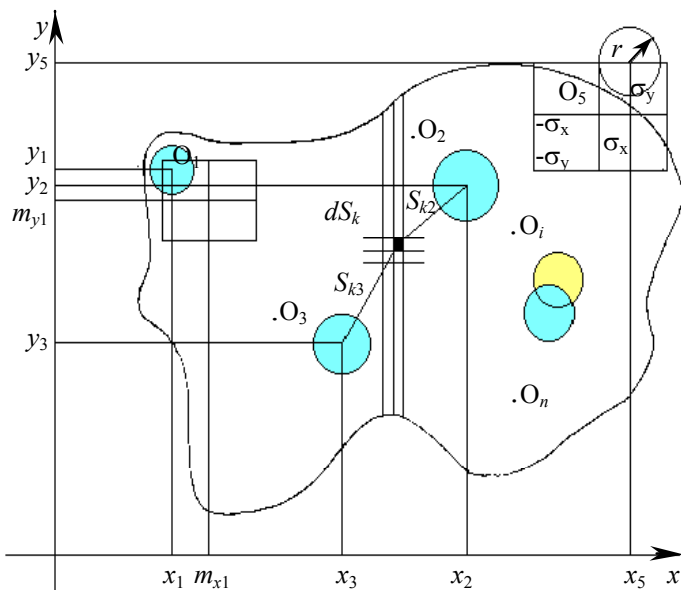


Рис. 2.3. Пример попадания  $n$  ракет по площадной цели для  $i$ -й реализации по методу Монте-Карло

Для моделирования из  $\xi_k$  нормально распределённых величин используют способ сумм. Допустим, что этап обстрела смоделирован и мы получили  $n$  точек попадания ракет, которые при заданном распределении точек прицеливания  $O_j$  попали так, как показано на рис. 2.3. Вокруг каждой точки попадания опишем круг радиуса  $r$  и подсчитаем площадь той части цели, которая покрыта хотя бы одним из кругов. Всю цель разделим на ряд элементарных площадок  $dS_j$ . Для каждой площадки  $j$  определяется расстояние  $\rho_{ij}$  от точки попадания  $j$ -й ракеты до центра площадки  $dS_j$ . Если хотя бы для одной из точек попадания  $j$  это расстояние оказалось равным или меньше  $r$  ( $\rho_{ij} \leq r$ ), то

считается  $dS_j$  поражённой, т. е. некоторая переменная-индикатор поражения площадки  $x_{ij}$ , которая в начале проверки площадок была равна 0, становится равной 1 ( $x_{ij} = 1$ , если  $\rho_{ij} \leq r$ ). После чего производится суммирование поражённых площадок по всей цели ( $S_n = \sum_{i,j} dS_j \cdot x_{ij}$ ). Далее для  $k$ -й реализации

процесса обстрела площади  $n$  ракетами ( $k = \overline{1, N}$ ) долю  $S_{цк}$  поражённой площади  $U_k$  находят как отношение числа площадок  $dS_{пj}$  к общему числу элементарных площадок в цели:

$$U_k = S_{цк} / S_{ц}. \quad (2.27)$$

Одновременно для  $k$ -й реализации вычисляют количество ракет, расстояние которых от точек попадания больше, чем границы площадки  $S_{ц}$  (т.е. находят число промахов  $r_{ок}$ ). Математическое ожидание и дисперсия доли поражённой площади определится из соотношений:

$$M_U = \frac{1}{N} \sum_{k=1}^N U_k; \quad D_U = \frac{1}{N} \sum_{k=1}^N U_k^2 - M_U^2. \quad (2.28)$$

Вероятность того, что доля поражённой площади будет не меньше, чем  $\tilde{U}$ , определяют следующим образом. С каждой  $k$ -й реализацией формируется число

$$Z_k = \begin{cases} 1 & \text{при } U_k \geq \tilde{U}, \\ 0 & \text{при } U_k < \tilde{U}. \end{cases} \quad (2.29)$$

Тогда  $P(U \geq \tilde{U}) = \frac{1}{N} \sum_{k=1}^N Z_k$ . Наконец, определяют среднее число промахов ракеты по цели:

$$\bar{r}_0 = \frac{1}{N} \sum_{k=1}^N r_{ок}. \quad (2.30)$$

Само же число реализаций  $N$  по методу Монте-Карло рассчитывается исходя из заданной доверительной вероятности  $\beta$ , т.е. находят такое число элементов выборки, чтобы была обеспечена требуемая доверительная вероятность  $\beta$ . Расчётная формула имеет вид

$$N \geq S^2 / (1 - \beta) \Delta^2, \quad (2.31)$$

где  $\Delta = k S_{\bar{x}}$ ;  $S_{\bar{x}} = S / \sqrt{N}$ ;  $k = 1, 2, 3, \dots$ ;  $S$  – выборочное среднее квадратическое отклонение;  $\Delta$  – допустимая (требуемая) точность оценки дисперсии вычислений.

## 2.6. Моделирование случайных процессов

Случайный процесс  $X(t)$  представляется некоторой случайной функцией, аргументом которой является время. Определённый вид этой функции  $X_i(t)$ , полученный в результате опыта, называют реализацией случайного процесса  $X(t)$ . Множество реализаций  $X_1(t), X_2(t), \dots, X_n(t)$  образует ансамбль. Совокупность мгновенных значений случайного процесса, заданного ансамблем в произвольный момент времени  $t_k$ , называют **сечением** случайного процесса (рис. 2.4). Различают пять основных видов случайного процесса:

1. Дискретную случайную последовательность (временной ряд), представляющую собой сложный процесс, у которого область значений и область определения являются дискретными множествами.

2. Непрерывная случайная последовательность – это случайный процесс, у которого область значений непрерывна, а область определения дискретна.

3. Дискретный случайный процесс – это область значений есть дискретное множество, а область определения – непрерывное множество.

4. Непрерывный случайный процесс – область значений и область определений есть непрерывное множество.

5. Поток, который представляет собой случайный процесс в виде последовательности точек, расположенных случайным образом на оси времени.

Выделим на временной оси произвольный момент времени  $t_k$ . Тогда случайная функция  $X(t)$  превратится в случайную величину  $x_k = X(t_k)$  с одномерным законом распределения  $F_1(x_k, t_k)$ . Отметим, что одномерные законы распределения могут быть получены для любых сечений  $t_i, i = \overline{1, m}$ , но они не могут ответить на вопрос о зависимости между случайными величинами  $x_1, x_2, \dots, x_m$  при различных  $m$ .

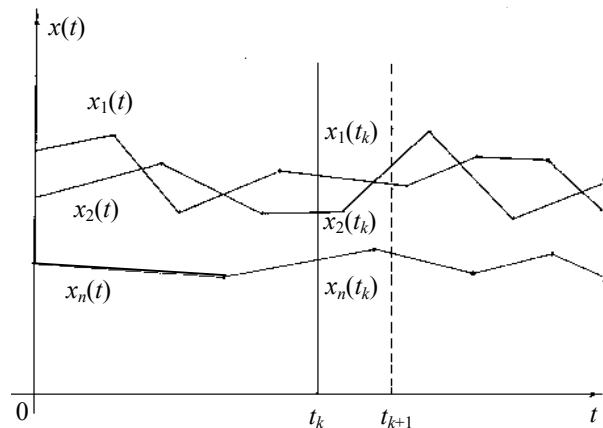


Рис 2.4. Пример проведения сечения случайного процесса

Полностью статистические свойства  $X(t)$  характеризует многомерный закон распределения случайной величины  $F_m(x_1, x_2, \dots, x_m; t_1, t_2, \dots, t_m)$ , практическое использование которого затруднительно. Поэтому обычно применяют следующие числовые характеристики случайного процесса  $X(t)$ .

1. Математическое ожидание процесса  $X(t)$ , которое является неслучайной функцией,  $m_x(t)$ . При каждом значении аргумента  $t$  эта функция равна математическому ожиданию соответствующего сечения

$$m_x(t) = M\{x(t)\}. \quad (2.32)$$

2. Дисперсия процесса  $X(t)$  – также неслучайная функция, значения  $D_x(t)$  которой для каждого  $t$  равно дисперсии соответствующего сечения:

$$D_x(t) = \sigma_x^2(t) = M\{[x(t) - m_x(t)]^2\}. \quad (2.33)$$

3. Автокорреляционная функция процесса  $X(t)$  – также неслучайная функция двух аргументов  $R_x(t, t')$ , которая при каждой паре значений времени  $(t, t')$  равна корреляционному моменту соответствующих сечений:

$$R_x(t, t') = M\{[x(t) - m_x(t)][x(t') - m_x(t')]\}. \quad (2.34)$$

Рассмотрим порядок определения характеристик случайной функции из опыта. На практике обычно регистрируются значения функции  $X(t)$  в моменты времени  $t_1, t_2, \dots, t_m$ . Пусть это будет последовательность значений функции

$$X(t) = (X_1(t), X_2(t), \dots, X_n(t)).$$

Результаты измерений представим следующей матрицей:

$X_i(t)$ \ $t$	$t_1$	$t_2$	$t_3$	$\dots$	$t_m$
$X_1(t)$	$x_1(t_1)$	$x_1(t_2)$	$x_1(t_3)$	$\dots$	$x_1(t_m)$
$X_2(t)$	$x_2(t_1)$	$x_2(t_2)$	$x_2(t_3)$	$\dots$	$x_2(t_m)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$X_n(t)$	$x_n(t_1)$	$x_n(t_2)$	$x_n(t_3)$	$\dots$	$x_n(t_m)$

Оценкой  $m_x(t_j)$  является среднее арифметическое

$$\bar{X}(t_j) = \frac{1}{n} \sum_{i=1}^n x_i(t_j). \quad (2.35)$$

Оценкой  $m_x(t)$  является вектор, компонентами которого будут

$$\bar{X}(t) = \langle \bar{X}(t_1), \bar{X}(t_2), \dots, \bar{X}(t_m) \rangle.$$

Оценкой  $\sigma^2(t_j)$  является эмпирическая дисперсия

$$S^2(t_j) = \sum_{i=1}^n [x_i(t_j) - \bar{x}(t_j)]^2 / (n-1). \quad (2.36)$$

Оценкой  $\sigma^2(t)$  также является вектор, у которого компоненты суть эмпирические дисперсии

$$S^2(t) = \langle S^2(t_1), S^2(t_2), \dots, S^2(t_m) \rangle.$$

Оценкой автокорреляционной функции  $R_x(t_k, t_l)$  будет величина

$$\hat{R}_x = \frac{\sum_{i=1}^n [x_i(t_k) - \bar{x}(t_k)][x_i(t_l) - \bar{x}(t_l)]}{(n-1)}. \quad (2.37)$$

Оценкой  $R_x(t, t')$  будет уже матрица фактических значений автокорреляционных функций вида:

	$t_1$	$t_2$	$t_3$	...	$t_{m-1}$	$t_m$
$t_1$	$S^2(t_1)$	$\hat{R}_x(t_1, t_2)$	$\hat{R}_x(t_1, t_3)$	...	$\hat{R}_x(t_1, t_{m-1})$	$\hat{R}_x(t_1, t_m)$
$t_2$		$S^2(t_2)$	$\hat{R}_x(t_2, t_3)$	...	$\hat{R}_x(t_2, t_{m-1})$	$\hat{R}_x(t_2, t_m)$
...	...	...	...	...	...	...
$t_{m-1}$				...	$S^2(t_{m-1})$	$\hat{R}_x(t_{m-1}, t_m)$
$t_m$				...		$S^2(t_m)$

Важным классом случайных процессов являются стационарные случайные процессы, которые характеризуются однородностью протекания во времени. Стационарные случайные процессы не имеют ни начала, ни конца. За начало процесса может быть выбран любой момент времени, поскольку вероятностные характеристики стационарного случайного процесса не зависят от времени. Случайный процесс называют стационарным в узком смысле, если все конечномерные функции распределения вероятностей любого порядка инвариантны относительно сдвига во времени. Поэтому для любого  $m$  и  $t$  справедливо соотношение

$$F_m(x_1, \dots, x_m; t_1 - t_0, \dots, t_m - t_0) = F_m(x_1, \dots, x_m; t_1, \dots, t_m).$$

Таким образом, одномерная функция распределения, математическое ожидание и дисперсия такого случайного процесса не зависят от времени. Двумерная же функция распределения и автокорреляционная функция стационарного случайного процесса являются функциями длительности временного промежутка между сечениями процесса ( $\tau$ ). Таким образом, для некоторого фиксированного  $\tau$  эта функция также постоянна:

$$\begin{aligned} F_1(x_1, t) &= F_1(x_1); & F_2(x_1, x_2, t, t') &= F_2(x_1, x_2, \tau); \\ m_x(t) &= m_x = \text{const}; & \sigma_x^2(t) &= \sigma_x^2 = \text{const}; & R_x(t, t') &= R_x(\tau). \end{aligned} \quad (2.38)$$

Если выполняются только три последних условия, то случайный процесс называют стационарным в широком смысле.

Для организации вычислений характеристик стационарного случайного процесса исследователь должен располагать следующими статистиками:

$$\sum_{i=1}^n x_i(t); \quad \sum_{i=1}^n [x_i(t_j)]^2; \quad \sum_{k=0}^{m-1} \sum_{i=1}^n x_i(t) x_i(t + k \Delta t). \quad (2.39)$$

Длинную реализацию стационарного случайного процесса обычно разбивают на участки одинаковой длительности. Для некоторых стационарных случайных процессов каждый из таких участков может рассматриваться как «полномочный представитель» отрезка реализации статистического ансамбля одинаковых систем. Стационарные случайные процессы, для которых это разбиение возможно, называются эргодическими. Их вероятностные характеристики не зависят ни от начала отсчёта, ни от номера реализации. Для такого случайного процесса результаты усреднения по времени (по одной реализации) совпадают с соответствующими результатами вычислений по ансамблю. При этом требуемые вероятностные характеристики процесса могут быть получены при обработке лишь одной реализации достаточно большой длительности. Неограниченное убывание автокорреляционной функции  $R_x(t, t')$  стационарного случайного процесса при  $[t, t'] \rightarrow \infty$  является достаточным условием определения математического ожидания как среднего по времени. Если же стационарная случайная функция к тому же имеет нормальное распределение, то условие  $\lim R_x[t, t'] \rightarrow \infty$  при  $[t, t'] \rightarrow \infty$  является достаточным для определения по одной реализации всей автокорреляционной функции. Для вычисления числовых характеристик стационарного эргодического процесса по одной реализации используются следующие формулы:

$$\bar{x} = \frac{1}{m} \sum_{j=1}^m x(t_j); \quad (2.40)$$

$$S^2 = \sum_{j=1}^m [x(t_j)]^2 / (m-1) - \left[ \sum_{j=1}^m x(t_j) \right]^2 / (m(m-1)); \quad (2.41)$$

$$\hat{R}_x(\tau) = \sum_{j=1}^r (x(t_j) x(t + \tau) - \bar{x}^2) / r. \quad (2.42)$$

Для стационарного эргодического многомерного случайного процесса

помимо  $\hat{R}'_x(\tau)$  также вычисляют взаимнокорреляционный момент:

$$S_{gk}^g = \sum_{j=1}^m [x^g(t_j) - \bar{x}^g] [x^k(t_j) - \bar{x}^k] / (m-1); \quad g = \overline{1, P}; \quad k = \overline{1, P}. \quad (2.43)$$

Нестационарный случайный процесс имеет тенденцию развития во времени. Его характеристики зависят от начала отсчёта процесса и времени. Они могут быть определены для каждого  $t_i$  по ансамблю. Обычно стараются нестационарный случайный процесс свести к стационарному путём деления оси времени на промежутки, на которых этот процесс оказывается стационарным. Характеристики нестационарной случайной функции могут быть вычислены и по одной реализации. Для этого применяют методы сглаживания на основе использования скользящей и экспоненциальной средней. Разбивают интервал  $(O, T_p)$  реализации этой случайной функции на  $m$  равных частей с шагом  $\Delta t$ . Для определения *скользящего среднего* (сглаженного значения) функции в точке  $t_k$  выделим из множества точек  $\{t_1, t_2, \dots, t_k, \dots, t_m\}$  такое подмножество  $(2v+1)$  точек  $\{t_{k-v}, \dots, t_{k-1}, t_k, t_{k+1}, \dots, t_{k+v}\}$ , в котором центральной является точка  $t_k$ . Тогда скользящая средняя  $(2v+1)$ -го порядка в точке  $t_k$  будет вычислена по формуле

$$\bar{x}_c(t_k) = \sum_{r=-v}^v x(t_{k+r}) / (2v+1). \quad (2.44)$$

Дисперсия скользящей средней определится по формуле

$$S^2 = \sum_{r=-v}^v [x(t_{k+r}) - \bar{x}_c(t_k)]^2 / (2v+1). \quad (2.45)$$

Можно также произвести учёт «старения» данных по мере их удаления от момента, для которого определено скользящее среднее. В этом случае рассчитывается экспоненциальное среднее по формуле

$$\bar{x}_3(t_k) = \alpha \bar{x}_c(t_k) + (1-\alpha) \bar{x}_c(t_{k-1}). \quad (2.46)$$

Обычно выбирают значение коэффициента  $\alpha$  в пределах  $\alpha=0,1 \dots 0,3$

$$S_3^2 = \alpha \bar{x}_c^2(t_k) + (1-\alpha) \bar{x}_c^2(t_{k-1}) - \bar{x}_3^2(t_k). \quad (2.47)$$

Случайный поток событий называется *стационарным*, если вероятность попадания того или иного числа событий на интервале времени длиной  $\tau$  зависит только от длины этого интервала и не зависит от того, где именно на оси времени он расположен. Поток событий считается ординарным, если вероятность попадания на интервал  $\Delta t$  двух и более событий мала по сравнению с вероятностью попадания на этот интервал одного события. Поток

событий называется без последствия, если для любых неперекрывающихся интервалов число событий не зависит от числа событий попадания на другой интервал. Поток событий называется простейшим, если он одновременно стационарен, ординарный и без последствия.

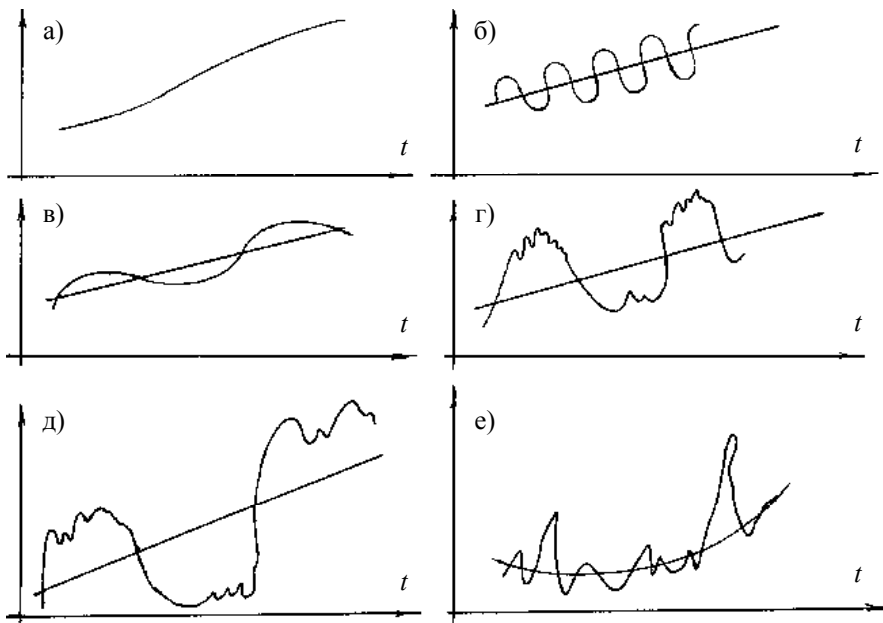
## 2.7. Представление случайных процессов временными рядами

Временные ряды (ВР) представляют собой упорядоченные во времени данные. Членами временного ряда являются либо суммы, либо средние, либо числовая информация, полученная в некоторые моменты времени. При этом все члены ВР в моменты  $t_1, \dots, t_k$  должны быть сравнимы. Поэтому первоначальные данные корректируются таким образом, чтобы их можно было сравнивать. Динамику ВР изучают в долго-, средне- и краткосрочной перспективах. При анализе ВР наиболее часто используется выделение: трендов движения ВР; краткосрочных трендов, длинных циклов, краткосрочных циклов, сезонных колебаний. Концепция тренда представляется в виде непрерывной линии, сглаживающей краткосрочные колебания. Для описания основной тенденции ВР и для сравнения различных трендов используют различные математические функции, сглаживающие исходные данные. Дополнительно для сглаживания циклической составляющей и сезонных колебаний используют метод скользящей средней.

Базовая концепция цикла состоит в изучении альтернативных фазовых состояний – фаз роста и падения ВР. Поэтому при анализе циклов любой длины используются такие термины, как «верхняя поворотная точка», «нижняя поворотная точка», фазы «подъёма» и «спада». Расстояние между двумя нижними поворотными точками называется длиной цикла ВР. Модели, описывающие поведение ВР, имеют вид:

- главного временного ряда (рис. 2.5, а);
- ВР на основе квартальных данных с сезонными колебаниями (рис. 2.5, б);
- циклических колебаний без сезонных колебаний (рис. 2.5, в);
- одновременно сезонных и циклических колебаний (рис. 2.5, г);
- длинные циклы с сезонными колебаниями (рис. 2.5, д);
- со сложной структурой (рис. 2.5, е).

Тренды ВР являются либо прямыми, либо линиями «без перегибов». Вместо описания трендов ВР в терминах абсолютных средних приростов широко используют описания в терминах средних приростов в процентах (или темпов). Такой подход особенно важен для тех случаев, когда исходные данные трудно сопоставимы друг с другом. Поэтому ВР могут расти (падать) с постоянным, возрастающим или убывающим данным. Для удобства сравнения ВР они зачастую аппроксимируются прямыми или другими линиями. Эти линии используются также для целей долгосрочного прогноза.



