

**МИНИСТЕРСТВО ТРАНСПОРТА И КОММУНИКАЦИЙ
РЕСПУБЛИКИ БЕЛАРУСЬ**
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТРАНСПОРТА»

Кафедра «Электрический подвижной состав»

В. Н. ГАЛУШКО

МАТЕМАТИЧЕСКИЕ МОДЕЛИ В ТРАНСПОРТНЫХ СИСТЕМАХ

Учебно-методическое пособие

Гомель 2015

МИНИСТЕРСТВО ТРАНСПОРТА И КОММУНИКАЦИЙ
РЕСПУБЛИКИ БЕЛАРУСЬ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТРАНСПОРТА»

Кафедра «Электрический подвижной состав»

В. Н. ГАЛУШКО

МАТЕМАТИЧЕСКИЕ МОДЕЛИ В ТРАНСПОРТНЫХ СИСТЕМАХ

*Одобрено методической комиссией механического факультета
в качестве учебно-методического пособия
для студентов специальности
“Городской электрический транспорт”*

Гомель 2015

УДК 51.001.57 (075.8)
ББК 22.1
Г16

Р е ц е н з е н т ы : канд. техн. наук, доцент кафедры «Математические проблемы управления» *С. Ф. Маслович* (УО «ГГУ им. Ф. Скорины»);
заведующий кафедрой «Электрический подвижной состав» канд. техн. наук, доцент *В. С. Могила* (УО «БелГУТ»)

Галушко, В. Н.

Г16 Математические модели в транспортных системах : учеб.-метод. пособие / В. Н. Галушко ; М-во трансп. и коммуникаций Респ. Беларусь, Белорус. гос. ун-т трансп. – Гомель : БелГУТ, 2015. – 154 с.
ISBN 978-985-554-404-4

Изложены основные сведения теории математического моделирования, проверки результатов моделирования, планирования экспериментов и обработки их результатов для получения научно обоснованных и достоверных выводов. Приведены конкретные примеры.

Предназначено для студентов механического факультета, обучающихся по специальности: «Городской электрический транспорт».

УДК 51.001.57 (075.8)
ББК 22.1

ISBN 978-985-554-404-4

© Галушко В. Н., 2015
© Оформление. УО «БелГУТ», 2015

ОГЛАВЛЕНИЕ

| | |
|--|-----|
| СПИСОК УСЛОВНЫХ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ | 5 |
| ВВЕДЕНИЕ | 6 |
| 1 МОДЕЛИ И МОДЕЛИРОВАНИЕ | 9 |
| 2 МЕТОДОЛОГИЯ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ | 12 |
| 2.1 Математические модели и их виды..... | 12 |
| 2.1.1 Теоретические математические модели аналитического типа..... | 12 |
| 2.1.2 Эмпирические математические модели..... | 20 |
| 2.2 Адекватность математических моделей..... | 25 |
| 2.3 Алгоритм научных исследований с помощью математического моделирования..... | 31 |
| 3 МЕТОДЫ РАЗРАБОТКИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ | 35 |
| 4 ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ И ПРИЕМЫ | 47 |
| 4.1 Вычислительные методы алгебры..... | 47 |
| 4.2 Вычислительные методы решения дифференциальных уравнений..... | 55 |
| 4.3 Приемы контроля математических моделей..... | 59 |
| 5 ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ | 61 |
| 5.1 Основы теории вероятностей..... | 61 |
| 5.1.1 Необходимые сведения из теории случайных событий..... | 61 |
| 5.1.2 Случайные величины и законы их распределения..... | 64 |
| 5.1.3 Предельные теоремы теории вероятностей..... | 76 |
| 5.2 Основы математической статистики..... | 79 |
| 5.2.1 Отбор информации..... | 80 |
| 5.2.2 Требования к точечным оценкам и объему выборки..... | 85 |
| 5.2.3 Проверка статистических гипотез..... | 88 |
| 5.2.4 Точность оценки. Доверительный интервал..... | 96 |
| 6 ОСНОВЫ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА | 97 |
| 6.1 Классификация задач многомерного статистического анализа..... | 97 |
| 6.2 Регрессионный анализ..... | 99 |
| 6.3 Корреляционный анализ..... | 103 |
| 6.4 Дисперсионный анализ..... | 108 |
| 6.5 Статистическая проверка адекватности математических моделей..... | 116 |
| 7 ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА | 125 |
| 7.1 Статистические методы планирования эксперимента..... | 125 |
| 7.1.1 Проблемы построения эксперимента..... | 125 |
| 7.1.2 Назначение плана эксперимента..... | 129 |
| 7.1.3 Планирование объема эксперимента..... | 131 |
| 7.2 Особые методы планирования эксперимента..... | 133 |
| 7.2.1 Специальные приемы планирования эксперимента..... | 133 |
| 7.2.2 Методы экспертных оценок..... | 135 |