Чаще всего для аппроксимации используют прямую, параболу и экспоненту.

Коэффициенты  $a$  и  $b$  прямой линии  $Y_x = a + bX$  можно подобрать с помощью метода наименьших квадратов. Для этой цели решают систему двух линейных уравнений

$$\begin{aligned} Na + (\Sigma X)b &= \Sigma Y; \\ (\Sigma X)a + (\Sigma X^2)b &= \Sigma XY. \end{aligned} \quad (2.48)$$

Тангенс угла наклона этой прямой указывает либо на рост  $Y$ , либо на падение  $Y$ , представляя собой при этом угол наклона прямой. Экстраполяция прогноза трендовых значений основана на предположении, что ряды будут расти так же, как и в базовом периоде. Если подгоняется к данным прямая, то делается допущение, что скорость изменения ВР в среднем постоянна. При подгонке экспонентой предполагается, что скорость изменения ряда логарифмов эмпирических значений в среднем постоянна. Обычно не рекомендуется экстраполировать на будущее тренды, если предполагается, что ряд может сильно поменять непрерывное движение (случай очень короткого периода). Обычно экстраполяция трендов относится к категории долгосрочных прогнозов. Так как тренды сглаживают циклические колебания, их экстраполяцию не используют для краткосрочных предсказаний.

ВР может возрастать (убывать) с переменной скоростью. Тогда его лучше аппроксимировать параболой  $Y_x = a + bX + cX^2$ . Параметры этой параболы ( $a$ ,  $b$ ,  $c$ ) находят решением системы трёх линейных уравнений

$$\begin{aligned} Na + (\Sigma X)b + (\Sigma X^2)c &= \Sigma Y; \\ (\Sigma X)a + (\Sigma X^2)b + (\Sigma X^3)c &= \Sigma XY; \\ (\Sigma X^2)a + (\Sigma X^3)b + (\Sigma X^4)c &= \Sigma X^2Y. \end{aligned} \quad (2.49)$$

Часто возникает проблема выбора не между прямой и параболой, а между параболой и некоторой другой кривой. Если используется тренд параболы, то первые разности ряда располагаются вокруг прямой. Но при этом сам ВР должен быть гладким.

В ряде случаев для экстраполяции ВР используется логарифмическая прямая, которая получается из экспоненциальной функции вида

$$c Y_x = ab^x; \quad \log_c Y_x = \log_c a + (\log_c b)X. \quad (2.50)$$

ВР такого типа изменяется в каждую единицу времени в среднем с постоянным темпом в процентах. Поэтому тангенс наклона этой прямой означает постоянное приращение ВР. Используя экспоненциальную функцию (2.50) мы найдём постоянный темп изменения в процентах относительно первого и последнего членов ряда, а также темп изменения в процентах для тренда, полученного методом наименьших квадратов.

Если необходимо сравнить графически два и более ВР по темпам их изменений, то используют следующую координатную систему. По ординатам указываются алгоритмы значений ВР, а по оси абсцисс – времена. Ось ординат такого графика маркируется числами от 1 до 10. Такой график называется двучиклическим полулогарифмическим. Эту кривую можно двигать вверх и вниз без изменения её вида, умножая исходную шкалу на числа 10, 20, ..., 100. Когда необходимо изобразить два и более ВР на одном графике, используют несколько шкал. При этом множители у элементов этих ВР выбираются такими, чтобы ВР не пересекались и в то же время не были слишком далеки друг от друга.

Различают следующие виды ВР: с предшествующими, совпадающими и запаздывающими поворотными точками.

К **предшествующим** ВР относятся: средняя рабочая неделя, число производственных рабочих, объём выпуска, средненедельное число заявлений на пособие по безработице.

К **совпадающим** рядам относятся: объём продаж, процент безработицы, число лиц, работающих по найму, и т. д.

Примеры **запаздывающих** рядов: затраты предпринимательского сектора, балансовая стоимость, банковские проценты по займам и т. д.

Чтобы увидеть циклы ВР в чистом виде, необходимо аппроксимировать ВР (прямой, параболой, экспонентой). Затем вычисляют значения тренда и вычитают их из исходного ряда либо делят значения ряда на трендовые. В первом случае получают разности  $(Y - cY_x)$ , а во втором – процент отноше-

ний. Чтобы легче увидеть циклические движения, необходимо, чтобы тренд сглаживал все циклы, пересекая их. Однако тренд без точек перегиба часто не удовлетворяет этому условию, поэтому используют такое понятие, как скользящее среднее ( $pM_x$ ), где  $p$  – число объединяемых временных интервалов. Скользящее среднее представляет собой тренд, сглаживающий цикл любого характера и длины. Его обычно используют для выделения циклических и сезонных колебаний. Расчёты ведутся по формуле

$$pM_{x'} = \sum_{j=1}^p \omega_j x_{x+i-2} / \sum_{j=1}^p \omega_j, \quad (2.51)$$

где  $x' = [x + (p-1)/2] > 0$ ;  $p$  – число временных интервалов,  $pW_j = (\omega_1, \omega_2, \dots, \omega_p)$ .

Результаты расчётов заносят в таблицу вида:

### Пример 1

$X$	$Y_x$	$3M_x$ при разных $\omega_j$
0	$Y_0$	—
1	$Y_1$	$3M_1 = (\omega_1 Y_0 + \omega_2 Y_1 + \omega_3 Y_2) / \sum_{j=1}^3 \omega_j$
2	$Y_2$	$3M_2 = (\omega_1 Y_1 + \omega_2 Y_2 + \omega_3 Y_3) / \sum_{j=1}^3 \omega_j$
3	$Y_3$	$3M_3 = (\omega_1 Y_2 + \omega_2 Y_3 + \omega_3 Y_4) / \sum_{j=1}^3 \omega_j$
		и т. д.

### Пример 2

$X$	$Y_x$	$4M_x$ при $\omega_j = 1$
0	$Y_0$	—
1	$Y_1$	$4M_1 = (Y_0 + Y_1 + Y_2 + Y_3) / 4$
2	$Y_2$	$4M_2 = (Y_1 + Y_2 + Y_3 + Y_4) / 4$
3	$Y_3$	$4M_3 = (Y_2 + Y_3 + Y_4 + Y_5) / 4$
		и т. д.

Как видим, скользящие средние могут быть взвешенными и простыми. Значение количества интервалов сложения ( $p$ ) и веса  $pW_j$  скользящих средних  $pM_x$  определяется длиной цикла и условием того, чтобы  $pM_x$  сглаживало циклы, пересекая их. При этом из ВР уже вычтен тренд. Поэтому при

анализе циклов исследуются все отклонения от скользящей средней. Когда выбирают скользящую среднюю, соответствующие ВР, у которого несколько циклов, то необходимо принимать во внимание длину циклов. Если ВР не содержит слишком длинных или слишком коротких циклов, то  $pM_x$  должно быть равно длине самого продолжительного цикла. В других случаях обычно удобно брать  $p$  немного меньше длины наибольшего цикла. Очень важно зафиксировать поворотные точки во ВР. Затем вычисляют средние арифметические данных для одного цикла и отдельные значения ряда в этом цикле выражают в виде  $q$ -го отношения к этому среднему. Эта операция означает анализ специфики циклов. Затем те же ВР расчленяют на сегменты, соответствующие опорным точкам ВР, вычисляют средние арифметические на интервалах и рассчитывают процентные отношения. Таким способом иногда удаётся освободиться от тренда и сделать сопоставимыми отдельные циклы, сравнивая при этом процентные отношения фактических значений к средним. Таким способом облегчается анализ фаз подъёма и спада во ВР.

Вторым подходом к сглаживанию циклов является построение трендовой линии «среднего цикла». Для этого фиксируют поворотные точки ВР, находят среднюю арифметическую циклов и соответственно им наносят горизонтальные прямые. Далее проводят вертикали к этим линиям в точках, лежащих посередине между нижними и верхними поворотными точками. Соединяют между собой точки пересечения этих вертикалей со средними линиями. Полученная таким способом ломаная линия и будет искомым трендом ВР (рис. 2.6).

Основная цель изложенного на рис. 2.6 метода состоит в выделении циклов таким образом, чтобы

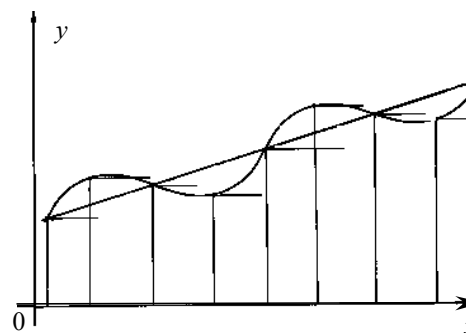


Рис. 2.6. Пример построения линии искомого тренда

длина каждого цикла прорабатывалась визуально.

ВР с сезонными колебаниями можно представить моделью  $Y = T + C + S + T$ . Существуют две основные причины, по которым проводится анализ сезонности ВР. Во-первых, появляется необходимость упрощения сезонных колебаний для того, чтобы в чистом виде можно было рассматривать циклические дви-

жения. Эта операция называется «сезонным сглаживанием». Во-вторых, иногда требуется изучить подробно само явление и вычислить средние характеристики сезонности для самого ВР. Проще всего подобные операции

реализуются путём вычисления индексов сезонности по следующей методике.

**Этап 1.** Поскольку данные являются квартальными, то ВР имеют период сезонности  $p=4$ . Предварительно из ВР исключают тренд. В идеале надо вычислить простую скользящую  $4M_X$ . Но так как среднюю следует центрировать в точках  $X$ , то применяют взвешенную скользящую  $5M_X$  с весами  $5W_j=(1, 2, 2, 2, 1)$ . Если ВР представлен месячными данными, то необходимо рассчитать  $12M_X$  с весами  $12W_j=(1, 2, \dots, 2, 1)$ . Но поскольку  $P$  велико, то в данном случае можно использовать простую скользящую среднюю.

**Этап 2.** Выражают исходные данные  $Y_x$  в виде процентных отношений  $(Y_x/M_X) \cdot 100$ .

**Этап 3.** Эти процентные отношения классифицируются по кварталам (месяцам) и годам. Для каждого квартала (или месяца) вычисляют среднее значение признака. Если сезонные колебания неоднородны от года к году, то для вычисления индексов сезонности нужно иметь длинные ряды.

**Этап 4.** Вычисляют преобразованные средние. Каждую среднюю умножают на некоторое значение  $K$  (которое подбирают исходя из общего масштаба), и после округления получают индексы сезонности  $S_{11}, S_{12}, S_{13}, S_{14}$ , где

$$K = \sum_{j=1}^4 S_{1j}.$$

Очистить ряды от сезонных колебаний означает получить кривую, похожую на кривую скользящей средней. Ряды делятся по индексам сезонности. Во многих ВР явления сезонности могут носить нерегулярный характер. В этих случаях разница между очищенным ВР и линией скользящего среднего  $5M_X$  существенна.

Циклические колебания анализировать легче, если предварительно исключить из ряда сезонные колебания. Такие ряды зачастую применяют при прогнозировании.

## 2.8. Моделирование поведения сложных технических систем марковскими процессами

Пусть имеется сложная техническая система, которая в произвольный момент времени  $t_k$  находится в одном из состояний  $S_i$ ,  $i = \overline{1, n}$  с вероятностью  $P_i(t_k)$ . Через некоторые промежутки времени система переходит из одного состояния в другое. Каждый такой переход называется шагом процесса. Случайный процесс, протекающий в СТС с переходом из одного состояния в другое состояние  $S$ , называется **марковским**, если для произвольного момента времени  $t_k$  вероятность любого состояния системы в будущем (при  $t > t_k$ ) зависит только от её состояния в настоящем (при  $t = t_k$ ) и не зависит от

того, когда и каким образом система пришла в это состояние. Различают следующие виды марковского процесса:

- с дискретными состояниями и дискретным временем (стохастическая последовательность, называемая дискретной цепью Маркова);
- с дискретными состояниями и непрерывным временем (непрерывная цепь Маркова).

Оба типа цепей Маркова могут быть однородными и неоднородными. Для дискретной цепи Маркова смены состояний случайного процесса происходят в дискретные моменты времени  $t_j$  с постоянным шагом  $\tau=t_{i-1}-t_i=\text{const}$ . Состояние системы  $S_i$  в момент  $t_k$  необходимо характеризовать условными вероятностями  $P_{ij}^{(k)}$  того, что система за один шаг перейдёт в какое-либо состояние  $S_j$  при условии, что в момент  $t_{k-1}$  она находилась в состоянии  $S_i$ . Вероятности  $P_{ij}^{(k)}$  являются основными характеристиками марковской цепи. Они называются вероятностями перехода или иногда **переходными вероятностями**. Поскольку система может находиться в одном из состояний, то для каждого момента времени  $t_k$  необходимо задать  $n^2$  вероятностей перехода в виде матрицы

$$P^k = \begin{pmatrix} P_{11}^k & P_{12}^k & \dots & P_{1n}^k \\ P_{21}^k & P_{22}^k & \dots & P_{2n}^k \\ \dots & \dots & \dots & \dots \\ P_{n1}^k & P_{n2}^k & \dots & P_{nn}^k \end{pmatrix}. \quad (2.52)$$

Эта матрица называется матрицей переходных вероятностей или матрицей переходов. При этом необходимо выполнение условий:

$$\sum_{j=1}^n P_{ij}^{(k)} = 1; \quad i = \overline{1, n}. \quad \text{Эта матрица обязательно является квадратной с неотри-$$

цательным элементом. Когда вероятности перехода не зависят от времени, а зависят только от величины  $\tau$  и не изменяются при сдвиге вдоль временной оси, то цепь Маркова называется однородной. Для задания такой цепи используется матрица вида

$$P = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{pmatrix}. \quad (2.53)$$

Вероятности перехода представляют собой важнейшие характеристики любой марковской цепи. Если отнести матрицу переходов к первому шагу, определяющему начало системы, то для исключения условностей необходи-

мо ещё задать вероятности начальных состояний. Вероятности начальных состояний  $P_1^{(0)}, P_2^{(0)}, \dots, P_n^{(0)}$  являются безусловными вероятностями и образуют в матрице строку  $P_0 = [P_1^{(0)}, P_2^{(0)}, \dots, P_n^{(0)}]$ , сумма элементов которой должна быть равна единице. Матрица переходов даёт исчерпывающее представление о вероятностях возможных переходов за один шаг. Возникает вопрос: как рассчитать вероятности того, что СТС, находящаяся в данный момент в состоянии  $S_1$ , переходит в состояние  $S_i$  за  $r$  шагов. Матрица переходов за  $r$  шагов  $P(r)$  вычисляется как  $r$ -я степень матрицы переходов за один шаг  $P(r) = P^r$ . Безусловные вероятности системы на  $r$ -м шаге определяются из выражений

$$P_r = P_0 P^r; \quad P_j^{(2)} = \sum_{i=1}^n P_i^{(0)} P_{ij}(r). \quad (2.54)$$

Различают два вида марковских цепей: эргодические и поглощающие. Это различие основано на классификации состояний. Состояние  $S_i$  называется невозвратным, если существуют такие состояния  $S_j$  ( $j \neq i$ ) и число шагов  $r$ , что  $P_{ij}(r) > 0$ , но  $P_{ij}(q) = 0$  для всех  $q$ . Все остальные состояния называются **возвратными**. Таким образом, из невозвратного состояния всегда можно с положительной вероятностью за какое-либо число шагов перейти в некоторое другое состояние. В то же время вернуться из этого состояния в первоначальное невозможно. Если выбрать такие состояния  $S_i$  и  $S_j$ , что для них при некоторых  $r$  и  $q$  выполняются неравенства  $P_{ij}(r) > 0$  и  $P_{ij}(q) > 0$ , то они называются сообщающимися состояниями. Если же  $S_j$  сообщается с  $S_i$ , а  $S_i$  сообщается с  $S_k$ , то  $S_j$  сообщается с  $S_k$ . Это обстоятельство позволяет разделить множество возвратных состояний на классы (подмножества) сообщающихся состояний. Состояния, принадлежащие к разным классам, не общаются между собой. Если множество возвратных состояний состоит из одного класса, то оно называется **эргодическим**. Существуют ещё и поглощающие состояния. Если  $S_i$  есть некоторое поглощающее состояние, то  $P_{ij} = 1$  или же  $P_{ij} = 0$ . Цепи Маркова, не содержащие невозвратные множества и образующие эргодическое множество, называют **эргодическим**. Свойства и методы расчёта параметров поглощающих и эргодических цепей Маркова различны.

Для эргодической цепи Маркова характерно, что она из каждого своего состояния может попасть в любое другое. Основным вопросом при её исследовании является изучение того, как изменяются с течением времени безусловные вероятности эргодической цепи. Вообще для эргодической цепи Маркова справедлива следующая теорема: при увеличении числа шагов  $r$  вероятности  $P_{ij}(r)$  стремятся к определённому пределу  $\pi_j$ , называемому финальной вероятностью:

$$\lim_{r \rightarrow \infty} P_{ij}(r) = \pi_j \text{ для всех } i \text{ и } j. \quad (2.55)$$

Безусловные вероятности  $P_j^{(r)}$  при увеличении  $r$  также стремятся к финальным вероятностям:

$$\lim_{r \rightarrow \infty} P_j^{(r)} = \pi_j. \quad (2.56)$$

Это означает, что при достаточно большом  $r$  безусловные вероятности состояний, составляющие в совокупности одномерные распределения, не зависят от времени. Поскольку у однородных цепей Маркова переходные вероятности  $\{P_{ij}\}$  не зависят от времени, то при  $r \rightarrow \infty$  можно считать, что двумерное распределение безусловных вероятностей состояний зависит лишь от разности моментов времени. Из корреляционной теории случайных процессов [6] известно, что если двумерная плотность вероятностей зависит от разности времён между сечениями случайного процесса, то такой процесс является стационарным в широком смысле. Очевидно, что стационарный режим возможен лишь у однородных цепей Маркова.

До наступления стационарного режима система находится в переходном режиме, длительность которого можно определить, выбрав некоторый критерий, зависящий от разности ( $\pi_j = P_j^{(r)}$ ). Например,  $|\pi_j - P_j^{(r)}| < \varepsilon$ ,  $\varepsilon > 0$ . Вычисление  $\pi_j$  для известного начального распределения  $\{P_j^{(0)}\}$  и заданной матрицы переходов  $\{P_{ij}\}$  является важной задачей при изучении эргодических цепей Маркова. Существуют два способа решения этой задачи. Первый способ основан на теореме Маркова, он предусматривает возведение матрицы переходов в достаточно большую степень  $r$ , что должно дать матрицу исходных вероятностей

$$\lim_{r \rightarrow \infty} P(r) = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_n \\ \pi_1 & \pi_2 & \dots & \pi_n \\ \dots & \dots & \dots & \dots \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix}. \quad (2.57)$$

Однако определить финальные вероятности  $\pi_j$  подобным образом весьма трудоёмко. Поэтому гораздо проще воспользоваться другим способом, а именно  $\pi_j$  находят путём решения системы линейных однородных алгебраических уравнений с  $n$  неизвестными вида

$$\pi_j = \sum_{i=1}^n \pi_i P_{ij}, \quad j = \overline{1, n}.$$

Для непрерывной цепи Маркова вводится понятие **плотности вероятностей перехода** (интенсивности потока событий) как предела отношения



вероятности перехода системы за время  $\Delta t$  из состояния  $i$  в состояние  $j$  к длине промежутка  $\Delta t$ :

$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} P_{ij}(\Delta t) / \Delta t, \quad i \neq j. \quad (2.58)$$

Если  $\lambda_{ij}$  не зависят от  $t$  (т. е. от того, в какой момент начинается интервал  $\Delta t$ ), то непрерывная цепь Маркова называется однородной. В противном случае цепь Маркова является неоднородной. Однородная непрерывная цепь Маркова характеризуется тем, что все потоки, переводящие систему из одного состояния в другое, являются простейшими (т. е. стационарными пуассоновскими потоками). При этом время непрерывного пребывания цепи в каждом состоянии распределено по экспоненциальному закону. Для неоднородной цепи Маркова эти промежутки времени не распределены по показательному закону. Если непрерывная цепь Маркова является однородной и между любыми её двумя состояниями существует маршрут, то она **эргодична**. Кроме того, если вероятности состояний системы  $P_j$  не зависят от времени наблюдения за системой и совпадают с её начальными вероятностями  $P_j^{(0)}$  и стационарными вероятностями ( $P^{(0)} = \pi$ ), то **режим** цепи является **стационарным**. Таким образом, однородная цепь Маркова определится:

1.  $P^{(0)} = (P_1^{(0)}, \dots, P_n^{(0)})$  – начальным распределением вероятностей состояний;
2.  $\|\lambda_{ij}\|$  – матрицей интенсивностей простейших потоков, где  $\lambda_{ij} = P_{ij} \lambda_i$ ;
3.  $(1/\mu_1, 1/\mu_2, \dots, 1/\mu_n)$  – вектором экспоненциально распределённых времён пребывания в состояниях с параметром  $\mu_i, i = \overline{1, n}$ .

Если время пребывания процесса в каждом состоянии не распределено по экспоненте, то такой случайный процесс называется **полумарковским**. Обычно он задаётся следующими характеристиками:

- $P^{(0)} = (P_1^{(0)}, \dots, P_n^{(0)})$  – начальным распределением вероятностей состояний;
- $\|P_{ij}\|$  – матрицей переходных вероятностей;
- $f_1, \dots, f_n$  – совокупностью произвольных функций распределения времени пребывания системы в состояниях  $i, i = \overline{1, n}$ .

В моменты переходов из одного состояния в другое полумарковский процесс обладает марковскими свойствами. Если рассматривать полумарковский процесс только в моменты переходов, то получающаяся при этом марковская цепь переходов с дискретным временем называется вложенной марковской цепью, поскольку она содержится в полумарковском процессе.

### 3. МОДЕЛИ НАДЁЖНОСТИ ТЕХНИЧЕСКИХ СИСТЕМ

#### 3.1. Основные понятия теории надёжности

Надёжность сложной технической системы (СТС) определяется следующими факторами:

- мобильностью и динамичностью работы СТС;
- взаимосвязью и взаимодействием компонентов СТС;
- иерархичностью построения самих СТС;
- сложностью устранения последствий отказов компонентов СТС;
- наличием или отсутствием регламентации функционирования СТС;
- непрерывностью или дискретностью процессов в СТС;
- многомерностью СТС и сложностью управления СТС.

Кроме перечисленных факторов на функционирование СТС сильное влияние оказывают информационные потоки, циркулирующие в СТС. Поэтому важное значение для функционирования СТС имеет надёжность информационных систем. Сбои и отказы по информации зачастую приводят к непоправимым последствиям в работе СТС. Это приводит к необходимости использования специальных информационно-управляющих систем с высокой степенью надёжности их функционирования. Дополнительными факторами, влияющими на надёжность функционирования СТС, являются зависимость от действий оператора и необходимость выделения некоторого интервала времени, требуемого на принятие решений и выработку управляющих воздействий.

Исходя из вышеперечисленных факторов, воздействующих на СТС, определим понятие надёжности как некое свойство СТС сохранять во времени в установленных пределах значения параметров, характеризующих способность СТС выполнять свои функции в заданных режимах и условиях. Таковыми условиями являются особенности: применение СТС, техническое обслуживание СТС, наличие ремонтов компонентов СТС, возможность хранения и транспортировки её компонентов.

В теории надёжности транспортных систем (ТРС) исследователи обычно оперируют следующими понятиями:

1. **Безотказность** – свойство ТРС непрерывно сохранять работоспособное состояние в течение некоторого периода времени.
2. **Долговечность** – свойство компонента СТС сохранять работоспособность до наступления предельного состояния при установленной системе технического обслуживания (ТО) и ремонтов элементов СТС.
3. **Ремонтпригодность** – свойство СТС по поддержанию и восстанов-

лению работоспособного состояния за счёт ТО и ремонтов элементов СТС.

4. **Сохраняемость** – свойство СТС сохранять значения первых трёх показателей в течение некоторого периода после срока хранения и транспортировки.

5. **Состояние** СТС, выражающееся как исправное или неисправное по критериям требований конструкторской документации (КД). Причём неисправное состояние означает нарушение как минимум одного из требований КД. Исправное или работоспособное состояние означает соответствие параметров СТС требованиям КД. Различают ещё одно понятие состояния компонентов СТС – **предельное**, когда дальнейшее их применение недопустимо или невозможно восстановление их функций.

6. **Повреждение** компонентов СТС означает такое особое их состояние, когда нарушена исправность СТС, но сохранено работоспособное состояние СТС.

7. **Отказ** в СТС представляет собой такое событие, когда нарушено работоспособное состояние СТС.

Для оценки надёжности СТС вводят набор показателей надёжности. Обычно это несколько количественных характеристик одного или нескольких свойств, составляющих в совокупности надёжность компонентов СТС. Наиболее часто используется наработка – продолжительность или объём работы, выполненной компонентом или всей СТС. Эксплуатационники СТС чаще используют понятие **наработка на отказ**. Обычно это время от начала эксплуатации компонента СТС до возникновения первого отказа в его работе.

Вторым по важности критерием является **срок службы**, определяемый как продолжительность от начала эксплуатации компонента СТС до перехода его в предельное состояние. Наконец, важным показателем является **ёмкость** СТС, определяющая все остальные характеристики надёжности СТС.

Математически перечисленные показатели определяются следующим образом.

Свойство **безотказности** СТС характеризуется семейством показателей:

1. **Вероятность безотказной работы**  $Q(t)$  – вероятность того, что в пределах заданной наработки на отказ  $t_n$  отказ в СТС не возникнет:

$$Q(t) = \int_{t_n}^{\infty} f(x) dx, \quad (3.1)$$

где  $f(t)$  – функция плотности вероятностей отказа функционирования компоненты СТС.

Используя характеристику  $t_n$ , можно вычислить и вероятность отказа

СТС:

$$P(t) = \int_0^{t_n} f(x) dx. \quad (3.2)$$

Очевидно, что  $Q(t) = 1 - P(t)$ . Таким образом, зная одну из этих характеристик, можно вычислить другую.

2. **Средняя наработка СТС до отказа**  $T_{cp}$  – среднее время безотказной работы СТС:

$$T_{cp} = \int_0^{t_n} t \varphi(t) dt. \quad (3.3)$$

Как видим, это – математическое ожидание критерия наработки компоненты на отказ до появления отказа в момент времени  $t$ .

3. **Р-процентная наработка** до отказа  $T_p$  – величина наработки, в течение которой отказ в СТС не возникает с вероятностью  $P_1$ , выраженной в процентах. Определяется эта характеристика из уравнения

$$1 - \int_0^{T_p} f(t) dt = \frac{P}{100}. \quad (3.4)$$

4. **Средняя наработка на отказ**  $t_n/n_{отк}$  – отношение наработки восстанавливаемой компоненты к математическому ожиданию числа её отказов  $n_{отк}$  в течение этой наработки  $t_n$ .

5. **Вероятность отказа** в заданном интервале времени  $(t_1, t_2)$  –

$$P_{отк}(t_1, t_2) = Q(t_2) - Q(t_1). \quad (3.5)$$

6. **Вероятность безотказной работы в заданном интервале**  $(t_1, t_2)$  исходя из зависимости (3.5), определяется соотношением

$$P_{ботк}(t_1, t_2) = Q(t_1) - Q(t_2). \quad (3.6)$$

7. Частота появления отказов в интервале  $(t_1, t_2)$  (иногда её называют **интенсивностью отказа**) определяется как условная плотность вероятности возникновения отказа невозстанавливаемого компонента СТС для момента времени  $t$  и при условии, что отказ компоненты СТС ещё не возникал. Обозначается интенсивность отказа через  $B(t)$ .

Обычно различают следующие виды отказов: полный и частичный, внезапный и постепенный, независимый и зависимый, производственные, конструкционные, сбои, перемежающиеся, эксплуатационные. При полном отказе использование компонента СТС невозможно до его восстановления. При частичном отказе компонента СТС значения его параметров выходят за допустимые пределы. При внезапных отказах происходят скачкообразные

изменения нескольких основных параметров СТС. Производственные отказы возникают из-за нарушения технологического процесса при эксплуатации СТС или некачественного ремонта компонентов СТС. Конструктивные отказы являются следствием ошибок изготовления или компоновки компонентов СТС. При временной утрате работоспособности СТС фиксируют наличие сбоя. Это, как правило, самоустраняющиеся отказы. Если нарушаются установленные правила эксплуатации СТС, возникают эксплуатационные отказы. По природе возникновения все эксплуатационные отказы делятся на две группы: отказы технических средств и организационно-технологические отказы. Последние могут быть результатом износа, поломки или ухудшения параметров работы компонентов СТС. Перемежающиеся отказы представляют собой многократно возникающие сбои СТС одного и того же характера.

По последствиям различают три типа отказов:

- катастрофические (аварии, катастрофы, невыполнение ответственной работы);
- с экономическим ущербом (приводят к росту времени простоя СТС, снижаются технологические мощности, система работает с ухудшенными параметрами функционирования);
- без последствий (изменяются только затраты на ремонт, которые обычно колеблются в пределах установленных норм).

Отказы бывают ещё параметрические и функциональные. При параметрических отказах возможно некачественное выполнение функций системы. Функциональные отказы влекут за собой утрату СТС ряда важных своих функций.

Свойство долговечности основывается на понятии **ресурса**, имеющегося или необходимого для СТС. **Долговечность** СТС определяется следующей группой показателей:

1.  $\bar{R}$  – **средний ресурс** – оценка математического ожидания расхода ресурса изделия (компонента СТС).
2.  **$P\%$ -ресурс** – наработка на отказ  $t_{пр}$ , в течение которой компонент или СТС не достигнет предельного состояния с заданной вероятностью  $P$ , выраженной в процентах. Находится  $P\%$  из уравнения

$$1 - \int_0^{t_n} f(t) dt = \frac{P}{100}. \quad (3.7)$$

3. **Назначенный ресурс**  $t_{наз}$  – суммарное время наработки на отказ, при достижении которого применение компонента должно быть прекращено.
4. **Средний срок службы компонента СТС**  $t_{сл}$ , представляющий собой оценку математического ожидания срока эксплуатации компонента СТС.

5.  **$P$ -процентный срок службы** – календарная продолжительность использования компоненты СТС, фиксируемая от начала эксплуатации СТС, в течение которой она не достигнет предельного состояния с заданной вероятностью  $P$ , выраженной в процентах.

6. **Назначенный срок службы**  $t_{назсл}$  – календарная продолжительность эксплуатации СТС, при достижении которой применение компонента должно быть прекращено.

7. **Средний межремонтный срок службы**  $t_{мрсл}$ , представляющий собой оценку математического ожидания межремонтного срока службы данного типа компонентов СТС.

Свойство **ремонтпригодности** характеризуется следующей группой показателей:

1. **Вероятность восстановления** работоспособного состояния  $P_{вост}$  представляет собой вероятность того, что длительность периода восстановления объекта  $t_{вост}$  не превысит заданного значения, и определяется из выражения

$$P_{вост} = \int_0^{t_{вост}} f(\theta) d\theta, \quad (3.8)$$

где  $f(\theta)$  – плотность вероятности восстановления работоспособности компонентов СТС.

2. **Среднее время простоя системы**  $\bar{t}_{пр}$  из-за отказа компонента СТС.

3. **Среднее время восстановления**  $\bar{t}_{вост}$  работоспособного состояния СТС представляет собой оценку математического ожидания времени восстановления работоспособности СТС.

4. **Интенсивность восстановления**  $\lambda_{восст}$  определяет число восстановлений работоспособности СТС в единицу времени.

Свойство **сохраняемости** характеризуется следующей группой показателей:

1. **Средний срок сохраняемости** компоненты СТС  $\bar{t}_{сох}$  представляет собой оценку математического ожидания срока сохраняемости данного класса компонентов системы.

2.  **$P$ -процентный срок сохраняемости**  $t_{сох}^p$  – срок сохраняемости, достигаемый СТС с заданной вероятностью  $P$  в процентах.

3. **Интенсивность отказов при хранении**  $\lambda_{отхр}$  – количество отказов в единицу времени при хранении СТС.

Кроме перечисленных выше показателей различных свойств СТС на практике зачастую используются комплексные показатели надёжности СТС. В зависимости от вида и назначения СТС могут использовать самые различ-

ные показатели в качестве комплексных. Наиболее часто используются следующие коэффициенты:

1. **Коэффициент готовности**  $r$  – вероятность того, что СТС окажется в работоспособном состоянии в произвольный момент времени  $t$  (кроме планируемых периодов, в течение которых применение СТС по назначению не предусмотрено).

2. **Нестационарный коэффициент готовности**  $v_i$  – вероятность того, что компонент СТС будет работоспособен в заданный момент времени  $t_i$ .

3. **Средний коэффициент готовности**  $\bar{v}_\tau$  – усреднение на заданном интервале времени  $\tau$  значений нестационарных коэффициентов готовности СТС.

4. **Коэффициент оперативной готовности**  $K_{ог}$  – вероятность того, что компонент СТС, находясь в режиме ожидания, окажется работоспособным в произвольный момент времени  $t$  в установившемся процессе эксплуатации и, начиная с этого момента времени, будет работать безотказно в течение заданного интервала времени.

5. **Нестационарный коэффициент оперативной готовности**  $K_{ног\tau}$  – вероятность того, что объект, находясь в режиме оперативного ожидания, окажется работоспособным в заданный момент времени  $t$  и, начиная с этого момента, будет работать безотказно в течение определённого заранее интервала времени  $\tau$ .

6. **Коэффициент технического использования**  $K_{ти}$  компонента СТС – отношение математического ожидания интервалов времён пребывания её в работоспособном состоянии за некоторый период эксплуатации  $T$  к сумме математических ожиданий интервалов времён пребывания СТС во всех состояниях (работоспособном, простоях из-за технического обслуживания и ремонтов) за тот же период эксплуатации.

7. **Коэффициент планируемого применения**  $K_{пп}$  – доля периода эксплуатации, в течение которой СТС не должна находиться на плановом техническом обслуживании и ремонте.

8. **Коэффициент сохранения эффективности СТС**  $K_{эфj}$  – показатель отношения эффективности СТС за определённую ( $j$ -ю) продолжительность эксплуатации к номинальному значению этого показателя, вычисленному при условии, что отказы СТС в течение этого периода эксплуатации не возникают.

Вторым важнейшим понятием теории надёжности является **резервирование**. Оно включает такие понятия, как избыточность СТС и разные виды резервирования.

Понятие **избыточности** означает наличие в СТС дополнительных компонентов или использование дополнительных возможностей у существующих компонентов сверх минимально необходимых для выполнения СТС

заданных функций.

Понятие **резервирование** используется при описании способа повышения надёжности СТС путём введения избыточности в составе компонентов СТС. Различают следующие виды резервирования: временное, структурное, скользящее, смешанное.

При **временном** резервировании устанавливают резерв времени работ из-за отказов  $T_{pi}$ , время для проведения профилактических работ  $T_0$  коэффициенты резерва (общий коэффициент резерва)  $\beta = \frac{T_0 + T_1 + T_2}{T}$ , где  $T$  – расчёт-

ный период работы СТС;  $T_1$  – резерв времени для компенсации отказов компонентов СТС;  $T_2$  – резерв времени на рост объёмов работ из-за отказов СТС.

При **структурном** резервировании вводится избыточность структурных элементов на случай выхода из строя основных элементов СТС.

При **скользящем** резервировании используется группа основных элементов, которые являются резервом для всех остальных элементов СТС.

Наконец, **смешанное** резервирование является комбинацией первых трёх типов резервирования.

Коэффициенты готовности СТС и резервирования находятся в следующей зависимости:

$$1. r_0 = \frac{T - T_0}{T} = 1 - \beta_0. \quad (3.9)$$

Это доля времени, в течение которого в системе не ведётся профилактика и проводятся восстановительные работы.

$$2. r_1 = \frac{T - (T_0 + T_1)}{T - T_0} = 1 - \beta_1. \quad (3.10)$$

Это доля времени, используемая СТС за вычетом потерь на восстановление работоспособности, связанных с ликвидацией последствий отказов.

$$3. r_2 = \frac{T - (T_0 + T_1 + T_2)}{T - (T_0 + T_1)} = 1 - \beta_2 - \quad (3.11)$$

означает готовность СТС к обеспечению её функций на ближайшую перспективу.

Итак, общий коэффициент готовности к безотказной работе и общий коэффициент резерва связаны следующей зависимостью:

$$r = \frac{T - (T_0 + T_1 + T_2)}{T} = r_0 r_1 r_2 = (1 - \beta_0)(1 - \beta_1)(1 - \beta_2) = 1 - \beta. \quad (3.12)$$

Отметим, что приведённые коэффициенты резерва и готовности на практике нашли более широкое применение по сравнению с вычислением других характеристик надёжности СТС. Именно простота и лёгкость вычислений явились причиной такого положения дел. Однако приближённый характер самих коэффициентов ставит ряд проблем, связанных с нахождением самих времён  $T_i$ .

### 3.2. Надёжность элемента системы, работающего до первого отказа

Под элементом системы обычно понимают любое устройство, надёжность работы которого изучается независимо от надёжности составляющих его частей. Предполагается, что в начальный момент времени  $t=0$  элемент начинает работу, а в момент  $t=\tau$  происходит отказ его работоспособности. Здесь  $\tau$  – время жизни элемента, представляющее собой случайную величину с заданным законом распределения:

$$F(t) = P\{\tau < t\}. \quad (3.13)$$

Эта функция определяет вероятность отказа элемента до момента времени  $t$ . Предполагают, что  $F(t)$  непрерывна и существует непрерывная функция плотности вероятности отказа  $f(t)$ . Эта функция полностью определяет надёжность элемента. Наряду с ней часто используется другая функция

$$P(t) = 1 - F(t) = P\{\tau > t\}, \quad (3.14)$$

представляющая собой вероятность безотказной работы элемента за время  $t$ . Её называют *функцией надёжности*.

Во многих случаях надёжность элемента пытаются определить некоторыми числовыми величинами:

1. *Среднее время безотрывной работы*  $T_0$ , являющееся математическим ожиданием случайной величины  $\tau$  и определяемое из выражения

$$T_0 = M_\tau = \int_0^\infty f(t)dt = -tP(t)|_0^\infty + \int_0^\infty P(t)dt. \quad (3.15)$$

2. *Дисперсия времени жизни элемента*  $D_\tau$ ,

$$D_\tau = M(\tau - T_0)^2 = M\tau^2 - (M\tau)^2 = 2 \int_0^\infty tP(t)dt - T_0^2. \quad (3.16)$$

Среднеквадратичное отклонение  $\tau$  от своего среднего  $\sigma_\tau = \sqrt{D_\tau}$  используется в тех случаях, когда необходимо отслеживать небольшой разброс величины  $\tau$  (т. е.  $\sigma_\tau < T_0$ ).

Пусть элемент проработал безотказно до момента  $t$ . Необходимо определить вероятность того, что он не откажет на временном интервале  $(t, t_1)$ . Обозначим её как  $P(t, t_1)$ . Пусть  $A$  – событие, означающее безотказную работу элемента на интервале времени  $(0, t)$ , а  $B$  – событие, означающее безотказную работу элемента на интервале времени  $(t, t_1)$ . Тогда вероятность  $P(t, t_1)$  есть условная вероятность события  $A/B$ :

$$P(t, t_1) = P\{A/B\} = \frac{P(AB)}{P\{A\}}. \quad (3.17)$$

Поскольку событие  $(AB)$  означает безотказную работу элемента на интервале  $(0, t_1)$ , то

$$P(t_0, t_1) = \frac{P(t_1)}{P(t)}. \quad (3.18)$$

Таким образом, вероятность отказа элемента на интервале  $(t, t_1)$

$$F(t, t_1) = 1 - P(t, t_1) = \frac{P(t) - P(t_1)}{P(t)}. \quad (3.19)$$

Допустим,  $t_1 = t + \Delta t$  при  $\Delta t \rightarrow \infty$ . Тогда

$$F(t, t + \Delta t) = \frac{P(t) - P(t + \Delta t)}{P(t)} = \frac{P'(t)}{P(t)} \Delta t + O(\Delta t). \quad (3.20)$$

При малом значении  $\Delta t$  имеет место равенство  $F(t, t + \Delta t) = \lambda(t)\Delta t$ .

Величина  $\lambda(t)$  является локальной характеристикой надёжности элемента в каждый данный момент времени. Итак,  $\lambda(t)$  – вероятность того, что элемент, проработавший безотказно до момента  $t$ , в последующую достаточно малую единицу времени  $\Delta t$  откажет в работе. Как видим  $\lambda(t)$  представляет собой плотность условной вероятности отказа в момент  $t$ , предполагая, что до этого момента элемент работал безотказно. Функцию  $\lambda(t)$  называют опасностью отказа. Уравнение (3.20) может быть разрешено относительно функции надёжности  $P(t)$ :

$$P(t) = e^{-\int_0^t \lambda(t)dt}. \quad (3.21)$$

Таким образом, вероятность безотказной работы элемента на интервале  $(t_1, t_2)$  определяется из выражения

$$P(t_1, t_2) = e^{-\int_{t_1}^{t_2} \lambda(t)dt}. \quad (3.22)$$

Если  $\Delta n$  – число отказов элемента на интервале  $(t, t + \Delta t)$ ,  $n(t)$  – число элементов данного типа, не отказавших к моменту  $t$ , то можно эксперименталь-

но определить **опасность отказа** из выражения  $\lambda(t) \approx \Delta n / \Delta t n(t)$ .

Для многих элементов СТС функция  $\lambda(t)$  имеет вид, представленный на рис. 3.1. Как видим, для многих случаев можно принять, что  $\lambda(t) = \lambda = \text{const}$ .

Из зависимости (3.21) следует, что для постоянной опасности отказов функция надёжности представляет собой экспоненту

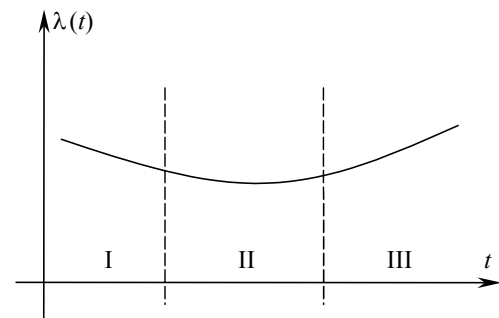


Рис. 3.1. Вид функции **опасность отказа** элементов СТС

$$P(t) = e^{-\lambda t}. \quad (3.23)$$

Тогда функции вероятности отказа и плотности вероятности отказов представляют собой:

$$F(t) = 1 - e^{-\lambda t}; \quad f(t) = \lambda e^{-\lambda t}. \quad (3.24)$$

Среднее время жизни элемента

$$T_0 = \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}, \quad (3.25)$$

т. е. опасность отказа для экспоненциального закона обратно пропорциональна среднему времени жизни элемента

$$\lambda = 1 / T_0. \quad (3.26)$$

Поэтому функцию надёжности можно записать в виде

$$P(t) = e^{-t/T_0}. \quad (3.27)$$

Обычно  $t \ll T_0$ , тогда  $t/T_0 \ll 1$ , и можно пользоваться простыми приближёнными формулами

$$P(t) \approx 1 - t/T_0; \quad F(t) \approx t/T_0. \quad (3.28)$$

Ошибка при таких расчётах не превосходит величины  $0,5(t/T_0)^2$ . Заметим, что это нижние оценки вероятностей, поскольку  $P(t) \geq 1 - t/T_0$ ,  $F(t) \leq t/T_0$  и  $(1-x) \leq e^{-x}$  (все точки линии  $(1-x)$  лежат ниже экспоненты  $e^{-x}$ ).