| | |
|--|-----|
| 8 ОСНОВЫ ОПТИМИЗАЦИИ ВАРИАНТОВ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ | 139 |
| 8.1 Методы векторной оптимизации сложных систем..... | 139 |
| 8.2 Оптимизация нелинейной функции отклика от одной переменной..... | 142 |
| 8.3 Максимизация функции отклика от многих переменных..... | 146 |
| 8.4 Нахождение экстремума функции отклика от многих переменных при наличии ограничений..... | 151 |
| СПИСОК ЛИТЕРАТУРЫ | 153 |

СПИСОК УСЛОВНЫХ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

- АТС – автоматическая телефонная станция
- ДА – дисперсионный анализ
- ДТП – дорожно-транспортное происшествие
- КА – корреляционный анализ
- МК – микропроцессор
- ОДУ – обыкновенные дифференциальные уравнения
- РА – регрессионный анализ
- СВ – случайная величина
- СЛС – сложная система
- СМО – система массового обслуживания
- СС – случайное событие
- ТС – транспортное средство
- ФА – факторный анализ
- ЦФ – целевая функция

ВВЕДЕНИЕ

Соотношение науки и практики всегда было главной философской проблемой всех исследователей. Вопрос о том, насколько верны те или иные рассуждения, расчеты, действия человека, отражают суть реальности, является коренным вопросом любой научной теории.

Познанное человечеством по отношению к реальности можно условно иллюстрировать рисунком 1. Реальность всегда бесконечнообразна и бесконечномерна, поэтому границы ее на рисунке обозначены пунктиром. Нам в силу ограниченности органов чувств и познаний дано лишь составлять себе то или иное представление о реальности. Для науки существенное значение имеет уверенная оценка близости такого ограниченного представления к реальности. Оценить эту близость практикой, полагая ее истиной, невозможно, так как мы можем исследовать опять же лишь тот ограниченный круг свойств, который нам доступен. Поэтому между нашим представлением и реальностью оказывается дистанция неконтролируемого размера.

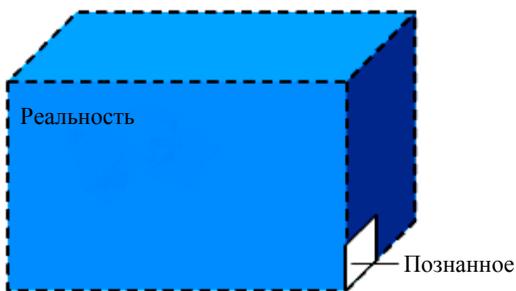


Рисунок 1 – Познанное человечеством по отношению к реальности

Но это еще не все трудности научного познания. Если один человек что-то новое познал о реальности, он стремится объяснить это другим людям. И здесь возникает необходимость использовать терминологию, единообразно понимаемую собеседниками, которая неизбежно содержит какие-то упрощения – абстракции. Абстракции позволяют отсечь,

отбросить из рассмотрения малозначительные факторы, ввести однозначные термины и представить себе объект в более простой форме, доступной формальной человеческой логике.

Однако неизбежно при этом познанное одним человеком передается другому в усеченном виде – в виде некоторых моделей. Такое соотношение

познанного и моделей можно иллюстрировать рисунком 2. На нем подчеркнута возможность существования для одного познанного явления нескольких моделей, отражающих те или иные его свойства в различных условиях.

Явления и объекты, достаточно точно и однозначно описываемые небольшим количеством факторов, получили название "хорошо организованных систем". Экспериментальные исследования таких систем заключались в наблюдении за результатом изменения одного фактора при постоянстве прочих. Законы природы, выявленные таким образом, непосредственно составляют *модель* явления. Примерами могут служить законы классической механики, генетики, химии и т.п.