Экспоненциальный закон распределения очень популярен в теории надёжности. Он естественен, прост и удобен в расчётах. Важное свойство это-

го закона состоит в том, что вероятность безотказной работы на интервале  $(t, t + \tau)$  не зависит от времени предшествующей работы элемента  $t$  и зависит только от длины интервала  $\tau$ . Это означает, что будущее поведение элемента не зависит от прошлого (нет эффекта последействия). Известно, что внезапные отказы, носящие случайный характер, хорошо описываются экспоненциальным законом распределения. Постепенные отказы же, возникающие из-за износа или старения, обычно описываются нормальным законом распределения вероятностей.

Для ряда случаев у элементов часто совмещаются оба типа отказов (внезапный и постепенный). Такие элементы можно рассматривать состоящими из двух частей: в одной происходит только внезапный отказ, а в другой формируется постепенный отказ. Функцию надёжности в этом случае можно представить в виде

$$P(t) = P_1(t)P_2(t) = e^{-\lambda t} \frac{1}{\sqrt{2\pi}} \int_{\frac{t-t_0}{\sigma}}^{\infty} e^{-\frac{x^2}{2}} dx. \quad (3.29)$$

Предположив, что  $\sigma \ll T_0$ , а  $T_0$  и  $1/\lambda$  одного порядка, можно установить, что среднее время жизни элемента

$$T_{\text{ср}} \approx \frac{1}{\lambda} (1 - e^{-\lambda T_0 + \frac{\lambda^2 \sigma^2}{2}}). \quad (3.30)$$

Для нахождения точного распределения можно использовать таблицы нормального закона распределения [1].

В более сложных случаях вместо экспоненциального закона используются распределение Вейбулла и логнормальное распределение. Если функция надёжности имеет распределение Вейбулла  $P(t) = e^{-\lambda t^2}$ , то основные числовые характеристики надёжности будут иметь вид

$$T_0 = \int_0^{\infty} e^{-\lambda t^2} dt = \frac{\Gamma(\frac{1}{\alpha} + 1)}{\lambda^{1/\alpha}}; \quad (3.31)$$

$$D\tau = \sigma^2 = \frac{\Gamma(\frac{2}{\alpha} + 1) - \Gamma^2(\frac{1}{\alpha} + 1)}{\lambda^{2/\alpha}}; \quad (3.32)$$

$$\lambda(t) = \alpha \lambda t^{\alpha-1}. \quad (3.33)$$

Закон Вейбулла обобщает экспоненту и при этом содержит параметр  $\alpha$ . На рис. 3.2 представлены графики плотности  $\lambda(t)$  при разных значениях  $\alpha$ .

Подбирая опытным путём  $\lambda$  и  $\alpha$ , можно получить наилучшее приближение к практике. Если имеют место скрытые дефекты, и нет долго старения,

то  $\alpha < 1$ . Если нет скрытых дефектов, но элемент быстро стареет, то опасность отказа быстро растёт ( $\alpha > 2$ ). При постепенном отказе используется логнормальное распределение для функции надёжности

$$P(t) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\ln(t/T_0)}{\sigma}}^{\infty} e^{-\frac{x^2}{2}} dx; \quad (3.34)$$

$$T_{cp} = T_0 e^{\frac{\sigma^2}{2}}. \quad (3.35)$$

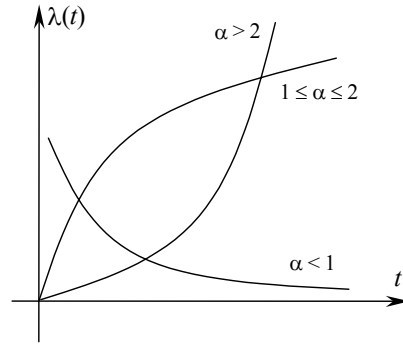


Рис. 3.2. Вид функции опасности отказа при разных значениях параметра  $\alpha$

В ряде случаев функция надёжности имеет степенное распределение. Эта функция всегда конечна. Через неё выражаются все основные характеристики нашего процесса:

$$D_v(t) = 2 \int_0^t H(t-\tau) dH(\tau) + H(t) - H^2(t). \quad (3.36)$$

На интервале времен  $(t_1, t_2)$  среднее число отказов

$$\bar{v}(t_1, t_2) = H(t_1) - H(t_2). \quad (3.37)$$

Иногда вместо функции восстановления используют дифференциальные характеристики потока отказов, плотность восстановления или интенсивность отказов. Обычно это среднее число отказов, происшедших за единицу времени в момент времени  $t$ .

При экспоненциальном законе распределения имеем следующие соотношения:

$$F(t) = 1 - e^{-\lambda t}; \quad P_n(t) = P\{v(t)=n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}; \quad (3.38)$$

$$H(t) = \lambda t; \quad h(t) = \lambda.$$

Для нормального закона распределения при  $\sigma \ll T_0$  эти же соотношения имеют вид

$$F_n(t) = \Phi\left(\frac{t - nT_0}{\sigma\sqrt{n}}\right); \quad \Phi(x) = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz; \quad H(t) = \sum_{n=1}^{\infty} \Phi\left(\frac{t - nT_0}{\sigma\sqrt{n}}\right); \quad (3.39)$$

$$P(t) = 1 / (1 + \frac{1}{T_0})^2. \quad (3.40)$$

Тогда при  $\alpha > 1$  (при преобладающем старении)  $T_{cp} = T_0 / (\alpha - 1)$ .

### 3.3. Надёжность восстанавливаемого элемента

Предполагается, что после отказа элемент восстанавливается (либо немедленно, либо через некоторый интервал времени). При этом либо элемент заменяется новым, либо ремонтируется. В первом случае время восстановления пренебрежимо мало по сравнению с временем жизни элемента ( $t_{восст} \ll t_{жиз}$ ), поскольку элемент заменяется немедленно. Допустим, что случайные времена отказов  $\tau_1, \tau_2, \dots$  независимы и имеют один и тот же закон распределения  $F(t) = P(\tau_n < t)$ .

Это означает, что моменты отказов (а значит, и восстановлений) образуют случайный поток, называемый в дальнейшем **потоком восстановления**. Среднее время жизни его  $T_0$  и дисперсия конечны и вычисляются по формулам

$$T_0 = M\tau_n = \int_0^{\infty} [1 - F(t)] dt; \quad \sigma^2 = D\tau_n = 2 \int_0^{\infty} t[1 - F(t)] dt - T_0^2. \quad (3.41)$$

**Число отказов**  $v(t)$  является базовой характеристикой, которая может принимать только целые неотрицательные значения ( $v(t) > 0$ ). Для нахождения среднего числа отказов используется функция восстановления  $H(t)$ , вычисляемая по формуле

$$H(t) = Mv_t = \sum_{n=1}^{\infty} n[F_n(t) - F_{n+1}(t)] = \sum_{n=1}^{\infty} F_n(t); \quad (3.42)$$

$$h(t) = \sum_{n=1}^{\infty} \frac{1}{\sqrt{2\pi n\sigma}} e^{-\frac{(t-n\tau_0)^2}{2\sigma^2 n}}. \quad (3.43)$$

Во втором случае имеет место процесс восстановления с конечным временем восстановления отказавшего элемента. В реальности любое восстановление элемента требует времени как для поиска вида и места неисправности в элементе, так и для ремонта. Обозначим времена безотказной работы и восстановления  $i$ -го элемента соответственно через  $\tau_i'$  и  $\tau_i''$ . Пусть моменты отказов определяются по формуле

$$t_n' = \tau_1' + \tau_1'' + \tau_2' + \tau_2'' + \dots. \quad (3.44)$$

Следовательно, моменты восстановления

$$t_n'' = \tau_1'' + \tau_2' + \tau_2'' + \tau_3' + \tau_3'' + \dots. \quad (3.45)$$

Предполагаем, что  $\tau_i'$  и  $\tau_i''$  независимы. Предполагаем также, что  $\tau_i'$  распределены по одному закону  $F(t) = P\{\tau_i' < t\}$  с параметрами  $\tau_i = M\tau_n'$ ,  $\sigma_1^2 = D\tau_n'$ , а все периоды восстановления также распределены одинаково –  $G(t) = P\{\tau_i'' < t\}$  с параметрами  $\tau_2 = M\tau_n''$ ,  $\sigma_2^2 = D\tau_n''$ . Пусть для простоты эти законы распределения имеют непрерывные функции плотности:  $f(t) = F'(t)$ ;

$$g(t) = G'(t); \sigma_1^2 + \sigma_2^2 \neq 0.$$

Тогда такой процесс назовём **процессом восстановления с конечным временем восстановления**. Основной характеристикой этого процесса является **коэффициент готовности** элемента  $K_r$ , который вычисляется следующим образом. Рассмотрим последовательность событий  $A_n = \{t_n'' < t_{n+1}''\}$ . Предполагается, что до момента  $t$  произошло ровно  $n$  восстановлений, элемент исправен в момент  $t$  и определяется следующим образом:

$$B = \bigcup_{n=0}^{\infty} A_n; \quad P(B) \sum_{n=0}^{\infty} P\{A_n\} = K_r. \quad (3.46)$$

На практике среднее значение коэффициента готовности элемента определяется как доля времени, в котором система пребывала в исправном состоянии:

$$K_r = T_1 / (T_1 + T_2), \quad (3.47)$$

где  $T_1$  и  $T_2$  – соответственно суммарные длительности нахождения элемента в исправном и неисправном состояниях. Обозначим через  $v_1(t)$  число отказов элемента на интервале  $(0, t)$ ;  $v_2(t)$  – число восстановлений элемента на том же интервале. Соответствующие математические ожидания (если их оценки – средние величины) значений  $v_1(t)$  и  $v_2(t)$ :

$$Mv_1(t) = H_1(t); \quad Mv_2(t) = H_2(t). \quad (3.48)$$

Поскольку нас интересует в основном большие значения  $t$  и  $v_1(t)$ ,  $v_2(t)$  асимптотически ведут себя одинаково, отличаясь друг от друга максимум на 1, то будем рассматривать только моменты восстановления. Эта характеристика представляет собой сумму времен:  $\tau_n = \tau_n' + \tau_n''$ . И эта характеристика тоже случайная величина, имеющая распределение вида

$$P\{\tau_n < t\} = \Phi(t) = \int_0^t F(t-x) dG(x). \quad (3.49)$$

Таким образом, функция восстановления и плотность восстановления определяются по формулам:  $H_2(t) = \sum_{n=1}^{\infty} \Phi_n(t)$ ;  $h_2(t) = \sum_{n=1}^{\infty} \varphi_n(t)$ .

Очевидно, что  $h_2(t) \rightarrow t / (T_1 + T_2)$  (среднему числу восстановлений в единицу времени). Отметим, что величина  $v_2$  при  $t \rightarrow \infty$  асимптотически имеет нормальное распределение с параметрами

$$\left( \frac{t}{T_1 + T_2}; \frac{(\sigma_1^2 + \sigma_2^2)t}{(T_1 + T_2)^2} \right). \quad (3.50)$$

Аналогично для  $v_1$  при  $t \rightarrow \infty$  имеем такую же асимптоту, и вероятность безотказной работы на интервале  $(t, t + \tau)$  определяется по формуле

$$P(\tau) = \underbrace{K_r}_{P_1} \underbrace{\frac{1}{T} \int_{\tau}^{\infty} [1 - F(x)] dx}_{P_2}. \quad (3.51)$$

Общий вывод: вероятность безотказной работы элемента в течение времени  $\tau$  равна произведению вероятности того, что в начальный момент времени мы застаём элемент исправным ( $P_1$ ) на вероятность исправной работы элемента в течение  $\tau$ .

На практике используют статистические подходы к вычислению критериев надёжности восстанавливаемых систем. В качестве примера рассмотрим следующую модель испытаний элемента. Пусть на испытании находится  $N$  элементов. При отказах элементы заменяются новыми. Испытания завершаются, когда число отказов статистически достаточно для оценки надёжности изделия с заданной доверительной вероятностью  $\beta$ . Без учёта  $t_{\text{восст}}$  элементов количественными характеристиками надёжности элементов можно считать:

$\omega(t)$  – параметр потока отказов (восстановлений);

$t_{\text{cp}}$  – наработка на отказ (т.е. интервал между двумя соседними отказами).

**Параметр потока отказов** определяется следующим образом: пусть  $n(\Delta t)$  – число отказавших элементов в интервале времени  $(t - \frac{\Delta t}{2}, t + \frac{\Delta t}{2})$ , а  $N$  – число испытанных элементов. Тогда параметр потока отказов

$$\bar{\omega}(t) = \frac{n(\Delta t)}{N \Delta t}. \quad (3.52)$$

Параметр потока отказов обладает следующими свойствами:

1. Для любого  $t$  независимо от закона распределения времени безотказной работы параметр потока отказов выше, чем частота отказов ( $\bar{\omega}(t) > a(t)$ ).

2. Независимо от функции распределения частоты отказов  $a(t)$  параметр потока отказов при  $t \rightarrow \infty$  асимптотически стремится к средней интенсивности отказов  $\bar{\omega}(t) \rightarrow 1 / T_{\text{cp}}$ .

Итак, при длительной эксплуатации ремонтируемого элемента поток его отказов становится стационарным.

3. Если  $\lambda(t)$  растёт во времени, то имеет место соотношение

$$\lambda(t) > \bar{\omega}(t) > a(t). \quad (3.53)$$

При убывании  $\lambda(t)$  во времени соотношение между этими характеристиками меняется:



$$\omega(t) > \lambda(t) > a(t). \quad (3.54)$$

4. При постоянном значении интенсивности отказов ( $\lambda(t) = \text{const}$ ) параметр потока отказов системы элементов не равен сумме значений параметров потока отказов элементов:

$$\omega_c(t) \neq \sum_{i=1}^N \omega_i(t), \quad (3.55)$$

т. е. при вычислении количественных характеристик суммы элементов (в виде системы) нельзя суммировать интенсивности отказа элементов.

5. При  $\lambda(t) = \text{const}$   $\omega(t) = \lambda(t) = \lambda$ , т. е. параметр потока отказов соответствует интенсивности отказов системы элементов.

Чаще всего в инженерной практике используется следующая методика нахождения характеристик  $\lambda(t)$  и  $\omega(t)$ .

1. По формуле (3.52) вычисляют значение потока отказов  $\omega(t)$  для каждой серии испытаний элементов. По каждой серии испытаний формируется гистограмма распределения  $\omega_i(t)$ .

2. Гистограмма заменяется кривой, которая описывается уравнением Вольтера:

$$\omega(t) = a(t) + \omega(t)a(t - \tau)d\tau. \quad (3.56)$$

3. Далее осуществляют преобразование Лапласа  $\omega_i(s)$  для функции  $\omega_i(t)$ :

$$\omega(s) = \frac{a(s)}{1 - a(s)}; \quad a(s) = \frac{\omega(s)}{1 + \omega(s)}. \quad (3.57)$$

4. По известной функции  $\omega_i(s)$  по формуле (3.57) вычисляют преобразование Лапласа  $a_i(s)$  для частоты отказов  $u_i(t)$ .

5. Вычисляют аналитическое выражение интенсивности отказов по формуле

$$\lambda_i(t) = \frac{a_i(t)}{\left(1 - \int_0^t a_i(t) dt\right)}. \quad (3.58)$$

6. Строят график  $\lambda(t)$  на всех  $i$ -х участках временной оси.

7. Если находится участок, где  $\lambda_i(t) = \text{const} = \lambda_i$ , то это значение и является оценкой вероятности безотказной работы элемента.

Наработка на отказ определяется по формуле

$$t_{\text{нот}} = \left(\sum_{i=1}^n t_i\right) n, \quad (3.59)$$

где  $t_i$  – время исправной работы элемента между  $(i-1)$ -м и  $i$ -м отказами;  $n$  – число отказов элемента за время  $t$ .

Как видим, наработка на отказ определяется по данным испытаний одного образца элемента. Если на испытании находятся  $N$  элементов в течение времени  $t$ , то наработка на отказ вычисляется по формуле

$$t_{\text{нот}} = \left(\sum_{j=1}^N \sum_{i=1}^n t_{ij}\right) / \sum_{j=1}^N n_j, \quad (3.60)$$

где  $t_{ij}$  – время исправной работы  $j$ -го образца между  $(i-1)$ -м и  $i$ -м отказами;  $n_j$  – число отказов за время  $t$  работы  $j$ -го элемента.

Параметры потока отказов  $w(t)$  и наработка на отказ  $t_{\text{нот}}$  характеризуют надёжность ремонтируемого элемента и не учитывают  $t_{\text{вост}}$ . Поэтому они не характеризуют готовности элемента к выполнению своих функций в нужное время  $t$ . Для этой цели вводятся такие критерии, как коэффициент готовности  $K_r$  и коэффициент вынужденного простоя элемента  $K_n$ . Расчёт этих характеристик ведут по следующим формулам:

$$K_r = t_p / (t_p + t_n), \quad (3.61)$$

$$K_n = t_n / (t_p + t_n). \quad (3.62)$$

где  $t_p = \sum_{i=1}^n t_{p_i}$  – суммарное время исправной работы элемента,  $t_n = \sum_{i=1}^n t_{n_i}$  – суммарное время простоя элемента,  $n$  – число отказов;  $i$  – номер отказа.

Очевидна связь между этими характеристиками:  $K_n = 1 - K_r$ .

При анализе надёжности восстанавливаемых систем коэффициент готовности рассчитывается по формуле

$$K_r = T_{\text{cp}} / (T_{\text{cp}} + t_{\text{вост}}), \quad (3.63)$$

где  $T_{\text{cp}}$  – среднее время безотказной работы элемента;  $t_{\text{вост}}$  – среднее время восстановления элемента.

Если интенсивности отказов  $\lambda$  и восстановления  $\mu$  элемента есть величины постоянные, и система при  $t=0$  находится в исправном состоянии ( $P(0)=1$ ), то вероятность готовности системы к функционированию  $P_r(t)$  определяется из выражений

$$P_r(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t}; \quad (3.64)$$

$$P_r(t) = K_r + (1 - K_r) e^{-t/(K_r t_b)}, \quad (3.65)$$

где  $\lambda = \frac{1}{T_{\text{cp}}}$ ;  $\mu = \frac{1}{t_{\text{вост}}}$ ;  $K_r = \frac{T_{\text{cp}}}{T_{\text{cp}} + t_{\text{вост}}}$ .

Установлено, что  $P_r(t) \rightarrow K_r$  при  $t \rightarrow \infty$ . Как видим, коэффициент готовности представляет собой вероятность того, что элемент будет в исправном состоянии при установившемся процессе его эксплуатации.

### 3.4. Меры надёжности транспортных систем

Отказ в работе всех типов станций, узлов, участков и других транспортных систем приводит к задержке поездов и вагонов и поэтому они необратимы. Основной количественной мерой их работоспособности и безотказности таких систем служит интервал времени между концом восстановления  $i$ -й компоненты транспортной системы (ТС) и появлением следующего отказа за  $t_{\text{бор}}$ . В общем случае  $t_{\text{бор}}$  есть случайная величина. Если физические, технические, технологические и информационно-управляющие качества элементов ТС и самой ТС в целом остаются неизменными, а нагрузка на систему (величина транспортного потока) с течением времени растёт, то функция надёжности  $P(t)$  обладает следующими свойствами.

1.  $P(t)$  является монотонно убывающей функцией времени  $t$ , если по оси  $t$  ставится в соответствие непрерывно возрастающая нагрузка.

2. Если нагрузка носит дискретный характер, то  $P(t_0)$  будет дискретно убывающей.

3. Для станций, узлов и других систем при  $t=0$  (в начальный момент времени) справедливо неравенство  $0 \leq P(0) \leq 1$ .

Свойства функции отказов  $F(t_0)$  носят противоположный характер из-за того, что  $F(t_0) = 1 - P(t_0)$ .

Статистическое определение функции отказов имеет вид

$$F(t_0) = n(t_0)/N(t_0), \quad (3.66)$$

где  $n(t_0)$  – число задержанных поездов;  $N(t_0)$  – общее число поездов.

Для объектов, свойства которых со временем ухудшаются, функция отказов имеет тот же вид. При этом под  $n(t_0)$  понимают число отказанных объектов в момент  $t_0$ , а  $N(t_0)$  – число исправных объектов. Если эксплуатационные характеристики ухудшаются со временем, то вероятность отказа ТС определяется по формуле

$$F(t, t+t_0) = \frac{n(t+t_0) - n(t_0)}{N(t)} = \frac{\Delta n(t, t_0)}{N(t)}, \quad (3.67)$$

где  $\Delta n(t_0)$  – приращение отказов от  $t_0$  до  $t$ ;  $N(t)$  – число исправных объектов на момент  $t$ .

При этом

$$\Delta n(t, t_0) = N(t) - N(t+t_0); \quad n(t+t_0) - n(t) = \Delta N(t, t_0). \quad (3.68)$$

На рис. 3.3 представлена геометрическая интерпретация функций надёж-

ности и отказов.

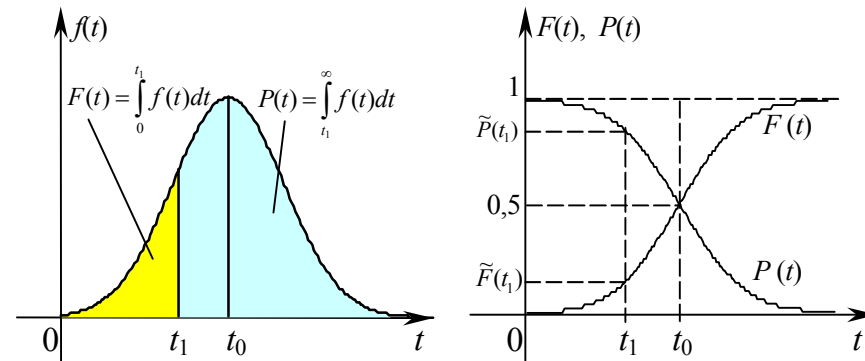


Рис. 3.3. Геометрическая интерпретация функций  $P(t)$  и  $F(t)$

Статистически плотность отказов определяется из соотношения

$$\tilde{f}(t) = \frac{\Delta n(t, \Delta t)}{N(t + \Delta t)\Delta t}, \quad (3.69)$$

где  $N(t + \Delta t)$  – число безотказно отработавших поездов на конец периода  $(t + \Delta t)$ ;  $\tilde{f}(t)$  – частота отказов в интервале  $(t, t + \Delta t)$ .

Как видим, плотность распределения определяется как отношение числа задержанных поездов за  $\Delta t$  к произведению числа поездов, пропущенных к концу периода  $(t + \Delta t)$  на длительность интервала  $\Delta t$ . Статистическое определение интенсивности отказов для систем, пропускающих транспортный поток с трансформацией и отказы которых выражаются через задержки поездов, практически совпадает со статистическим определением плотности распределения отказов, т. е.

$$\tilde{\lambda}(t) = \frac{\Delta n(t, \Delta t)}{N(t + \Delta t)\Delta t}. \quad (3.70)$$

Для транспортных систем поток отказов может быть однозначно задан либо статистической выборкой об отказах в периоде времени  $T$ , либо законом распределения потока отказов (распределение интенсивности отказов в этом периоде).

Вспользуемся традиционным подходом в теории надёжности: представлением случайного потока отказов как случайного процесса в реальном масштабе времени, т. е. чередование интервалов  $\tau_{\text{работ}}$  и  $\tau_{\text{вост}}$  (или  $\tau_{\text{простоя}}$ ). Под восстановлением понимают не только ремонт, но и замену оборудования в вагонах, устранение коммерческих неисправностей, освобождение путей подвода локомотивов для отправления скопившихся поездов. Извест-

но, что среднее время работы транспортной системы до появления отказа можно вычислить по формуле

$$T_p = \int_0^{\infty} P(x) dx, \quad (3.71)$$

При статистическом анализе  $\tilde{T}_p$  определяют по начальному циклу объектов, работоспособность которых анализируется:

$$\tilde{T}_p = \frac{1}{n} \sum_{i=1}^n \theta_i, \quad (3.72)$$

где  $n$  – число отказов системы;  $\theta_i$  – длительность работы  $i$ -го объекта до его отказа.

Среднее время простоя и восстановления безотказной работы системы

$$\tilde{T}_{\text{прост}} = \frac{1}{n} \sum_{i=1}^n t_{\text{прост}_i}, \quad (3.73)$$

где  $t_{\text{прост}_i}$  – длительность восстановления и простоя  $i$ -го объекта после отказа. Интенсивность восстановления безотказной работы в момент  $t$  определяется по формуле

$$\mu(t) = \frac{\Delta n_3(t, t + \Delta t)}{n(t) \Delta t}. \quad (3.74)$$

где  $\mu(t)$  – отношение числа задержанных, но уже принятых поездов к произведению числа задержанных, но не принятых поездов на длительность  $\Delta t$ ;  $n(t)$  – число задержанных поездов к моменту времени  $t$ ;  $\Delta n_3(t, t + \Delta t)$  – число задержанных поездов на интервале  $(t, t + \Delta t)$ .

Для транспортных систем имеет место ещё одна особенность: время простоя и восстановления безотказности ( $t_{\text{вос}}$ ) не равно времени задержки и простоя самого транспортного средства. Поэтому зачастую приходится определять производные характеристики последствий из-за отказов. Например, поездо-часы

$$n_t = \sum_{i=1}^n t_{\text{зди}}. \quad (3.75)$$

Как видим, количественные характеристики восстанавливаемых систем принципиально не отличаются от показателей надёжности невосстанавливаемых систем. Например, вероятность безотказной работы имеет тот же смысл:  $P(t_0) = 1 - F(t_0)$ . Она означает возможность безотказной работы элемента в течение заданного времени  $t_0$  с момента окончания  $(n - 1)$  восстанов-

ления:

$$P_n(t_0) = P \left\{ t + t_0 \leq \sum_{i=1}^{n-1} (\theta_i - t_{\text{вос}i}) + \theta_n / \sum_{i=1}^{n-1} (\theta_i - t_{\%oi}) = t_0 \right\} = \\ = P \{ \theta_n \geq t_0 \} = 1 - F_n(t_0). \quad (3.76)$$

Статистическое определение этой характеристики применительно к транспортным системам имеет вид

$$\tilde{P}(t, t + t_0) = \frac{N(t + t_0)}{N(t) + N(t_0)}, \quad (3.77)$$

где  $N(t, t_0)$  – число поездов, не имеющих отказов к моменту времени  $(t, t_0)$ ;  $N(t)$  – число поездов, не имевших отказов к моменту  $t$ ;  $N(t_0)$  – общее число принятых станций и отправленных поездов за период  $(t, t_0)$ .

Широкое распространение в станционной практике получило использование коэффициента оперативной готовности  $R(t_0)$ , вычисляемое по формуле

$$R(t_0) = \frac{T_B}{T_p + T_B}, \quad (3.78)$$

где  $T_B$  – математическое ожидание  $t_{\text{вос}}$ ;  $T_p$  – математическое ожидание времени безотказной работы станции.

Статистическое определение функций отказов для восстанавливаемых систем осуществляется по формулам:

$$\tilde{F}(t_0) = \frac{n_K(t_0)}{N(t_0)} = 1 - \frac{N_B(t_0)}{N(t_0)}, \quad (3.79)$$

где  $n_K(t_0)$  – число поездов, задержанных в связи с отказами работы станции;  $N(t_0)$  – общее число поездов, пропущенных через станцию за период  $t_0$ ;  $N_B(t_0)$  – число безотказно пропущенных поездов.

Коэффициент готовности станции к безотказной работе статистически определяется следующим образом:

$$K(t) = \frac{N_B(t)}{N(t)} = 1 - \frac{n(t)}{N(t)}, \quad (3.80)$$

где  $N_B(t)$  – число поездов в системе в момент времени  $t$ , на которых не оказали влияния отказы станции;  $n(t)$  – число задержанных поездов к моменту  $t$ ;  $N(t)$  – число поездов, находящихся в системе в момент времени  $t$ .

Нестационарный коэффициент простоя системы определяет долю поездов, находящихся в состоянии отказа (или задержки), и вычисляется по формуле

$$R(t) = 1 - K(t). \quad (3.81)$$

### 3.5. Математические модели расчёта параметров надёжности систем

Функции надёжности  $P(t)$ , отказов  $F(t)$ , плотность вероятностей отказа  $f(t)$  и интенсивность отказов  $\lambda(t)$  взаимосвязаны друг с другом. Поэтому если известен один из этих показателей, то двое других можно вычислить по формулам, представленным в табл. 3.1.

Таблица 3.1. Формулы взаимосвязи параметров надёжности элементов

Исходная функция	Формулы для определения остальных функций			
	$P(t)$	$F(t)$	$f(t)$	$\lambda(t)$
$P(t)$	—	$1 - P(t)$	$-\frac{dP(t)}{dt}$	$-\frac{1}{P(t)} \frac{dP(t)}{dt}$
$F(t)$	$1 - F(t)$	—	$\frac{dF(t)}{dt}$	$\frac{1}{1 - F(t)} \frac{dF(t)}{dt}$
$f(t)$	$\int_t^\infty f(x)dx$	$\int_0^t f(x)dx$	—	$\frac{P(t)}{\int_t^\infty f(x)dx}$
$\lambda(t)$	$e^{-\int_0^t \lambda(x)dx}$	$1 - e^{-\int_0^t \lambda(x)dx}$	$\lambda(t)e^{-\int_0^t \lambda(x)dx}$	—

Определение показателей надёжности станций в виде сложных динамических систем и их подсистем даёт возможность решать практические задачи эксплуатационной работы с учётом характеристик надёжности. Для этой цели удобно использовать типовые математические модели, позволяющие рассчитывать параметры надёжности транспортных систем. Для оценки надёжности транспортной системы необходимо исследовать: её структуру; характер образования отказов; статистические закономерности распределений параметров  $t_{\text{бот}}$  и  $t_{\text{вос}}$ .

В зависимости от сложности и соответствия реальным замерам исходных данных осуществляется выбор математических типовых аналитических моделей. Например, работу станции, узла или участка можно характеризовать последовательностью пар  $\{\theta_i, t_{\text{вос}i}\}$ . У каждой такой  $i$ -й пары замеряется интервал безотказной работы узла на  $i$ -м участке временной оси ( $\theta_i$ ) и интервал восстановления ( $t_{\text{вос}i}$ ). При этом предполагают, что в периоды  $t_{\text{вос}i}$  система задерживает транспортный поток, и поэтому считается, что узел работает с отказами. Под  $t_{\text{вос}i}$  обычно понимают время простоя узла, отводимое на ремонт и восстановление работоспособности узла.

При произвольном законе распределения времён между отказами и восстановлениями для вычисления показателей надёжности рекомендуются

следующие формулы:

1. Среднее время работы системы, функционирующей до первого отказа

$$T_p = \int_0^\infty xf(x)dx = \int_0^\infty xdF(x) = \int_0^\infty P(x)dx. \quad (3.82)$$

2. Среднее время восстановления

$$T_{\text{вос}} = \int_0^\infty xf_{\text{вос}}(x)dx = \int_0^\infty xdF_{\text{вос}}(x) = \int_0^\infty [1 - F_{\text{вос}}(x)]dx. \quad (3.83)$$

3. Вероятность безотказной работ системы к моменту  $t_0$

$$P(t_0) = 1 - F(t_0). \quad (3.84)$$

4. Вероятность отказа системы

$$Q(t_0) = F(t_0). \quad (3.85)$$

5. Коэффициент готовности системы

$$K_T = \frac{T_{\text{вос}}}{T_p + T_{\text{вос}}}. \quad (3.86)$$

6. Коэффициент оперативной готовности системы

$$R(t_0) = \frac{1}{T_p - T_{\text{н}}} \int_{t_0}^\infty F(x)dx. \quad (3.87)$$

Поскольку при аналитических расчётах показателей надёжности широко используется экспоненциальное распределение для времён безотказной работы  $F(t_{\text{бор}})$  и времён восстановления работоспособности  $F(t_{\text{вос}})$ , ниже приводятся формулы расчёта основных показателей надёжности для случая, когда предполагается экспоненциальный характер этих распределений.

1. Функция надёжности

$$P(t_0) = e^{-\lambda t_0}, \quad (3.88)$$

где  $\lambda = 1/T_p$  – интенсивность потока отказов.

2. Функция отказов

$$Q(t_0) = 1 - e^{-\lambda t_0}. \quad (3.89)$$

3. Функции распределения вероятностей отказов и восстановлений:

$$F(t) = 1 - e^{-\lambda t}; \quad F_n(t) = 1 - e^{-\mu t}, \quad (3.90)$$

где  $\mu = 1/T_{\text{вос}}$  – интенсивность потока восстановлений;

$$\gamma = \frac{\lambda}{\mu}; \quad T_p = \frac{1}{\lambda}; \quad T_{\text{вос}} = \frac{1}{\mu};$$

$$K_r = \frac{\mu}{\lambda + \mu} = \frac{T_p}{T_p + T_{\text{вос}}} = \frac{1}{1 + \gamma}; \quad R(t_0) = ke^{-\lambda t_0}. \quad (3.91)$$

В ряде случаев имеет смысл использовать приближённые формулы, представленные в табл. 3.2. Там же указаны условия применения формулы и погрешность вычислений ( $\sigma$ ).

Таблица 3.2. Погрешности расчёта вычислений

Показатель	Приближённая формула	Условия применения	Погрешности
$P(t_0)$	$1 - \lambda t_0$	$\gamma t_0 \ll 1$	$\delta < \frac{1}{2}(\lambda t_0)^2$
$Q(t_0)$	$\lambda t_0$	$\lambda t_0 \ll 1$	$\delta < \frac{1}{2}(\lambda t_0)^2$
$K_r$	$1 - \gamma$	$\gamma \ll 1$	$\delta < \gamma$
$R(t_0)$	$1 - \gamma - \lambda t_0$	$\gamma \ll 1$ $\lambda t_0 \ll 1$	$\delta < 2,5 \{ \lambda \max(t_0, T_{\text{ш}}) \}$

В качестве примера ниже излагается методика расчёта эмпирических характеристик железнодорожного узла (ЖУ). Исследуем надёжность ЖУ, связывающего станции  $X$ ,  $K$ ,  $\Pi$ , и  $M$  по схеме, представленной на рис. 3.4.

На интервале времени  $(0, t_0)$  графики подвода поездов к станции имеют вид, представленный на рис. 3.5. Весь временной интервал работы ЖУ разбит на кванты длительностью  $\Delta t$ . Поезда, приходящие от станции  $X$  к станции  $K$ , имеют номера  $(x_i, i = \overline{1, 5})$ . Аналогично нумеруются поезда, приходящие от станции  $\Pi$  к станции  $K$  ( $\Pi_j = \overline{0, 12}$ ). Кружочками обведены номера задержанных поездов.

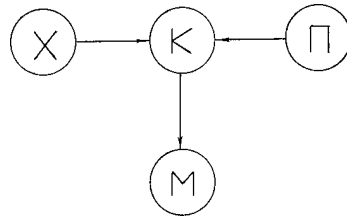


Рис. 3.4. Схема связей станций на железнодорожном узле

Со станции  $X$  были отправлены поезда  $x_1 - x_5$ , а со станции  $\Pi$  на станцию  $K$  пришли поезда  $\Pi_0 - \Pi_{12}$ . Все эти 18 поездов транзитом прошли через станцию  $K$ . Из-за технологических трудностей на станции  $K$  были последовательно задержаны поезда  $x_2, \Pi_5, \Pi_6, \Pi_9$ . Процесс подвода поездов к станциям представлен на рис. 3.5. По оси времен для станции  $X$  указаны моменты отправления поездов  $X_i$ , на оси времен  $K_1$  – моменты прихода поездов  $X_i$  на станцию  $K$ . По оси времен  $\Pi$  указаны моменты отправления поездов  $\Pi_j$  со станции  $\Pi$ , на оси времен  $K_2$  – моменты прихода поездов  $\Pi_j$  на станцию  $K$ . На оси времен  $M$  показаны моменты прихода поездов  $X_i$  и  $\Pi_j$  на станцию  $M$ . Сделаем проек-

цию прихода поездов  $X_i$  и  $\Pi_j$  на станцию  $K$  и ухода их со станции  $M$ .

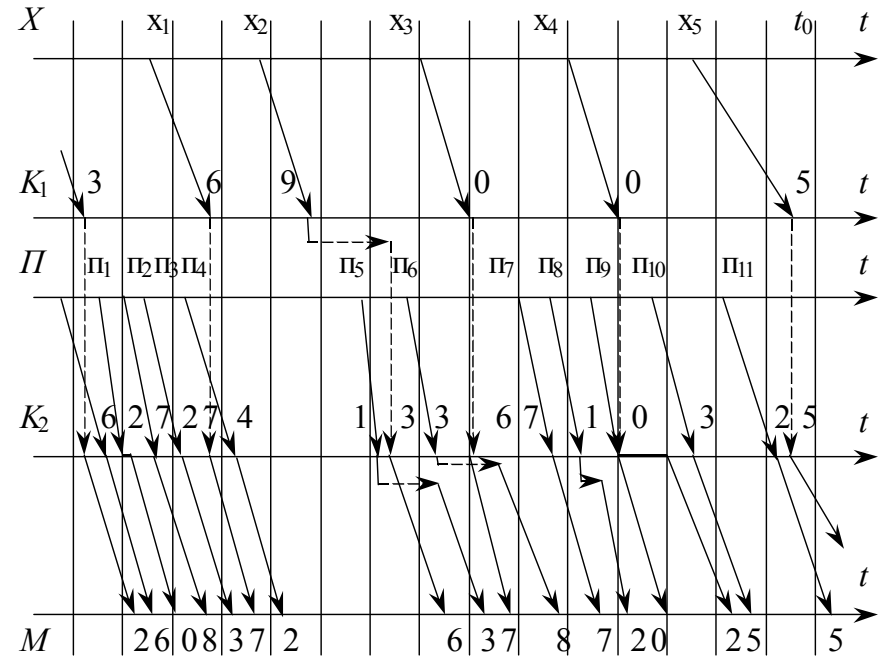


Рис. 3.5. График подвода поездов к железнодорожному узлу

Выполним анализ этих временных диаграмм и вычислим основные характеристики надёжности ЖУ.

1. За период  $t_0$  со станции  $X$  подведено 5 поездов. Из них поезд  $x_3$  задержан (обведен кружком). При этом  $N(0, t_0) = 5, n(t_0) = 1$ . Поэтому эмпирическое значение функции надёжности  $\tilde{P}_{XK}(t_0) = 1 - \frac{1}{5} = 0,8$ .

2. Для поездов, подведенных со станции  $\Pi$ , имеем следующие характеристики:  $N(0, t_0) = 12, n(t_0) = 3$ . Определим эмпирическое значение функции надёжности:  $\tilde{P}_{XK}(t_0) = 1 - \frac{3}{12} = 0,75$ .

3. Теперь суммарно по транспортному потоку (через станцию  $M$ ) имеем:  $N_{M}(0, t_0) = 18; n_M(t_0) = 4; \tilde{P}_{XK}(t_0) = 1 - \frac{4}{18} = 0,778$ .

Заметим, что задержки четырёх поездов на станции  $K$  могут быть вовсе не из-за низкой надёжности станции  $K$ . Они могут быть обусловлены работой крупной сортировочной станции. Вычислим остальные характеристики

надёжности ЖУ, ориентируясь на станцию  $M$ . Для функции отказов находим  $Q_M(t_0) = 1 - P_M(t_0) = 1 - 0,778 = 0,222$ .

$$\text{Плотность распределения отказов } \tilde{f}_M(t) = \frac{\Delta n_M(0, t_0)}{N(0, t_0)t_0} = \frac{4}{18t_0}.$$

Поскольку  $t_0$  в нашем случае равна 140 минут, то интенсивность отказов ЖУ находится ближе  $\tilde{\lambda}_M(t) = \frac{4}{18 \cdot 140}$  поездов в минуту. Как видим, имеет место совпадение плотности распределения отказов с интенсивностью появления отказов в работе ЖУ.

### 3.6. Надёжность невосстанавливаемых резервированных элементов системы

Резервированным соединением элементов в системе будем называть такое соединение, при котором отказ наступает только после отказа основного и всех резервированных элементов системы. Различаем общее и раздельное резервирования. При общем резервировании резервируются отдельные части системы или даже элементов системы. Кроме того, различают резервирование с целой и дробной кратностью. По способу включения резервных элементов резервирование разделяется на постепенное и замещение. При постепенном резервировании элементы, находящиеся в резерве, подключаются к основным в течение всего времени работы системы и находятся в одинаковых режимах ожидания. При резервировании замещением резервные элементы замещают основные элементы только после их отказа функционировать в системе. По способу замещения элементы могут находиться в трёх состояниях резерва: нагруженном, облегчённом, ненагруженном.