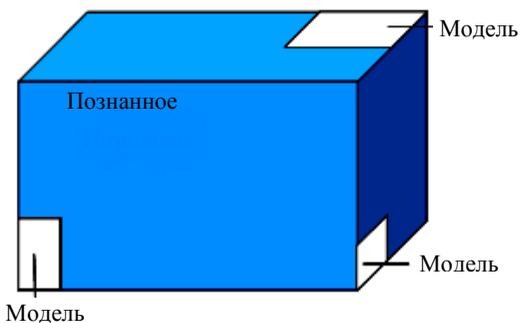


Рисунок 2 – Представление о соотношении познанного и моделей

Однако результаты, полученные с помощью умозрительно построенных моделей, не всегда хорошо соответствуют действительности – на результат действия выявленных законов накладывается влияние и других неучтенных факторов и погрешностей эксперимента. Попытка учета этих факторов приводит к усложнению модели. Если таких факторов много, модель становится **сложной** и трудно воспринимаемой. Например, расчет скорости движения транспортного средства (ТС) по длине следа торможения можно рассчитать по формуле

$$v_a = 1,8t_3j + \sqrt{25,92js_{10}},$$

где t_3 – время нарастания замедления (с); j – замедление (m/c^2); s_{10} – длина пути юза (м).

При программном моделировании (PC-Crash) скорости движения транспортного средства (ТС) по длине следа торможения реализована более адекватная и сложная модель за счет того, что учитываются характеристики работы двигателя, трансмиссии, модель шин каждого колеса, параметры работы подвески, распределение нагрузки в ТС, время срабатывания тормозной системы и рулевого привода, скорость вращения рулевого колеса, угол поворота управляемых колес, параметры работы тормозной системы.

Другим ярким примером влияния неучтенных факторов является исследование расхода топлива автомобилем, который зависит от технического состояния транспортного средства, состояния дороги и самое сложное – манеры вождения водителя и реакции на дорожную обстановку.

В XX веке стало ясно, что без изучения сложных систем, в том числе и созданных человеком, дальнейший прогресс невозможен. Простейшим примером такого типа систем является распространение разнообразных видов возмущений от ядерного взрыва: здесь есть и ударная волна, которая подчиняется одному закону, и световое излучение, подчиняющееся другому закону, и распространение радиации. Некоторые объекты вообще не могут быть изучены непосредственно: эксперименты с экономикой страны в познавательных целях, аварии на атомной электростанции или на действующей железной дороге, эксперименты с прошлым или планетами Солнечной системы.

Процессы в сложных системах нельзя описать законами, умозрительно построенными или полученными с помощью простых экспериментов. Для описания "*плохо организованных систем*" такой подход не всегда приемлем – необходимо учитывать не только множество разнообразных по своей природе связей – закономерностей, но и возможность различных методологических подходов и глубины отражения реальности. Примерами закономерностей могут служить курсы валют, сравнение активной и пассивной безопасности автомобилей, оценка надежности и т.п.

Как мы выяснили, умозрительные исследования могут привести к выявлению некоторых законов. Проверка их практикой тоже возможна лишь на ограниченном круге свойств. Постановка же исследовательского эксперимента нуждается не только в четкой формулировке цели исследований, но и в знании основных свойств оригинала. То есть перед постановкой и проведением эксперимента нужно не только формально провести его *планирование*, но и изучить объект, построить его описание и выбрать *модель*, хотя бы пробную. Если этого не делать, то можно совершить ошибку в отборе и *обработке информации*, ее оценке и прийти к выводам, прямо противоположным действительности [1, 2].

В итоге можно сказать, что целью инженерных исследований следует считать познание закономерностей, свойственных продуктам человеческой деятельности.

Математические модели часто использовались и раньше. Они позволяли анализировать недоступные или несуществующие объекты и процессы. Например, планета Уран была открыта путем анализа возмущений орбит трех планет (Леврье); К.Э. Циолковский показал, что для преодоления земного притяжения требуется первая космическая скорость, а не скорость света.

В настоящее время не существует объектов, при изучении которых не применялись бы **методы математического моделирования**. Разработаны и активно используются математические модели технических устройств, модели разнообразных технологических процессов, экономические модели предприятий, регионов и целых государств, экологические модели, модели геологических и геофизических процессов, модели социальных систем, биологические и медицинские модели.

1 МОДЕЛИ И МОДЕЛИРОВАНИЕ

Если с точки зрения целей исследования между двумя объектами есть сходство, то вместо одного можно исследовать другой. Первый называется *оригиналом*, а второй – *моделью*. Модель – это заместитель оригинала, позволяющий изучить некоторые его свойства в определенных условиях. При этом следует подчеркнуть, что сходство может быть не по всем характеристикам: форме, цвету, структуре и т.п. Достаточно, чтобы сходство было лишь в тех свойствах, которые являются объектом данного исследования.