Рассмотрим принципы организации общего резервирования с постоянно включенным резервом с целой кратностью. Пусть в системе имеется  $m$  параллельных ветвей элементов, а в каждой ветви имеется  $n$  элементов, из которых один элемент является основным, а остальные  $n-1$  элементов – резервными. Вероятность безотказной работы элементов  $P_c(t)$  определяется по формуле

$$P_c(t) = \{1 - [1 - e^{-\lambda t}]^{m+1}\}^n. \quad (3.92)$$

Среднее время безотказной работы системы

$$T_{\text{срп}} = \int_0^{\infty} P_c(t) dt. \quad (3.93)$$

Интенсивность отказов  $i$ -го элемента определяется из выражения

$$v_i = (i+1)/(m+1). \quad (3.94)$$

Тогда среднее время безотказной работы  $i$ -го элемента

$$T_{\text{срп}} = \frac{(n-1)!}{\lambda(m+1)} \sum_{i=0}^m \frac{1}{v_i(v_i+1)\dots(v_i+n)}. \quad (3.95)$$

При общем резервировании способа замещения с целой кратностью  $m$  и  $(m+1)$  вероятности безотказной работы системы будут соответственно  $P_m(t)$  и  $P_{m+1}(t)$ . Обозначим через  $P(t-\tau)$  вероятность безотказной работы основной системы в течение времени  $(t-\tau)$ , а через  $a_m(\tau)$  – частоту отказов резервированной системы кратности  $m$  в момент времени  $\tau$ . Для расчётов обычно используется рекуррентная формула

$$P_{m+1}(t) = P_m(t) + \int_0^t P(t-\tau)a_m(\tau)d\tau. \quad (3.96)$$

В случае экспоненциального закона распределения надёжности возможны три случая состояния резерва системы: ненагруженный, недогруженный, нагруженный. Рекуррентная формула (3.96) приобретает следующий вид:

а) для первого случая

$$P_c(t) = e^{-\lambda_0 t} \sum_{i=0}^m \frac{(\lambda_0 t)^i}{i!}; \quad T_{\text{срп}} = T_{\text{срп0}}(m+1), \quad (3.97)$$

где  $\lambda_0$  и  $T_{\text{срп0}}$  – соответственно интенсивность отказов и средняя наработка элемента до первого отказа основного (нерезервированного) элемента;

б) для второго случая частота отказов  $a_i$  и интенсивность отказов резервного устройства до замещения

$$a_i = \prod_{j=1}^{i-1} \left(1 + \frac{\lambda_0}{\lambda_1}\right); \quad K = \frac{\lambda_1}{\lambda_0}. \quad (3.98)$$

Тогда основные характеристики надёжности системы определяются по формулам

$$P_c(t) = e^{-\lambda_0 t} \left[1 + \sum_{i=1}^m \frac{a_i}{i!} (1 - e^{-\lambda_1 t})^i\right]; \quad T_{\text{срп}} = \frac{1}{\lambda_0} \sum_{i=0}^m \frac{1}{1+ik}; \quad (3.99)$$

в) для третьего случая расчётные формулы параметров надёжности системы имеют вид

$$P_c(t) = 1 - [1 - e^{-\lambda_0 t}]^{m+1}; \quad T_{\text{срп}} = \frac{1}{\lambda_0} \sum_{i=1}^m \frac{1}{1+i}. \quad (3.100)$$

Наконец, рассмотрим раздельное резервирование элементов системы с целой кратностью. Очевидно, что в этом случае вероятность безотказной работы системы  $P_c(t)$  в момент  $t$  равна произведению вероятностей элемен-

тов системы

$$P_c(t) = \prod_{i=1}^n P_i(t). \quad (3.101)$$

Сами значения  $P_i(t)$  вычисляются по формулам общего резервирования замещением:

$$P_{m+1}(t) = P_m(t) + \int_0^t P(t-\tau) a_m(\tau) d\tau. \quad (3.102)$$

В случае экспоненциального распределения расчётные формулы параметров надёжности системы имеют вид

$$P_c(t) = e^{-\lambda_0 t} \sum_{i=0}^m \frac{(\lambda_0 t)^i}{i!}; \quad T_{\text{ср}} = T_{\text{ср}0}(m+1). \quad (3.103)$$

### 3.7. Моделирование надёжности систем с монотонной структурой

Реальные системы имеют более сложную структуру, которую нельзя свести к параллельно-последовательным соединениям. К таким системам относятся: распределение системы связи, системы управления объектами, трубопроводы, транспортные системы, энергосистемы. Подобные системы называют системами с монотонной структурой. Основное их свойство состоит в том, что отказ любого из элементов может привести к ухудшению надёжности или отказу всей системы. Введём булеву функцию состояния элемента:

$$x_i = \begin{cases} 1 - \text{элемент работоспособен;} \\ 0 - \text{элемент неработоспособен.} \end{cases}$$

Тогда вероятность работоспособного состояния  $i$ -го элемента  $P_i = Mx_i$ .

Состояния системы представим в виде булевого вектора  $X = (x_1, x_2, \dots, x_n)$ .

Если в этом векторе отсутствует  $i$ -я координата, то используется следующее обозначение для  $(n-1)$ -мерного вектора:  $X = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ .

Наконец, обозначим через  $X_\alpha$  вектор, у которого отсутствуют компоненты  $x_i$ , для которых  $i \in \alpha$ . В этом случае размерность вектора  $n = |\alpha|$ .

Для системы вводится понятие структурной функции:

$$\Phi(X) = \begin{cases} 1, & \text{если } X \text{ соответствует работоспособности системы;} \\ 0, & \text{в противном случае.} \end{cases}$$

Вероятность работоспособного состояния системы  $h$  определим как математическое ожидание структурной функции  $h = M\Phi(X)$ .

Эту величину обычно находят с помощью функции надёжности системы:

$$h(P) = h(P_1, \dots, P_n), \quad (3.104)$$

где  $P = (P_1, \dots, P_n)$  – вектор, компонентами которого являются вероятности работоспособного состояния системы.

Структура системы называется монотонной, если выполняются следующие условия:

$$\begin{aligned} \Phi(1) &= 1, \text{ где } 1 = (1, \dots, 1); \quad \Phi(0) = 0, \text{ где } 0 = (0, \dots, 0); \\ \Phi(X) &\geq \Phi(Y), \text{ если } X > Y, 1 = (1, \dots, 1), \end{aligned} \quad (3.105)$$

т. е. когда выполняется условие, что  $X \geq Y$  при  $i = \overline{1, n}$ .

Для расчёта характеристик надёжности подобных систем используются различные методики. Рассмотрим две из них: прямого перебора и аналитико-статистическую.

**Методика прямого перебора** используется в тех случаях, когда каждый из  $n$  элементов системы может быть только в двух состояниях (работоспособный – 1; отказа – 0). В результате система может находиться в  $2^n$  различных состояниях. Введём следующие обозначения для состояний системы:

$H_0$  – все  $n$  элементов работоспособны;

$H_i$  – отказал  $i$ -й элемент, а все остальные элементы работоспособны;

$H_{ij}$  – отказали  $i$ -й и  $j$ -й элементы, а все остальные работоспособны;

$H_{1,2,\dots,n}$  – отказали все  $n$  элементов системы.

Всё множество состояний системы разделим на два подмножества: работоспособности  $\varphi$  и состояния отказа системы  $\psi$ . Если вероятность любого состояния  $H_a$  работоспособности системы представляет собой  $P_a$ , тогда вероятность работоспособности системы

$$P\{H_a \in \varphi\} = \sum_{H_a \in \varphi} P_a. \quad (3.106)$$

Допустим, что система состоит из взаимно независимых элементов и известны вероятности состояния работоспособности  $i$ -го элемента ( $p_i$ ) и его неработоспособности ( $q_i$ ). Допустим также, что эти вероятности постоянны и не зависят от времени  $t$ . Тогда расчётные формулы надёжности системы имеют вид

$$P_0 = \prod_{i=1}^n P_i; \quad (3.107)$$

$$P_i = q_i \prod_{\substack{k=1 \\ k \neq i}}^n P_k = \frac{q_i}{p_i} P_0 = \gamma_i P_0; \quad (3.108)$$

$$P_{ij} = q_i q_j \prod_{\substack{k=1 \\ k \neq i, j}}^n P_k = \gamma_i \gamma_j P_0; \quad (3.109)$$

$$P_{1, \dots, n} = P_0 \prod_{i=1}^n \gamma_i = \prod_{i=1}^n q_i. \quad (3.110)$$

Здесь под  $P_i$  понимают либо вероятности работы  $i$ -го элемента до первого отказа, либо любой из видов коэффициента готовности (нестационарный коэффициент оперативной готовности или просто коэффициент готовности элемента). Если под  $P_i$  понимать вероятность работы  $i$ -го элемента до отказа

$$P_i(t) = P\{\xi_i \geq t\}, \quad (3.111)$$

где  $\xi_i$  – случайная величина времени наработки системы до отказа  $i$ -го элемента, то вероятность безотказной работы системы

$$P(t) = P\{\xi \geq t\}, \quad (3.112)$$

где  $\xi$  – случайная величина времени наработки до первого отказа уже всей системы.

Далее расчёт этих вероятностей проводится по формуле (3.52). Средняя наработка системы до отказа

$$T_{cp} = \int_0^{\infty} P(t) dt. \quad (3.113)$$

Если же под  $P_i$  понимают коэффициент готовности  $i$ -го элемента; то и вероятность  $P(t)$  будет коэффициентом готовности системы в целом.

Рассмотрим пример моделирования надёжности системы с монотонной структурой, состоящей из пяти элементов (рис. 3.6). Предполагается, что все элементы идентичны и соединены в виде типового моста. Вероятность безотказной работы системы определяется по экспоненциальному закону  $\varphi(t) = e^{-\lambda t}$  с параметром  $\lambda = 0,01$   $1/\text{час}$ . Расчётное время  $t = 10$  часов.

Составим таблицу возможных состояний системы и определим принадлежность этих состояний к подмножествам  $\varphi$  и  $\psi$ . Здесь

$$x_i = \begin{cases} 1 & - i\text{-й элемент исправен;} \\ 0 & - i\text{-й элемент неисправен.} \end{cases}$$

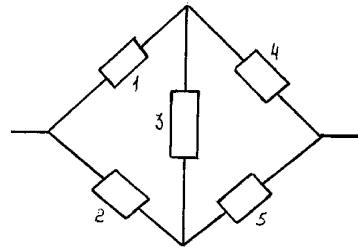


Рис. 3.6. Пример системы с монотонной структурой

Таблица 3.3. Таблица возможных состояний мостовой системы

Индекс состояния $a_i$	Состояния элементов $x_i$					Вид подмножества	Вероятность состояния системы
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$		
0	1	1	1	1	1	$\varphi$	$p^5$
1	0	1	1	1	1	$\varphi$	$q \cdot p^4$
2	1	0	1	1	1	$\varphi$	$q \cdot p^4$
3	1	1	0	1	1	$\varphi$	$q \cdot p^4$
4	1	1	1	0	1	$\varphi$	$q \cdot p^4$
5	1	1	1	1	0	$\varphi$	$q \cdot p^4$
12	0	0	1	1	1	$\psi$	$q^2 \cdot p^3$
13	0	1	0	1	1	$\varphi$	$q^2 \cdot p^3$
14	0	1	1	0	1	$\varphi$	$q^2 \cdot p^3$
23	1	0	0	1	1	$\varphi$	$q^2 \cdot p^3$
24	1	0	1	0	1	$\varphi$	$q^2 \cdot p^3$
25	1	0	1	1	0	$\varphi$	$q^2 \cdot p^3$
34	1	1	0	0	1	$\varphi$	$q^2 \cdot p^3$
35	1	1	0	1	0	$\varphi$	$q^2 \cdot p^3$
45	1	1	1	0	0	$\psi$	$q^2 \cdot p^3$
134	0	1	0	0	1	$\varphi$	$q^3 \cdot p^2$
135	0	1	0	1	0	$\psi$	$q^3 \cdot p^2$
145	0	1	1	0	0	$\psi$	$q^3 \cdot p^2$
234	1	0	0	0	1	$\psi$	$q^3 \cdot p^2$
235	1	0	0	1	0	$\varphi$	$q^3 \cdot p^2$
245	1	0	1	0	0	$\psi$	$q^3 \cdot p^2$
345	1	1	0	0	0	$\psi$	$q^3 \cdot p^2$
1345	0	1	0	0	0	$\psi$	$q^4 \cdot p$
15	0	1	1	1	0	$\varphi$	$q^2 \cdot p^3$

Согласно табл. 3.3 составим формулу расчёта:

$$\Phi(t) = \sum p_i q_i = p^5(t) + 5q(t)p^4(t) + 8q^2(t)p^3(t) + q^3(t)p^2(t) + q^4(t)p(t). \quad (3.114)$$

Заменим  $p = 1 - q$  и подставим в данную формулу. Для краткости опустим также индекс  $t$ :  $P = 1 - 2q^2 - 2q^3 + 5q^4 - 2q^5$ .

Пусть  $q(t = 10) = 1 - e^{-0,1} \approx 0,1$ .

Тогда  $P(t = 10) = 1 - 2 \cdot 0,01 - 2 \cdot 0,001 + 5 \cdot 0,0001 - 2 \cdot 0,00001 \approx 0,978$ .

Для вычисления средней наработки системы до отказа представим функцию  $P$  зависящей от  $p$ :

$$P(t) = 2p^2(t) - 5p^4(t) + 2p^3(t) + 2p^2(t). \quad (3.115)$$

Для экспоненциального закона  $P^k = e^{-k\lambda t}$ ;  $\lambda = 1/T_i$ .



Тогда средняя наработка системы

$$T = \int_0^{\infty} \{2e^{-5t/T_i} - 5e^{-4t/T_i} + 2e^{-3t/T_i} + 2e^{5t/T_i}\} dt. \quad (3.116)$$

Интегрируя это выражение для  $T_i = 10$  часов, получим  $T = 82$  часа.

Рассмотрим теперь **методику аналитико-статистического моделирования** системы с монотонной структурой. Если количество элементов системы велико, то переборный алгоритм расчёта параметров её надёжности не эффективен. Лучше использовать аналитико-статистическую методику расчётов, в основе которой лежит метод Монте-Карло. Однако когда элементы достаточно надёжны и система обладает избыточностью, то напрямую метод Монте-Карло использовать также неэффективно, поскольку многие реализации трудно достижимы либо неинформативны. Если все элементы системы равнонадёжны, то обычно используют смешанные методики аналитико-статистической оценки параметров надёжности системы.

Для примера рассмотрим моделирование отказов системы в фиксированных состояниях по следующему алгоритму:

1. Аналитически вычисляются вероятности различных типов состояний системы. Устанавливаем следующие состояния системы:

$H_0$  – работоспособность всех элементов системы;

$H_1$  – отказ ровно одного элемента в системе;

$H_k$  – отказ ровно  $k$  элементов в системе.

Вероятность события  $H_k$

$$P_k = C_n^k q^k p^{n-k}, \quad q = 1 - p. \quad (3.117)$$

2. Методом статистических испытаний (Монте-Карло) находят оценки условных вероятностей  $\Phi_k$  того, что система, находящаяся в состоянии  $H_k$ , работоспособна по выбранному критерию.

3. Вычисляется полная вероятность работоспособности системы:

$$P = P_0 + \sum_{k=1}^n \Phi_k P_k. \quad (3.118)$$

Зачастую в качестве критерия работоспособности системы выбирается наличие связи между всеми элементами. В этих случаях  $\Phi_k$  есть условная вероятность связности системы при отказе ровно  $k$  ребёр сети. Для этого осуществляется конкретная реализация ровно  $k$  отказов тех или иных ребер по методу Монте-Карло, т. е. разыгрывается ровно  $k$  равномерно распределённых величин  $\xi_1, \dots, \xi_k$  в интервале  $[1, n]$ . Эта конкретная реализация сети проверяется на связность. Если сеть окажется при  $i$ -й реализации  $\xi_1$  связной, то к счётчику числа успешных исходов прибавляется единица:

$$v_k(N_k) = v_k(N_k) + 1, \quad (3.119)$$

где  $N_k$  – полное число реализаций. Тогда

$$\Phi_k = v_k(N_k) / N_k. \quad (3.120)$$

В ряде случаев при исследовании систем с монотонной структурой моделируют процесс накопления отказов элемента до отказа системы. Обычно операция выполняется тогда, когда оценивается живучесть сети. Случайным образом переводятся элементы сети в состояние отказа до тех пор, пока по выбранному критерию (например, по связности элементов в системе) система переходит в состояние отказа. При этом запоминается число элементов в  $S$ -м испытании  $K_s$ , при котором впервые в  $S$ -й реализации получен отказ. Объём числа испытаний  $N$  рассчитывается исходя из доверительной вероятности

$$\Phi_k = \frac{1}{N} \sum_{k=1}^N \delta(K_s), \quad (3.121)$$

$$\text{где } \delta(K_s) = \begin{cases} 1, & \text{если } K_s \leq K; \\ 0, & \text{если } K_s > K. \end{cases}$$

Далее вычисляется полная вероятность работоспособности системы:

$$P = P_0 + \sum_{k=1}^n \Phi_k P_k. \quad (3.122)$$

При экспериментальных оценках надёжности систем всё многообразие показателей сводится к двум типам:

- разные виды наработки (средняя до отказа или между отказами, до предельного состояния, срок сохраняемости, время восстановления и т. д.);
- вероятности отказа системы (безотказная работа, исправного состояния в произвольный момент времени  $t$ , восстановления за заданное время и т.д.).

В первом случае наблюдаемыми величинами являются случайные интервалы, во втором – число событий в испытаниях (число отказов, число восстановлений, число предельных состояний). Экспериментальные данные разделяются на два типа: полученные при стратегиях с непрерывным контролем; с периодическим контролем.

При этом на характер и продолжительность испытаний влияют следующие факторы:  $N$  – число элементов;  $U$  – отсутствие замены элементов или их восстановления;  $R$  – наличие замены отказавших элементов;  $M$  – присутствие восстановлений отказавших элементов;  $T$  – испытания завершаются при стечении фиксированного отрезка времени (наработки);  $Z$  – испытания заканчиваются по достижении фиксированного числа реализаций.

## 4. МОДЕЛИ МАССОВОГО ОБСЛУЖИВАНИЯ

### 4.1. Предпосылки описания сложной технической системы аппаратом теории массового обслуживания

Своеобразный класс вероятностных моделей образуют модели массового обслуживания, отражающие особенности поведения СТС, подвергающихся воздействию потока тех или иных событий. Предполагается, что СТС состоит из однотипных элементов, каждый из которых выполняет простейшие действия, сводящиеся в модели к одной операции по обслуживанию требований. Необходимо исследовать взаимодействие компонент СТС, найти узкие места в структуре СТС. В таких случаях в качестве аппаратов исследования используются системы массового обслуживания (СМО). Это время – логическая система, состоящая из пяти типов описателей: *ВХП* – входящего потока требований на обслуживание; *ПП* – обслуживающего прибора с параметром «Время обслуживания» ( $t_{обс}$ ); *DISC* – дисциплины обслуживания требований приборами; *ОЧ* – очереди обслуживания; *ВЫХ* – выходящего потока требований.

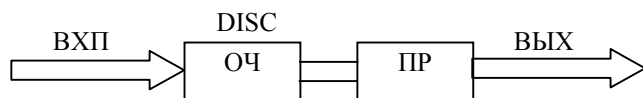


Рис. 4.1. Структура простейшей СМО

На рис. 4.1 представлена простейшая структура СМО. В простейшем случае СМО может состоять из одного обслуживающего прибора *ПП*, на вход которого поступает единственный поток требований *ВХП*. Различаем три вида *DISC*: с потерями требований; с ожиданием требований в очереди; смешанные. При ожидании необходимо указать порядок выбора требований *DISC* из очереди *ОЧ*.

СМО по **характеру поступления требований** в систему подразделяются на системы с регулярными и случайными потоками требований. Случайные потоки требований в систему подразделяются на стационарные и нестационарные. По **количеству поступающих требований** в один момент времени бывают СМО с ординарными и неординарными потоками, по **связи** между требованиями – СМО с последствием и без последствия. По **характеру поведения** требований различают СМО с отказами, с ограниченным ожиданием, с неограниченным ожиданием. По **способу выбора требований** на обслуживание исследуются СМО с приоритетами, в порядке поступления, со случайным вектором, с дообслуживанием, по **характеру обслуживания** требований – СМО с детерминированным временем обслужи-

вания, случайным временем обслуживания, смешанного типа. По **числу каналов обслуживания** изучаются СМО обыкновенные и многоканальные. По **количеству этапов** различают однофазные и многофазные СМО, по **однородности** требований – однородные и неоднородные. По **ограниченности потока требований** исследуются замкнутые и разомкнутые СМО. Рассмотрим более подробно способы формализации СМО и их составляющих.

**Входящий поток.** Задаётся время между двумя соседними моментами времени поступления требований в систему  $\tau$ . Поток требований в систему считается **ординарным**, если они поступают поодиночке. Если число требований зависит только от длительности периода  $\tau$  и не зависит от его положения на интервале  $(0, t)$ , то поток является **стационарным**. Если распределение величины  $\tau$  не зависит от времени наблюдения (положения на временной оси), то говорят, что отсутствует последствие. В теории СМО часто предполагают, что поток требований является простейшим. По определению это такой поток требований, когда он ординарен, стационарен и без последствия. Установлено, что для простейшего потока количество требований  $n$ , поступающих в систему на интервале времени  $(0, t)$ , имеет распределение Пуассона:

$$P(n, t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad (4.1)$$

где  $\lambda(t)$  – среднее значение интенсивности потока требований. Для пуассоновского потока среднее время между интервалами поступления требований в систему  $\bar{\tau} = 1/\lambda = \sigma$ .

**Обслуживающие приборы.** Основной характеристикой *ПП* будет время, затрачиваемое на обслуживание одного требования  $t_{обсл}$ , являющееся чаще всего случайной величиной. Время обслуживания требования *ПП* определяется видом требований, интенсивностью их поступления в *ОЧ*  $\lambda$  и порядком их поступления из *ОЧ* на прибор *DISC*. В теории СМО обычно полагают, что  $t_{обсл}$  распределено по экспонентному закону. Для определения характеристик *ПП* к заданию функции распределения  $F(t_{обсл})$  необходимо указать число каналов обслуживания (число требований, которые одновременно могут обслуживаться в СМО). В этом случае говорят об одноканальной или многоканальной СМО. В многоканальной СМО предполагается, что каждое требование обслуживается одновременно на своём *ПП*. Требования могут обслуживаться последовательно на нескольких *ПП* до момента выхода из системы в виде выходящего потока *ВЫХ* и затем поглощаться. В этом случае считают, что имеет место **многофазная** СМО. Для каждого требования последовательность фаз обслуживания называется путём обслуживания требований в СМО. Если для обслуживания требований используется несколько *ПП*, то СМО считается многоузловым *ПП*, через который проходят требова-

ния, находящиеся на различных фазах обслуживания. Такой прибор называют многовходовым.

Цель моделирования СМО – выбор оптимального количества каналов обслуживания у многоканального узла и установление наиболее рациональной дисциплины обслуживания *DISC* в СМО. В многофазной СМО в процессе обслуживания перед каждой фазой обслуживания могут образовываться очереди требований.

**Дисциплины обслуживания.** Различают следующие *DISC*:

- *FIFO* (первым пришёл, первым обслуживается);
- *LIFO* (последним пришёл, первым обслуживается);
- *RANDOM* (случайный выбор требований из очереди);
- *RR* (кругового обслуживания);
- *PRIOR* (приоритетные, с абсолютным, относительным и смешанным приоритетами).

**Очереди.** Служат для согласования скоростей поступления требований на *ПП* и скоростей обслуживания требований в системах с ожиданием. Предполагается, что у *ПП* имеется бесконечный по размерам буфер, используемый для хранения требований. В ряде случаев налагается ряд ограничений на пребывание требований в очереди. Наиболее часто используются следующие ограничения:

- по максимальному времени нахождения в очереди  $t_{\text{нах}}$ ;
- по длине очереди  $l_{\text{очер}}$ ;
- по вероятности того, что требование покинет систему необслуженным

$P_{\text{от}}$ .

Существуют СМО, у которых отсутствуют очереди, которые называются «системы с отказами». В таких системах, если требование застаёт *ПП* занятым обслуживанием предыдущего требования, то требование покидает СМО необслуженным. Зачастую встречаются СМО смешанного типа.

**Модели СМО.** Используются для решения задач анализа и задач синтеза СТС. Задачи анализа предполагают оценку эффективности функционирования СМО при неизменных наперёд заданных характеристиках СТС и структуре СТС, дисциплинах обслуживания требований, потоках требований и законах распределения времени их обслуживания на *ПП*. Задачи синтеза направлены на поиск оптимальных параметров СМО. Допускается варьирование структурой и параметрами компонентов СМО. Если изучены или заданы предварительно входящие потоки требований, механизм и дисциплина их обслуживания, то это даёт основание для построения математической модели СТС.

В задачах анализа СМО в качестве основных показателей функционирования системы могут быть использованы:

$P_{\text{отк}}$  – вероятность потери требования в системе;

$P_0$  – вероятность простоя каналов обслуживания (все *ПП* свободны);

$P_k$  – вероятность того, что обслуживанием заняты именно  $k$  приборов;

$P_n$  – вероятность того, что обслуживанием заняты все  $n$  приборов;

$N_3 = \sum_{k=1}^n kP_k$  – среднее число занятых приборов, характеризующее степень загрузки системы;

пень загрузки системы;

$N_0 = \sum_{k=0}^{n-1} (n-k)P_k$  – среднее число *ПП*, свободных от обслуживания;

$K_n = N_0 / n$  – коэффициент простоя приборов;

$K_3 = N_3 / n$  – коэффициент загрузки приборов.

Для систем с ожиданием используют дополнительные характеристики:

$\bar{t}_{\text{ож}} = \int_0^{\infty} t dP_1(t_{\text{ож}} > t)$  – среднее время ожидания в очереди, где

$P_1(t_{\text{ож}} > t) = \sum_{k=n}^{\infty} P_k P_1(t_{\text{ож}} > t)$  – условная вероятность того, что за время ожидания в момент поступления очередного требования в ней уже было обслужено  $k$  требований,

$P_2(t_{\text{ож}} < t) = \sum_{k=n}^{\infty} P_k P_1(t_{\text{ож}} < t)$  – вероятность того, что время пребывания в очереди не продлится больше величины  $t$ ;

$M_{\text{ож}} = \sum_{k=n}^{\infty} (k-n)P_k$ , для  $k \geq n$  – средняя длина очереди к прибору;

$M = \sum_{k=1}^{\infty} kP_k = M_{\text{ож}} + N_3$  – среднее число требований, находящихся в среде обслуживания СМО;

$P_m = \sum_{k=m+1}^{\infty} P_k$  – вероятность того, что число требований, ожидающих начала обслуживания, больше некоторого числа  $m$ .

Если входные потоки требований простейшие, а времена их обслуживания на *ПП* имеют экспоненциальное распределение, то для большинства задач анализа СМО зачастую удаётся построить аналитические модели СТС в виде одноканальных или многоканальных СМО. На таких моделях легко определить перечисленные выше характеристики эффективности систем для любого заданного момента времени. Предсказательные возможности таких моделей велики, несмотря на вероятностный характер параметров СМО. Когда же либо входные потоки требований не являются простейшими, либо

времена обслуживания требований на *ПП* не распределены экспоненциально, то аналитически подобного рода задачи не решаемы. В этих случаях исследователи вынуждены обращаться к построению численных или имитационных моделей СМО (эти классы моделей рассматриваются в третьей части данного пособия).

#### 4.2. Аналитические модели СТС в виде одноканальных систем массового обслуживания

Все рассматриваемые ниже СМО исследуются аналитическими методами с помощью моделей, построенных на основе представления процесса функционирования СТС как марковского процесса с непрерывным временем и дискретными состояниями. Рассмотрим следующие наиболее часто встречающиеся задачи анализа одноканальных СМО:

- анализ детерминированной системы;
- исследование одноканальной разомкнутой СМО;
- анализ одноканальной замкнутой СМО с ожиданием.

**Задача анализа детерминированной системы** формулируется следующим образом. Пусть исследуется производственный процесс, в котором поступление требований происходит через равные промежутки  $\Delta t_n = \text{const}$  (т. е. интенсивность потока  $\lambda = 1/\Delta t_n = \text{const}$ ) и обслуживание производится через равные промежутки времени  $\Delta t_{\text{обс}} = \text{const}$  (т. е. интенсивность обслуживания  $\mu = 1/\Delta t_{\text{обс}} = \text{const}$ ). Имеется один канал обслуживания. Предполагается, что  $\Delta t_{\text{обс}}/\Delta t_n = \lambda/\mu < 1$  (ибо в противном случае очередь будет бесконечно расти) и что к каналу обслуживания в системе уже имеется  $n$  требований. Необходимо определить время, когда очередь к каналу исчезнет.

**Построение модели СМО.** Величину  $\psi = \lambda/\mu$  называют коэффициентом использования. Пока обслуживается очередь из  $n$  требований в течение времени  $t = n\Delta t_{\text{обс}}$ , в систему вновь поступит на обслуживание  $n_1$  требований:

$$n_1 = \frac{t}{\Delta t_n} = \frac{n\Delta t_{\text{обс}}}{\Delta t_n} = n \frac{\lambda}{\mu} = n\psi.$$

Аналогично пока будут обслуживаться эти  $n_1$  требований в течение времени  $t_1 = n_1 \cdot \Delta t_{\text{обс}}$ , в СМО дополнительно поступит на обслуживание  $n_2$  требований:

$$n_2 = \frac{t_1}{\Delta t_n} = \frac{n_1 \Delta t_{\text{обс}}}{\Delta t_n} = n_1 \frac{\lambda}{\mu} = n_1 \psi = n\psi^2.$$

Это будет происходить до тех пор, пока  $t_k > \Delta t_n$ , после чего очередь исчезнет. Весь процесс функционирования СМО представим в аналитическом

виде. Время, через которое очередь исчезнет, представим в виде

$$T = t + t_1 + t_2 + \dots + t_k = t + \sum_{i=1}^k t_i,$$

$$T = n\Delta t_{\text{обс}} + n_1\Delta t_{\text{обс}} + n_2\Delta t_{\text{обс}} + \dots + n_k\Delta t_{\text{обс}} = \frac{n}{\mu} (1 + \psi + \psi^2 + \dots + \psi^k),$$

$$\text{т. е. } T = \frac{n(1 - \psi^{k+1})}{\mu(1 - \psi)}. \quad (4.2)$$

В модели использована формула суммы геометрической прогрессии. Чем ближе интенсивность потока требований  $\lambda$  к интенсивности их обслуживания  $\mu$ , тем за больший промежуток времени исчезнет очередь. Для практических расчётов используется приближённая формула

$$T \approx n/(\mu - \lambda). \quad (4.3)$$

**Задача анализа разомкнутой системы с ожиданием** формулируется следующим образом. Пусть задана СТС, для которой справедливы следующие гипотезы:

– вероятность поступления требований зависит не от принятого начала отсчёта времени, а только от продолжительности периода наблюдений (стационарность потока);

– поток требований ординарен и отсутствует последствие. Известны также интенсивности поступления  $\lambda$  и интенсивности обслуживания требований  $\mu$ . Необходимо определить основные характеристики системы:  $P_0, P_n, M_{\text{ож}}, N_3, t_{\text{ож}}$ . Поскольку входной поток является простейшим, то на вход СМО поступает пуассоновский поток с интенсивностью  $\lambda$ . Если все потоки событий, переводящие систему из одного состояния в другое, являются пуассоновскими, то для этих систем вероятности состояний описываются с помощью системы обыкновенных дифференциальных уравнений. Размеченный граф состояний такой СМО представлен на рис. 4.2.

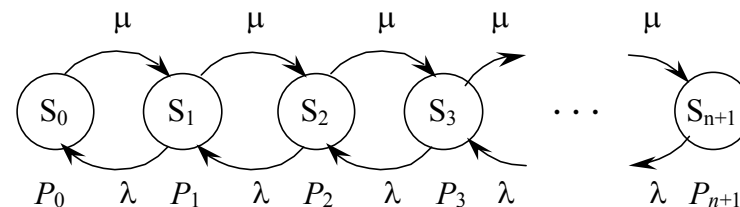


Рис. 4.2. Граф перехода разомкнутой СТС из состояния в состояние



В соответствии с этим графом, используя мнемоническое правило, запишем систему дифференциальных уравнений для вероятностей состояний:

$$\begin{aligned} \frac{dP_0(t)}{dt} &= -m\lambda P_0(t) + \mu P_1(t); \\ &\dots \\ \frac{dP_n(t)}{dt} &= -[(m-n)\lambda + \mu]P_n(t) + (m-n+1)\lambda P_{n-1}(t) + \mu P_{n+1}(t); \\ &\dots \\ \frac{dP_m(t)}{dt} &= -\mu P_m(t) + \lambda P_{m-1}(t). \end{aligned}$$

Ограничимся исследованием установившегося режима работы СМО. Тогда  $dP_n(t)/dt = 0$  при  $n=0, 1, \dots, m$ . Вместо системы обыкновенных дифференциальных уравнений получаем систему алгебраических уравнений:

$$\left. \begin{aligned} m\lambda P_0 - \mu P_1 &= 0; \\ \dots \\ [(m-n)\lambda + \mu]P_n - (m-n+1)\lambda P_{n-1} - \mu P_{n+1} &= 0; \\ \dots \\ \mu P_m - \lambda P_{m-1} &= 0. \end{aligned} \right\} \quad (4.8)$$

Для  $0 < n \leq m$  рекуррентная формула имеет вид

$$\left. \begin{aligned} \text{при } n=0 \quad P_1 &= \frac{m\lambda}{\mu} P_0 = m\psi P_0; \\ \text{при } n=1 \quad P_2 &= \frac{1}{\mu} \{[(m-1)\lambda + \mu]P_1 - m\lambda P_0\} = (m-1)\psi P_1; \\ \text{при } n=2 \quad P_3 &= \frac{1}{\mu} \{[(m-2)\lambda + \mu]P_2 - (m-1)\lambda P_1\} = (m-2)\psi P_2; \\ &\dots \\ &P_n = (m-n+1)\psi P_{n-1} \end{aligned} \right\} \quad (4.9)$$

Вероятность того, что в системе находится  $n$  требований,

$$P_n = (m-n+1)\psi(m-n+2)\psi \dots (m-1)\psi m\psi P_0 = \frac{m!\psi^n}{(m-n)!} P_0. \quad (4.10)$$

Используя равенство  $\sum_{n=0}^m P_n = \sum_{n=0}^m \frac{m!\psi^n}{(m-n)!} P_0 = 1$ , найдём вероятность про-

стоя канала обслуживания:

$$P_0 = \left[ 1 + \sum_{n=1}^m \frac{m!\psi^n}{(m-n)!} \right]^{-1}. \quad (4.11)$$

Среднее число требований в очереди

$$M_{\text{ож}} = \sum_{n=0}^m (n-1)P_n = m! \sum_{n=1}^m \frac{n!\psi^n}{(m-n)!} P_n = m \frac{1+\psi}{\psi} (1-P_0). \quad (4.12)$$

Среднее число требований, находящихся в системе,

$$M = \sum_{n=1}^m nP_n = m! P_0 \sum_{n=1}^m \frac{n!\psi^n}{(m-n)!} = m - \frac{1}{\psi} (1-P_0). \quad (4.13)$$

Среднее время ожидания требования в очереди

$$\bar{t}_{\text{ож}} = \frac{M_{\text{ож}}}{\lambda(m-M_{\text{ож}})} = \frac{1}{\mu} \left[ \frac{m}{1-P_0} - \frac{1+\psi}{\psi} \right]. \quad (4.14)$$

Среднее время ожидания требования в системе

$$T_{\text{сист}} = \frac{M}{\lambda(m-M)} = \frac{1}{\mu} \left[ \frac{m}{1-P_0} - \frac{1}{\psi} \right]. \quad (4.15)$$

### 4.3. Анализ многоканальных систем массового обслуживания

Рассмотренные ниже СМО также исследуются аналитическими методами при предположении пуассоновских потоков требований на входе системы и экспоненциального характера времени обслуживания требований в системе. Рассматриваются наиболее часто встречающиеся задачи исследования многоканальных СМО:

- анализ разомкнутой СМО с ожиданием;
- анализ разомкнутой СМО с отказом;
- анализ замкнутой СМО с ожиданием.

**Задача анализа разомкнутой СМО с ожиданием** формулируется следующим образом. Пусть известны интенсивность  $\lambda$  поступления потока требований и интенсивность  $\mu$  обслуживания системой этих требований. Число каналов обслуживания постоянно и равно  $N_k$ . Требуется определить основные характеристики СМО:  $P_n(t)$ ,  $P_0(t)$ ,  $M_{\text{ож}}$ ,  $\bar{t}_{\text{ож}}$ . Кроме того, необходимо найти среднее число свободных каналов обслуживания  $N_{\text{св}}$ . В такой СМО возможны два случая:

- число требований в системе  $n$  меньше числа каналов ( $0 \leq n < N_k$ );

– число требований в СМО больше числа каналов ( $n \geq N_k$ )

В первом случае все требования, находящиеся в системе, одновременно обслуживаются и при этом не все каналы заняты. Общая интенсивность обслуживания будет равна  $n\mu$ . Размеченный граф состояний представлен на рис. 4.4.

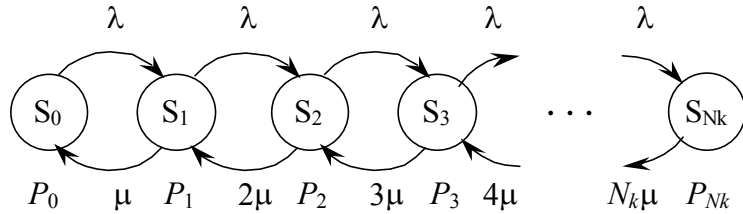


Рис. 4.4. Граф переходов разомкнутой СМО с ожиданием

В соответствии с размеченным графом состояний, используя мнемоническое правило, запишем систему обыкновенных дифференциальных уравнений для вероятностей состояний системы:

$$\begin{aligned} \text{для } n = 0: \quad & \frac{dP_0(t)}{dt} = -P_0(t)\lambda + P_1(t)\mu; \\ \dots & \dots \\ \text{для } i < N_k: \quad & \frac{dP_i(t)}{dt} = P_{i-1}(t)\lambda - (\lambda + i\mu)P_i(t) + (i+1)\mu P_{i+1}(t); \\ \dots & \dots \\ \text{для } n \geq N_k: \quad & \frac{dP_n(t)}{dt} = P_{n-1}(t)\lambda - (\lambda + N_k\mu)P_n(t) + N_k\mu P_{n+1}(t). \end{aligned}$$

Ограничимся исследованием установившегося режима работы СМО при  $\lambda = \text{const}$  и  $\mu = \text{const}$ . При  $t \rightarrow \infty$   $P_n = \text{const}$ . Тогда вместо системы обыкновенных дифференциальных уравнений получим систему алгебраических уравнений

$$\left. \begin{aligned} -\lambda P_0 + \mu P_1 &= 0; \\ \dots & \dots \\ \lambda P_{i-1} - (\lambda + i\mu)P_i + (i+1)\mu P_{i+1} &= 0; \\ \dots & \dots \\ \lambda P_{n-1} - (\lambda + N_k\mu)P_n + N_k\mu P_{n+1} &= 0. \end{aligned} \right\} \quad (4.16)$$

Определим рекуррентные выражения для вычисления вероятности  $P_n$  нахождения СМО в состояниях  $n$ :

$$\left. \begin{aligned} P_1 &= \frac{\lambda}{\mu} P_0 = \psi P_0; \\ P_2 &= \frac{1}{2\mu} [(\lambda + \mu)P_1 - \lambda P_0] = \frac{\psi^2}{2} P_0; \\ \dots & \dots \\ P_{N_k} &= \frac{1}{N_k\mu} \{[\lambda + N_k - 1]\mu P_{N_k-1} - \lambda P_{N_k-2}\} = \frac{\psi^{N_k}}{N_k!} P_0. \end{aligned} \right\} \quad (4.17)$$

Из этих выражений видно, что для всех  $n < N_k$  вероятность нахождения в системе  $n$  требований определится по формуле

$$P'_n = \frac{\psi^n}{n!} P_0. \quad (4.18)$$

Для состояний системы  $n \geq N_k$  расчётные формулы имеют вид

$$\left. \begin{aligned} \text{для } n = N_k: \quad P_{N_k+1} &= \frac{\psi}{N_k} \frac{\psi^{N_k}}{N_k!} P_0; \\ \text{для } n = N_k + 1: \quad P_{N_k+2} &= \frac{\psi^2}{N_k^2} \frac{\psi^{N_k}}{N_k!} P_0. \end{aligned} \right\} \quad (4.19)$$

Из полученных выражений видно, что для состояний системы  $n \geq N_k$  вероятность нахождения в системе  $n$  требований определится по формуле

$$P''_n = \frac{\psi^2}{N_k^{n-N_k} N_k!} P_0. \quad (4.20)$$

Имея аналитические выражения для всех состояний СМО, а также используя очевидное равенство  $\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{N_k-1} P'_n + \sum_{n=N_k}^{\infty} P''_n = 1$ , определим расчётные формулы для нахождения основных характеристик СМО:

– вероятность простоя каналов обслуживания

$$P_0 = \left[ \sum_{n=0}^{N_k} \frac{\psi^n}{n!} + \frac{\psi^{N_k}}{N_k!(1 - \psi/N_k)} \right]^{-1}; \quad (4.21)$$

– среднее число требований, находящихся в очереди,

$$M_{\text{ож}} = \frac{\psi^{N_k+1}}{N_k! N_k (1 - \psi/N_k)^2} P_0; \quad (4.22)$$

– средняя длина очереди

$$\bar{t}_{\text{ож}} = M_{\text{ож}} / \lambda ; \quad (4.23)$$

– среднее число каналов, свободных от обслуживания:

$$N_{\text{св}} = \sum_{n=0}^{N_k-1} (N_k - n) P_n = \sum_{n=0}^{N_k-1} (N_k - n) \frac{\Psi^n}{n!} P_0 . \quad (4.24)$$

**Задача анализа разомкнутой СМО с отказом** формулируется следующим образом. Пусть исследуется любая СТС, которая может быть представлена как разомкнутая СМО. Интенсивность поступления требований известна и равна  $\lambda\mu$ . Интенсивность обслуживания каждым каналом также известна и равна  $\mu$ . Если требование застало все  $N_k$  каналов занятыми, то оно получает отказ и покидает систему необслуженным. Требуется определить следующие характеристики СМО:  $P_0, P_n$ . Кроме того, надо найти среднее число занятых каналов обслуживания  $N_{\text{зк}}$ . Состояния системы  $S_i$  будем связывать с числом занятых каналов обслуживания требований. Возможны  $N_k$  состояний системы:

- $S_0$  – все каналы свободны и ни одно требование не обслуживается;
  - $S_1$  – один канал занят и одно требование обслуживается;
  - ...
  - $S_n$  – все  $n$  каналов занято и обслуживается  $n$  требований;
  - $S_{N_k}$  – все  $N_k$  каналов занято и обслуживается  $N_k$  требований.
- Размеченный граф состояний СМО приведён на рис. 4.5.

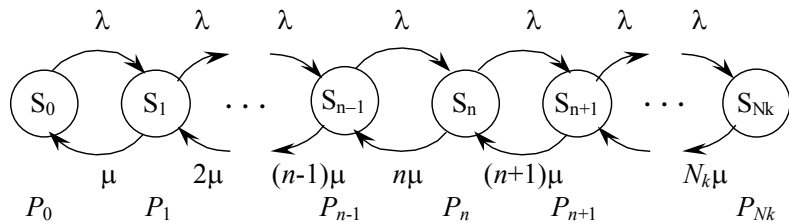


Рис. 4.5. Размеченный граф состояний СМО с отказами

В соответствии с этим графом, используя мнемоническое правило, запишем систему обыкновенных дифференциальных уравнений для вероятностей состояний:

$$\begin{aligned} dP_0(t) / dt &= -\lambda P_0(t) + \mu P_1(t); \\ &\dots \dots \dots \\ dP_n(t) / dt &= \lambda P_{n-1}(t) - (\lambda + n\mu) P_n(t) + (n+1)\mu P_{n+1}(t); \\ &\dots \dots \dots \\ dP_{N_k}(t) / dt &= -N_k \mu P_{N_k} + \lambda P_{N_k-1}(t). \end{aligned} \quad (4.25)$$

Эти уравнения в теории массового обслуживания называются уравнениями Эрланга. Исследуем стационарный режим работы системы (т. е. при  $t \rightarrow \infty$ ). Получим систему рекуррентных алгебраических уравнений

$$\left. \begin{aligned} -\lambda P_0 + \mu P_1 &= 0; \\ -(\lambda + \mu) P_1 + \lambda P_0 + 2\mu P_2 &= 0; \\ \dots \dots \dots \\ -(\lambda + n\mu) P_n + \lambda P_{n-1} + (n+1)\mu P_{n+1} &= 0; \\ \dots \dots \dots \\ -N_k \mu P_{N_k} + \lambda P_{N_k-1} &= 0. \end{aligned} \right\} \quad (4.26)$$

Из этих уравнений находим систему расчётных формул

$$P_1 = \Psi P_0; \quad P_2 = \frac{\Psi^2}{2} P_0; \dots; \quad P_n = \frac{\Psi^n}{n!} P_0. \quad (4.27)$$

Вероятность того, что все каналы обслуживания свободны,

$$P_0 = 1 / \sum_{n=0}^{N_k} \frac{\Psi^n}{n!}. \quad (4.28)$$

Вероятность того, что занято равно  $n$  каналов обслуживания,

$$P_n = \Psi^n / \left( n! \sum_{n=0}^{N_k} \frac{\Psi^n}{n!} \right). \quad (4.29)$$

Среднее число занятых каналов обслуживания

$$N_{\text{зк}} = \sum_{n=0}^{N_k} n P_n = \sum_{n=0}^{N_k} \frac{\Psi^n}{(n-1)! \sum_{n=0}^{N_k} \frac{\Psi^n}{n!}}. \quad (4.30)$$

**Задача анализа замкнутой многоканальной СМО с ожиданием** формулируется следующим образом. Пусть исследуется СТС, представляемая замкнутой СМО, в которой обслуживаемые требования возвращаются для повторного дообслуживания. Интенсивность поступления одного требования в СМО известна и равна  $\lambda\mu$ . Интенсивность обслуживания каждого канала также известна и равна  $\mu$ . Число каналов обслуживания  $N_k$  постоянно, число требований, циркулирующих в режиме дообслуживания, постоянно и равно  $m$ . Будем считать, что  $N_k \leq m$ . Требуется определить основные характеристики СМО:  $P_0, P_n, \bar{t}_{\text{ож}}, M_{\text{ож}}$ , а также среднее число свободных каналов



обслуживания ( $N_{ck}$ ). Аналогично предыдущим примерам, состояния системы  $S_i$  ( $i = \overline{1, N_k}$ ) будем связывать с числом требований, находящихся в системе.