Приведем примеры математических моделей:

- в химии: закон радиоактивного распада, скорость химической реакции;
- в физике: закон Кулона для системы зарядов, электропроводность проводника, сопротивление цепи, емкость конденсатора, освещенность;
- в биологии: закон динамики популяции (модель “хищник – жертва”).

Не существует таких моделей, которые воспроизводили бы все свойства оригинала. Во-первых, таких свойств бесконечно много, и мы многие из них даже не представляем себе. А, во-вторых, воспроизвести все свойства оригинала в состоянии только сам оригинал – как известно, даже два автомобиля различаются весьма существенно. Выбор необходимых для исследования свойств и условий дает возможность на основании предварительного изучения оригинала планомерно строить модель, удовлетворяющую поставленным целям (определенным требованиям точности, широты применения, ответа на поставленные вопросы и т.п.). Например, глобус – модель земного шара.

Моделирование – это процесс выбора или построения модели для исследования определенных свойств оригинала в определенных условиях. Моделирование – творческий процесс познания, который в первом приближении можно представить рисунком 1.1, отражающим необходимые стадии.

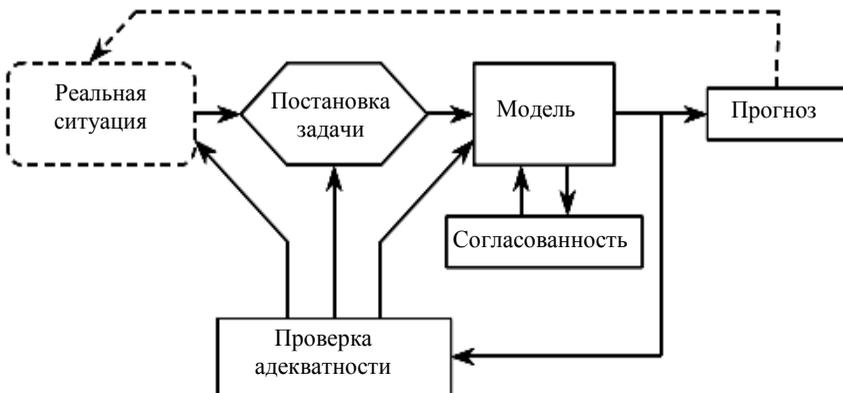


Рисунок 1.1 – Основные стадии процесса моделирования

Основная задача математического моделирования – выделение законов в природе, обществе и технике и запись их на языке математики.

Постановка задачи исходит из знаний, полученных в результате наблюдения, изучения объекта, а также из той практической проблемы, которую требуется решить. При этом из всего множества влияющих на объект факторов надо суметь отобрать **существенные** и определить **диапазоны** их изменения и **особенности** влияния на конечный результат. Это уже искусство, здесь не существует общих приемов и рекомендаций. Кроме того, к этой стадии относится оценка требуемой **точности** результатов, диктуемая целью исследования.

Пример результатов расчета программного комплекса «Кортэс-3», разработанного во ВНИИЖТ, приведен на рисунке 1.2. На основе имитационного моделирования движения поездов при заданных межпоездных интервалах и массах поездов определяются расходы и потери электроэнергии по участку, средние и максимальный токи плеч каждой подстанции, токи фидеров и температуры нагрева проводов подвески.

The screenshot shows the 'Тяговая нагрузка (тяговая нагрузка.tea)' application window. It includes a menu bar with 'Создать', 'Открыть', 'Из TRELX', and 'Сохранить'. Below the menu, there are input fields for 'Род тока: Переменный, 25 кВ', 'Период усреднения тока, мин: 1.0', and 'Принадлежность: Проектируемая дорога'. At the bottom, a table displays the results of the calculations.

| Направление | Категория поезда | Участок движения | Длина, км | Масса поезда | Время хода, мин | Расход энергии кВт·ч / кВ·А·ч | Ц, В |
|-------------|------------------|---------------------------------|-----------|--------------|-----------------|-------------------------------|------|
| ✓ Нечт. | Пассажирский | Красный берег – Жлобин Западный | 2,0 | 1100 | 2,0 | 335,4 409,2 | |
| ✓ Четн. | Пассажирский | Жлобин Западный – Красный берег | 2,0 | 1100 | 2,0 | 335,4 409,2 | |

Рисунок 1.2 – Результаты расчетов программного комплекса КОРТЭС

Проверка адекватности модели – это проверка соответствия результатов, получаемых с помощью модели, реальному поведению исследуемого объекта. На этой стадии проводится исследование и уточнение самой модели в соответствии с поставленной задачей, а также может корректироваться и постановка задачи, и общий подход к восприятию реальной ситуации.

Одно и то же явление может быть описано множеством моделей в зависимости от целей, которые ставили перед собой их авторы. Любая модель адекватна одним экспериментам и неадекватна другим. Поэтому в научных исследованиях проверка адекватности модели должна стать рутинным и удобным средством формулирования обоснованных и корректных гипотез.

Сутью решения практических прикладных задач является прогноз поведения объекта в различных ситуациях с требуемой точностью.

К построению алгоритма прогнозирования реальной ситуации в других случаях, отличающихся от исследованных во время процесса разработки модели, можно приступать только после завершения всех стадий, описанных выше. Каждая стадия этого процесса существенна. Пренебрежение любой из них может приводить к неверным выводам по существу решаемой практической задачи в результате таких ошибок, как вычисление с недопус-

тимой, неконтролируемой погрешностью, несоответствие полученных результатов поставленной задаче, неоднозначность решения, отсутствие решения (алгоритм расходится или не может завершиться).

Чем уже круг моделируемых свойств, условий и уже диапазон значений параметров, тем проще модель и легче добиться ее согласованности и адекватности, тем достовернее результаты и выводы исследования. Поэтому научные методы исследования основываются на замене оригинала моделью *в четко оговоренной области свойств и условий*, определяемой задачей исследования.

Моделирование не только удобный, но в некоторых условиях и необходимый научный прием. Среди таких особых условий можно выделить основные причины, вынуждающие применять моделирование, без которого изучение оригинала невозможно:

- сложность или дороговизна натурального исследования (например, в экономике, в экологии);
- невозможность натурального исследования по причинам аварийности или бесконечного времени ожидания результатов (например, аварийные ситуации на дороге или на железнодорожном транспорте, астрофизические явления);
- прогнозирование свойств и процессов создаваемого объекта (автомобиль, участок системы электроснабжения и т.д.)
- изучение быстропротекающих или медленных процессов (модель работы двигателя внутреннего сгорания, геологические модели).

Из всего вышесказанного следует, что любая наука представляет собой непрерывный процесс моделирования – творческий процесс познания реальности до такого уровня, который позволяет прогнозировать определенные свойства оригинала в определенных условиях.

Математическая модель – математическое представление реальности, один из вариантов модели, как системы, исследование которой позволяет получать информацию о некоторой другой системе.

В других вариантах математическая модель определяется как объект-заместитель объекта-оригинала, обеспечивающий изучение некоторых свойств оригинала, как совокупность математических соотношений, уравнений, неравенств, описывающих основные закономерности, присущие изучаемому процессу, объекту или системе [3].

Процесс построения и изучения математических моделей называется математическим моделированием.

Все естественные и общественные науки, использующие математический аппарат, по сути занимаются математическим моделированием: заменяют объект исследования его математической моделью и затем изучают последнюю. Связь математической модели с реальностью осуществляется с помощью цепочки гипотез, идеализаций и упрощений.

2 МЕТОДОЛОГИЯ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

2.1 Математические модели и их виды

Математические модели основываются на математическом описании объекта. В математическое описание входят взаимосвязи параметров объекта. Математическое описание включает в себя не только взаимосвязь элементов и параметров объекта (законы и закономерности), но и полный набор числовых и функциональных данных объекта (характеристики; начальные, граничные, конечные условия; ограничения), а также методы вычисления выходных параметров модели. То есть под математическим описанием понимается полная совокупность данных, функций и методов вычисления, позволяющая получать результат.

Со своей стороны в математическую модель может не входить часть математического описания (чаще всего – некоторые исходные данные), но в этих случаях должны присутствовать описания всех допущений, использованных для ее построения, а также алгоритмы перевода исходных и выходных данных с модели на оригинал и обратно.

В зависимости от математического описания математических моделей можно привести следующую их классификацию [4]:

1 Теоретические математические модели аналитического типа: линейные, нелинейные, в виде обыкновенных дифференциальных уравнений, в виде уравнений в частных производных, стохастические модели.

2 Эмпирические математические модели.

2.1.1 Теоретические математические модели аналитического типа

Простейшие аналитические модели могут быть заданы явно в виде функции одной или нескольких переменных. Модель, заданная в явном виде, дает исчерпывающее описание исследуемого объекта. Она позволяет построить зависимость его характеристик от управляющих факторов, взять производные и найти экстремумы модели, определить характеристики модели в окрестности экстремумов и т.д.