При этом возможны два случая:

– число требований  $n$ , поступивших в систему, меньше числа каналов обслуживания, т. е. все они находятся на обслуживании ( $n \leq N_k$ );

– число требований больше числа каналов обслуживания ( $n \geq N_k$ ), из них  $N_k$  обслуживаются, а  $r$  требований ожидает в очереди ( $r = 1, 2, \dots, m - N_k$ ).

Граф состояний СМО представлен на рис. 4.6.

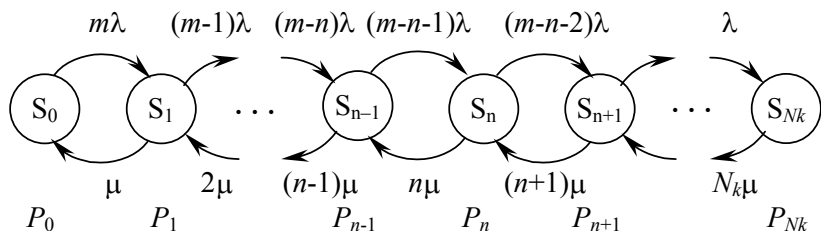


Рис. 4.6. Граф состояний многоканальной СМО с ожиданием

В соответствии с этим графом, используя мнемоническое правило, запишем систему дифференциальных уравнений для вероятностей состояний:

для  $0 \leq n \leq N_k$ :

$$\frac{dP_0(t)}{dt} = -m\lambda P_0(t) + \mu P_1(t);$$

$$\frac{dP_n(t)}{dt} = -[(m-n)\lambda + n\mu]P_n(t) + (m-n+1)\lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t);$$

для  $N_k \leq n \leq m$ :

$$\frac{dP_n(t)}{dt} = -[(m-n)\lambda + N_k\mu]P_n(t) + (m-n+1)\lambda P_{n-1}(t) + N_k\mu P_{n+1}(t);$$

$$\frac{dP_m(t)}{dt} = -N_k\mu P_m(t) + \lambda P_{m-1}(t).$$

Ограничимся рассмотрением установившегося режима. В этом случае  $\lambda = \text{const}$ ,  $\mu = \text{const}$ ,  $t \rightarrow \infty$ ,  $P_n = \text{const}$ ,  $dP_n(t)/dt = 0$  ( $n = \overline{1, m}$ ). Поэтому вместо системы обыкновенных дифференциальных уравнений получаем систему рекуррентных алгебраических уравнений, из которых находим вероятность

того, что в системе находится  $n$  требований:

$$\text{для } 0 \leq n \leq N_k : \quad P'_n = \frac{m!\psi^n}{n!(m-n)!} P_0; \tag{4.31}$$

$$\text{для } N_k \leq n : \quad P''_n = \frac{m!\psi^n}{N_k^{n-N_k} N_k!(m-n)!} P_0.$$

Для определения вероятности простоя каналов обслуживания  $P_0$  используем равенство  $\sum_{n=0}^m P_n = \sum_{n=0}^{N_k-1} P'_n + \sum_{n=N_k}^m P''_n = 1$ , откуда

$$P_0 = \left[ \sum_{n=0}^{N_k-1} \frac{m!\psi^n}{n!(m-n)!} + \sum_{n=N_k}^m \frac{m!\psi^n}{N_k^{n-N_k} N_k!(m-n)!} \right]^{-1}. \tag{4.32}$$

Среднее число требований, ожидающих начала обслуживания,

$$M_{ож} = \sum_{n=N_k}^m (n - N_k) P''_n = \sum_{n=N_k}^m \frac{(n - N_k) m!\psi^n}{N_k^{n-N_k} N_k!(m-n)!} P_0. \tag{4.33}$$

Среднее время ожидания требований в очереди

$$\bar{t}_{ож} = \frac{M_{ож}}{M(N_k - N_{ck})}. \tag{4.34}$$

Среднее число свободных каналов обслуживания

$$N_{ck} = \sum_{n=0}^{N_k-1} (n - N_k) P'_n = \sum_{n=0}^{N_k-1} \frac{(N_k - n) m!\psi^n}{n!(m-n)!} P_0. \tag{4.35}$$

#### 4.4. Примеры использования систем массового обслуживания при решении эксплуатационных задач транспорта

Рассмотрим несколько задач анализа транспортных потоков, использующих в качестве аппарата формализации системы массового обслуживания.

**Задача 1.** Необходимо рассчитать число маневровых локомотивов при формировании поездов на сортировочной станции. Содержательное описание процесса формирования поездов формулируется следующим образом. Объектом анализа является односторонняя сортировочная станция с пропускной способностью  $\theta = 36$  составов/сутки. В сортировочном парке станции имеется 12 путей. Средняя продолжительность цикла формирования одного состава  $t_{оф} = 0,3$  ч. Стоимость одного вагоно-часа – 10 руб., а одного маневрового локомотиво-часа с составительской бригадой – 684 руб. Длитель-

ность всех перерывов в работе одного маневрового локомотива за сутки  $\Sigma t_{\text{пер}} = 40$  мин. Для эффективного использования каждого дополнительного маневрового локомотива необходимо 2 млн руб. капитальных затрат на путевое развитие. Расходы на содержание этого путевого развития составляют 800 тыс. руб. в год при нормативном сроке окупаемости капиталовложений в 10 лет. Основное допущение для решения задачи в аналитическом виде: число составов, ожидающих окончания формирования в сортировочном парке, является случайной величиной, распределённой по закону Пуассона. Необходимо определить оптимальное количество маневровых локомотивов, необходимых для окончания формирования составов, при условии, что локомотивы не закреплены за маневровыми районами. Формальное описание процесса формирования поездов выполним поэтапно.

1. Определим минимальное число маневровых локомотивов, потребное для обеспечения окончания формирования поездов. Поскольку маневровый район в хвосте сортировочного парка представляет собой вытяжку с закреплённой за ней группой сортировочных путей, то его функционирование можно представить в виде СМО без потерь, но с ограничением по числу требований в очереди при ожидании обслуживания. Расчётная формула для нахождения минимального числа маневровых локомотивов имеет вид

$$M_{\min} = \frac{v t_{\text{оф}} \cdot 60}{1440 - \sum t_{\text{пер}}} = \frac{36 \cdot 0,3 \cdot 60}{1440 - 40} \approx 0,5 \text{ лок.}$$

Округлим  $M_{\min} = 1$ .

2. Определим среднее число составов, накапливающихся в сортировочном парке за 1 час:  $N_{\text{Н}}^{\text{час}} = \frac{v}{24} = \frac{36}{24} = 1,5$  состава.

Округлим  $N_{\text{Н}}^{\text{час}} = 2$ .

3. Оценим число составов, накапливающихся в сортировочном парке за время окончания формирования одного состава:

$$a = N_{\text{Н}}^{\text{час}} t_{\text{оф}} = 1,5 \cdot 0,3 = 0,45 \text{ состава.}$$

4. Максимальное число составов, ожидающих окончания формирования во всём сортировочном парке, выбирают равным 1/3 числа сортировочных путей. Поэтому максимальное число составов, ожидающих окончания формирования,  $m = 12/3 = 4$ .

5. Среднее число составов  $S_{\text{ср}}$ , ожидающих окончания формирования в сортировочном парке, может быть определено по формуле Эрланга для системы с ограничением по длине очереди:

$$S_{\text{ср}} = \frac{\frac{\alpha^M}{M!} \sum_{S=1}^m S \left(\frac{\alpha}{M}\right)^S}{\sum_{K=0}^M \frac{\alpha^K}{K!} + \frac{\alpha^M}{M!} \sum_{S=1}^m S \left(\frac{\alpha}{M}\right)^S}, \quad (4.36)$$

где  $M$  – число маневровых локомотивов, работающих в сортировочном парке одновременно до окончания формирования составов;  $K$  – число одновременно занятых локомотивов;  $S$  – число составов, ожидающих формирования.

При  $M = 1$  маневровых локомотивах имеем

$$S_{\text{ср}} = \frac{\frac{(0,45)^1}{1!} \left[ \left( \frac{(0,45)}{1!} \right) \cdot 1 + (0,45)^2 \cdot 2 + (0,45)^3 \cdot 3 + (0,45)^4 \cdot 4 \right]}{\left[ \frac{(0,45)^0}{0!} + \frac{(0,45)^1}{1!} \right] + \frac{(0,45)^1}{1!} [(0,45)^1 + (0,45)^2 + (0,45)^3 + (0,45)^4]} = 0,32 \text{ состава.}$$

6. Оценим среднее значение времени простоя при ожидании окончания формирования каждого состава, накопившегося в сортировочном парке:

$$t_{\text{оф}}^{\text{ож}} = S_{\text{ср}} / N_{\text{Н}}^{\text{сут}} = 0,32 / 1,50 = 0,2 \text{ часа.}$$

7. Приведённые суточные расходы при данном числе маневровых локомотивов найдём по формуле  $\sum E_{\text{сут}} = M C_{\text{АМ}} \cdot 24 + N_{\text{Н}}^{\text{сут}} t_{\text{оф}}^{\text{ож}} m_{\psi} C_{\text{вч}}$ , где  $C_{\text{АМ}}$  – приведённая стоимость маневрового локомотиво-часа с составительской бригадой;  $N_{\text{Н}}^{\text{сут}}$  – число составов, формируемых за сутки;  $m_{\psi}$  – число вагонов в формируемом составе;  $C_{\text{вч}}$  – приведённая стоимость одного вагоно-часа.

При вычисленных и заданных значениях составляющих в этой формуле фактическое значение приведённых затрат

$$\sum E_{\text{сут}} = 1 \cdot 684 \cdot 24 + 36 \cdot 0,2 \cdot 50 \cdot 10 = 16766 \text{ руб/сут.}$$

Далее можно провести аналогичные расчёты при  $M=2$ ;  $M=3$  и т. д. до тех пор, пока сумма общих приведённых расходов  $\sum E_{\text{сут}}$  не перейдёт точку минимума. При этом по каждому из этих вариантов к  $\sum E_{\text{сут}}$  будут добавляться ещё дополнительно приведённые расходы, руб/сут, на путевое разви-

тие:  $\Delta \sum E_{\text{сут}} = \frac{1}{365} \left( \frac{K_{\text{ОН}}}{t_{\text{ОК}}} + \mathcal{E}_{\text{П}} \right) \Delta M$ , где  $K_{\text{ОН}}$ ,  $\mathcal{E}_{\text{П}}$  – соответственно капитальные затраты на путевое развитие и расходы на его содержание, связанные с введением каждого дополнительного маневрового локомотива ( $\Delta M$ ).

Одновременно с вычислением общих расходов на организацию работы сортировочной станции ( $\Sigma E_{\text{сут}} + \Delta \Sigma E_{\text{сут}}$ ) находим и коэффициент использования маневровых локомотивов  $K_{\text{исп}}$ . В итоге формируется таблица, в которой для каждого значения  $M$  указаны значения вектора параметров работы сортировочной станции ( $S_{\text{ср}}, t_{\text{оф}}, \Sigma E_{\text{сут}} + \Delta \Sigma E_{\text{сут}}, K_{\text{исп}}$ ). Анализ этой таблицы позволяет определить оптимальное значение параметров этого вектора качества работы сортировочной станции. Методика выбора оптимума при векторной оптимизации представлена в третьей части данного пособия в разделе, посвящённом принятию решений при исследовании СТС с помощью моделей ИСО.

**Задача 2.** Необходимо рассчитать потребную складскую площадь для размещения грузов, поступающих на склад краткосрочного хранения. Содержательное описание процесса размещения грузов на складе формулируется следующим образом. На склад краткосрочного хранения (СКХ) в течение года поступает  $v = 130000$  т грузов. СКХ работает круглосуточно весь год. Средняя масса груза в одной поступающей партии  $g = 400$  т. Расчётная нагрузка на  $1 \text{ м}^2$  площади склада  $\rho = 1 \text{ т/м}^2$ . Грузы поступают на хранение в склад неравномерно, т. е. моменты их прибытия имеют вероятностный характер. Входящий поток партий грузов является простейшим в терминах теории случайных процессов. Математическое ожидание длительности хранения различных партий грузов на складе  $M t_{\text{обс}} = 10$  сут. при показательном законе распределения длительности хранения  $t_{\text{обс}}$  с нулевым значением параметра сдвига экспоненты. Требуется определить потребную полезную площадь склада  $F$ , обеспечивающую пропуск заданного грузооборота  $v$ . Формализацию процесса размещения грузов выполним поэтапно на основе использования аппарата теории массового обслуживания.

1. Среднесуточное поступление на склад отдельных партий грузов (интенсивность входящего потока)  $\lambda = \frac{v}{\rho \cdot 365} = \frac{130000}{400 \cdot 365} = 0,9$ .

2. Интенсивность обслуживания требований (партий грузов) равна обратному значению среднего срока обслуживания:  $\mu = \frac{v}{M t_{\text{обс}}} = \frac{1}{10} = 0,1$ .

3. Вероятность отказа при заданном грузообороте и имеющемся числе ячеек (приборов обслуживания)  $n$

$$P_n = \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n \cdot \sum_{m=0}^n \frac{1}{m!} \left( \frac{\lambda}{\mu} \right)^m,$$

где  $n$  – расчётное число условных ячеек-площадок хранения на складе;  $m$  –

число одновременно занятых ячеек склада. Расчётное число ячеек определяется из условия  $\lambda / (\mu n) < 1$ , откуда  $n > \lambda / \mu$ . Таким образом, минимальное расчётное число ячеек равно целой части отношения  $\lambda / \mu$ . На практике максимальное расчётное число ячеек принимается примерно равным утроенному значению их минимального числа. В данном случае  $n_{\text{min}} = \lambda / \mu = 0,9 / 0,1 = 9$ . Тогда вероятность отказа при 9 ячейках на СКХ определится из формулы

$$P_{n=9} = \frac{\frac{1}{9!} \cdot 9^9}{\frac{1}{0!} \cdot 9^0 + \frac{1}{1!} \cdot 9^1 + \frac{1}{2!} \cdot 9^2 + \dots + \frac{1}{9!} \cdot 9^9} = \frac{1067}{4759} = 0,224.$$

4. Поскольку значение вероятности отказа не может превышать 1, то фактическая пропускная способность склада при данном числе ячеек

$$N_{\text{скл}} = v(1 - P_n) = 130000 \cdot (1 - 0,224) = 100000 \text{ т/год.}$$

5. Полезная площадь склада  $F$  при заданном размере ячейки-площадки  $f$ , м, определится из выражения  $F = n f$ .

При нагрузке на  $1 \text{ м}^2$  площади склада в 1 т и средней массе груза в одной партии  $g = 400$  т размер ячейки-площадки  $f = g / \rho = 400 / 1 = 400 \text{ м}^2$ . Тогда  $F = 9 \cdot 400 = 3600 \text{ м}^2$ .

Далее, увеличивая количество ячеек склада на 1, вычислим: вероятность отказа  $P_n$ , пропускную способность склада  $N_{\text{скл}}$ , т/год, и требуемую полезную площадь  $F$ ,  $\text{м}^2$ . Результаты расчётов для очередного значения  $n$ , выполненные путём повторения этапов 1–5, приведены в табл. 4.1.

Таблица 4.1. Результаты расчётов для различных значений  $n$

Число ячеек $n$	Вероятность отказа $P_n$	Пропускная способность $N_{\text{скл}}$ , т/год	Полезная площадь $F$ , $\text{м}^2$
9	0,224	100000	3600
10	0,179	106000	4000
11	0,140	112000	4400
12	0,078	120000	4800
13	0,042	125000	5200
14	0,030	126000	5600

Как видно из табл. 4.1, практически безотказная работа склада обеспечивается при вероятности отказа  $P_n = 0,04 \dots 0,05$ . В данном случае только 4–5% поступающих на склад грузов могут получить отказ в приёме из-за занятости складской площади. При 13 ячейках склад обеспечивает пропуск заданного грузооборота с вероятностью отказа 0,042. За год склад может не принять около 5 тыс. т грузов, т. е. практически он будет действовать безотказно. Полезная площадь склада в этом случае  $F = 5200 \text{ м}^2$ .

## ЗАКЛЮЧЕНИЕ

В данной части пособия рассмотрены наиболее известные или типовые подходы к построению и использованию аналитических ММ вероятностных процессов в СТС.

В третьей части излагаются методики построения ММ на основе имитации реальных процессов, известные алгоритмы и методики из теории принятия решений.

Студентам, желающим глубже познакомиться с практикой использования аналитических ММ вероятностных процессов при анализе СТС, авторы рекомендуют представленный ниже список литературы.

## СПИСОК ЛИТЕРАТУРЫ

1. *Ермаков С. М., Михайлов Г. А.* Курс статистического моделирования. – М.: Наука, 1976. – 319 с.
2. *Максимей И. В.* Математическое моделирование больших систем: Учеб. пособие. – Мн.: Вышэйшая школа, 1985. – 118 с.
3. *Буриков А. Д., Малинковский Ю. В., Металыцкий М. А.* Теория массового обслуживания: Учеб. пособие по спецкурсу. – Гродно: ГрГУ, 1984. – 320 с.
4. *Гнеденко Б. В., Коваленко И. Н.* Введение в теорию массового обслуживания. – М.: Наука, 1966. – 512 с.
5. *Ярошевич В. П.* Расчётно-графические работы по дисциплине «Применение математических методов в инженерных и экономических расчётах». – Гомель: БЕЛИИЖТ, 1987. – 120 с.
6. *Сакович В. А.* Исследование операций. – Мн.: Вышэйшая школа, 1985. – 256 с.
7. *Фиакко А., Мак-Кормик Г.* Нелинейное программирование. – М.: Мир, 1972. – 240 с.
8. *Зангвилл У. И.* Нелинейное программирование. – М.: Советское радио, 1973. – 312 с.

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	3
1. МОДЕЛИ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ .....	4
1.1. Основные понятия нелинейного программирования .....	4
1.2. Особенности задач нелинейного программирования и их геометрическая интерпретация .....	4
1.3. Классификация задач и методов нелинейной оптимизации .....	5
1.4. Типовые задачи нелинейного программирования, в которых целевая функция зависит от одной переменной .....	5
1.5. Решение задач нелинейного программирования методом неопределённых множителей Лагранжа .....	8
1.6. Использование градиентных методов для решения задач нелинейного программирования .....	9
2. ВЕРОЯТНОСТНЫЕ МОДЕЛИ СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМ .....	11
2.1. Особенности использования вероятностных моделей .....	11
2.2. Использование метода статистических испытаний при исследовании СТС .....	11
2.3. Виды и алгоритмы реализации единичных жребиев .....	13
2.4. Моделирование псевдослучайных величин .....	14
2.5. Примеры построения и использования псевдослучайных величин .....	15
2.6. Моделирование случайных процессов .....	18
2.7. Представление случайных процессов временными рядами .....	20
2.8. Моделирование поведения сложных технических систем марковскими процессами .....	23
3. МОДЕЛИ НАДЁЖНОСТИ ТЕХНИЧЕСКИХ СИСТЕМ .....	25
3.1. Основные понятия теории надёжности .....	25
3.2. Надёжность элемента системы, работающего до первого отказа .....	29
3.3. Надёжность восстанавливаемого элемента .....	31
3.4. Меры надёжности транспортных систем .....	34
3.5. Математические модели расчёта параметров надёжности систем .....	36
3.6. Надёжность невосстанавливаемых резервированных элементов системы .....	38
3.7. Моделирование надёжности систем с монотонной структурой .....	39
4. МОДЕЛИ МАССОВОГО ОБСЛУЖИВАНИЯ .....	42
4.1. Предпосылки описания сложной технической системы аппаратом теории массового обслуживания .....	42
4.2. Аналитические модели СТС в виде одноканальных систем массового обслуживания .....	44
4.3. Анализ многоканальных систем массового обслуживания .....	46
4.4. Примеры использования систем массового обслуживания при решении эксплуатационных задач транспорта .....	49
ЗАКЛЮЧЕНИЕ .....	52
СПИСОК ЛИТЕРАТУРЫ .....	52