Очень удобна графическая интерпретация таких моделей. Однако модели в виде формул могут быть разработаны только для очень простых объектов. Примером такой модели может служить Пуассоновское распределение для k отказов в работе светофоров, поступающих на диспетчерский пульт городской системы управления дорожным движением в течение часа, представленное формулой

$$P_k = P\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Линейные математические модели. Наиболее простыми являются так называемые линейные детерминированные модели. Они задаются в виде линейной формы управляющих переменных (x):

$$W = a_0 + a_1x_1 + \dots + a_kx_k$$

при линейных ограничениях вида

$$b_{1j}x_1 + b_{2j}x_2 + \dots + b_{kj}x_k \geq b_j, \quad j = 1, \dots, q_1;$$

$$c_{1j}x_1 + c_{2j}x_2 + \dots + c_{kj}x_k = c_j, \quad j = 1, \dots, q_2;$$

$$d_{1j}x_1 + d_{2j}x_2 + \dots + d_{kj}x_k \leq d_j, \quad j = 1, \dots, q_3.$$

Общее число ограничений $m = q_1 + q_2 + q_3$ может превосходить число переменных ($m > k$). Кроме того, обычно вводится условие положительности переменных ($x_i \geq 0$).

Исследование линейных моделей не представляет труда. В частности влияние каждой из переменных на характеристики модели вида

$$W = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k$$

задается ее коэффициентами:

$$\frac{\partial W}{\partial x_i} = a_i, \quad i = 1, \dots, k.$$

Для нахождения оптимума линейной модели $W_{\text{опт}}$ разработан эффективный симплекс-метод [4].

К линейным иногда сводятся простейшие модели стоимости, рассматриваемые как совокупность производимых затрат. Примером такой модели является классическая *модель стоимости перевозок* (транспортная задача) (рисунок 2.1).

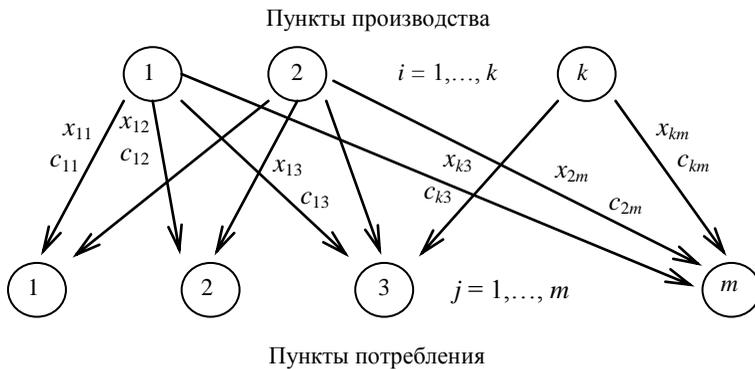


Рисунок 2.1 – Пример модели стоимости перевозок

Имеется k пунктов производства ($i = 1, \dots, k$) и m пунктов потребления ($j = 1, \dots, m$) некоторого продукта. Количество продукта, произведенного в

каждом из k пунктов производства, равно a_i ; количество продукта, необходимого в каждом из m пунктов потребления, равно b_j .

Предполагается равенство общего производства и потребления:

$$\sum_{i=1}^k a_i = \sum_{j=1}^m b_j.$$

Количество продукта, перевозимого из i -го пункта производства в j -й пункт потребления, равно x_{ij} ; стоимость перевозки единицы этого продукта – c_{ij} .

Суммарная стоимость перевозок C_Σ задается *линейной моделью*

$$C_\Sigma = \sum_{i=1}^k \sum_{j=1}^m c_{ij} x_{ij}$$

при следующих ограничениях:

$$\sum_{j=1}^m x_{ij} = a_i; \quad \sum_{i=1}^k x_{ij} = b_j; \quad x_{ij} \geq 0.$$

Нелинейные детерминированные модели. Они обладают бóльшей точностью и гибкостью. Они могут быть заданы в виде нелинейной функции одной или нескольких переменных или в виде дифференциальных уравнений (обыкновенных или в частных производных). Наиболее распространенными среди нелинейных моделей при описании дифференциальных уравнений являются:

- полиномиальные функции;
- позиномные функции;
- тригонометрические функции;
- экспоненциальные функции;
- обыкновенные дифференциальные уравнения;
- дифференциальные уравнения в частных производных др.

Нелинейные модели могут быть записаны в виде функционала, зависящего от управляющих переменных x и некоторых функций $f(x)$ всех или части этих переменных: $W = W(x, f(x))$.

Область определения модели может быть ограничена с помощью равенств или неравенств:

$$x_i = c_i, i = 1, \dots, m; \quad f_j(x) = c_j, j = 1, \dots, l; \\ x_{i \min} \leq x_i \leq x_{i \max}, i = 1, \dots, k; \quad f_j(x) \leq c_j, j = 1, \dots, n.$$

По существу, под определение нелинейной модели подпадает любое математическое описание дифференциальных уравнений, не укладывающееся в рамки простых моделей.

Полиномиальные модели. Полиномиальные модели основаны на идее приближенного представления модели конечным числом членов ряда Тейлора:

$$W(x) = W(x_0) + \sum_{i=1}^k \frac{\partial W(x_0)}{\partial x_i} (x_i - x_{i0}) + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \frac{\partial^2 W(x_0)}{\partial x_i \partial x_j} (x_i - x_{i0})(x_j - x_{j0}) + \dots$$

Наиболее простой является *квадратичная модель*

$$W(x) = a_0 + \sum_{i=1}^k a_i x_i + \sum_{\substack{i=1 \\ j \geq i}}^k a_{ij} x_i x_j$$

при ограничениях

$$\sum_{i=1}^k b_{ij} x_i \geq b_j, \quad j = 1, \dots, q_1; \quad \sum_{i=1}^k c_{ij} x_i = c_j, \quad j = 1, \dots, q_2; \quad \sum_{i=1}^k d_{ij} x_i \leq d_j, \quad j = 1, \dots, q_3.$$

Квадратичные модели широко используются для представления экспериментальных данных и аппроксимации отдельных участков поверхности отклика. Если квадратичная модель также оказывается недостаточно точной, то используются полиномиальные модели более высоких порядков. Например, полиномиальная модель применяется для реализации сред экономических прогнозов в условиях рынка [6]. Суммарное электропотребление троллейбуса для некоторых исследований [7] может описываться полиномиальной моделью следующего вида:

$$w = (\alpha C_{\text{сложн}} + C_{\text{с}} t^{\beta} + C_v v) L,$$

где α, β – коэффициенты линейной и степенной регрессии; $C_{\text{сложн}}$ – удельная сложность маршрута; $C_{\text{с}} t^{\beta}$ – удельное потребление электроэнергии, пропорциональное температуре окружающей среды; $C_v v$ – удельное потребление электроэнергии, пропорциональное средней эксплуатационной скорости; L – пробег троллейбуса.

Исследование полиномиальных моделей частично можно осуществить аналитическими методами. Например, аналитически можно определить степень влияния отдельных переменных на характеристики модели.

Позиномные модели. Они основаны на представлении модели в виде суммы произведений степенных функций:

$$W(x) = \sum_{j=1}^m c_j x_1^{\alpha_{1j}} x_2^{\alpha_{2j}} \dots x_k^{\alpha_{kj}} = \sum_{j=1}^m c_j \prod_{i=1}^k x_i^{\alpha_{ij}},$$

где x_i – управляющие переменные; α_{ij} – произвольные положительные числа; $c_j \geq 0$ – коэффициенты, обеспечивающие выпуклость модели.

Позиномные модели можно использовать для описания стоимости сложных систем. К позиномным моделям сводится задача выбора геометрических характеристик ряда технических устройств, например, электромагнитов, силовых ферм и т.д.

Исследование позиномных моделей сложнее, чем моделей полиномиального типа, и осуществляется в основном численными методами. Для поиска оптимальных решений на основе позиномных моделей разработан специальный аппарат – так называемое геометрическое программирование.

Математическая модель в виде обыкновенных дифференциальных уравнений. Математическую модель в виде одного или нескольких обыкновенных дифференциальных уравнений (ОДУ) широко применяют при изучении переходных процессов в системах автоматического регулирования, при описании процессов движения (поток, частицы, транспортные средства).

В простейшем случае модель может иметь вид линейного дифференциального уравнения n -го порядка

$$a_n \frac{d^n x(t)}{dt^n} + a_{n-1} \frac{d^{n-1} x(t)}{dt^{n-1}} + \dots + a_1 \frac{dx(t)}{dt} + a_0 x(t) = f(t)$$

или системы дифференциальных уравнений 1-го порядка

$$\frac{dx_1}{dt} = f_1(t, x_1, \dots, x_n);$$

$$\frac{dx_2}{dt} = f_2(t, x_1, \dots, x_n);$$

.....

$$\frac{dx_n}{dt} = f_n(t, x_1, \dots, x_n).$$

Модель, заданная в виде дифференциальных уравнений, должна включать в себя необходимый набор начальных условий:

$$x(0) = C_0, \quad x'(0) = C_1, \quad x''(0) = C_2, \dots, \quad x^{(n-1)}(0) = C_{n-1}$$

или

$$x_1(0) = C_1, \quad x_2(0) = C_2, \dots, \quad x_n(0) = C_n.$$

Исследование моделей, заданных в виде обыкновенных дифференциальных уравнений, осуществляется аналитическими и численными методами. Наиболее полными являются аналитические решения, обеспечивающие всесторонний анализ полученных результатов. Но такие решения получены лишь для ограниченного числа дифференциальных уравнений. Численные методы решения позволяют найти лишь конкретные значения изучаемой функции при заданной комбинации исходных данных. Для анализа модели можно использовать некоторую совокупность решений.

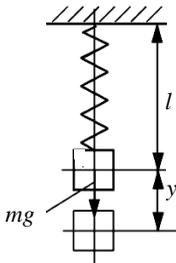


Рисунок 2.2 – Схема модели вертикальных перемещений груза на пружине

В качестве практического примера применения дифференциальных уравнений в приложении к расчету автомобильных конструкций [5] рассмотрим задачу о вынужденных колебаниях линейной системы с одной степенью свободы без учета сил сопротивления. Такой системой моделируются вертикальные перемещения груза массой m на пружине (рисунок 2.2).

Уравнение движения груза при его свободных колебаниях, согласно второму закону Ньютона, запишется в виде

$$m \frac{d^2 y}{dt^2} = mg - k(y + \delta_{ст}),$$

где y – вертикальное перемещение колеблющегося груза; t – время; $\delta_{ст}$ – удлинение пружины при действии статически приложенного груза; k – коэффициент жесткости пружины.

Модели, заданные в виде уравнений в частных производных. Ряд задач, связанных с использованием физических полей, приводит к моделям в виде дифференциальных уравнений в частных производных. Особенностью таких задач является то, что изучаемые параметры не только изменяются во времени, но и зависят от координат x , y , z рассматриваемого пространства. Такие модели называются нестационарными. Модели, в которых параметры не зависят от времени, называются стационарными. К таким моделям сводятся описания полей температур в элементах конструкции двигателя и полей скоростей при течении жидкости (газа). Уравнениями в частных производных описываются колебания элементов конструкции и поля напряжений, возникающих при работе этих элементов.

Линейное дифференциальное уравнение в частных производных имеет вид

$$a_0 \frac{\partial \Phi(t)}{\partial t} + a_1 \frac{\partial \Phi(t)}{\partial x_1} + a_2 \frac{\partial \Phi(t)}{\partial x_2} + \dots + a_k \frac{\partial \Phi(t)}{\partial x_k} = f(x_1, x_2, \dots, x_k, t).$$

Математическая модель, описанная дифференциальными уравнениями в частных производных, должна включать в себя краевые условия:

1 Должна быть задана область D , ограниченная поверхностью (на плоскости – кривой) Γ , в которой определяется решение.

2 Должны быть заданы условия на границе Γ этой области.

В случае нестационарного поля эти граничные условия, так же как и сама область, могут меняться во времени.

Граничные условия могут быть 1, 2 и 3-го рода. Граничные условия *1-го рода* предусматривают задание на границе величины искомой функции:

$\Phi|_{\Gamma} = f_1(\Gamma)$ – для стационарного поля; $\Phi(t)|_{\Gamma} = f_1(\Gamma, t)$ – для нестационарного поля. Граничные условия *2-го рода* предусматривают задание производной

искомой функции: $\left. \frac{\partial \Phi}{\partial x_i} \right|_{\Gamma} = f_2(\Gamma)$ – для стационарного; $\left. \frac{\partial \Phi}{\partial x_i}(t) \right|_{\Gamma} = f_2(\Gamma, t)$ – для нестационарного поля. Граничные условия *3-го рода* предусматривают комбинации функции и ее производной:

$\left. \left(a\Phi + b \frac{\partial \Phi}{\partial x_i} \right) \right|_{\Gamma} = f_3(\Gamma)$ – для стационарного поля;

$\left. \left(a(t)\Phi(t) + b(t) \frac{\partial \Phi}{\partial x_i}(t) \right) \right|_{\Gamma} = f_3(\Gamma, t)$ – для нестационарного поля.

3 Для нестационарных полей должны быть заданы одно или два начальных условия, характеризующих состояние поля в начальный момент времени:

$$\begin{aligned} \Phi(x_i)|_{t=0} &= f_4(x_i); \\ \frac{\partial \Phi}{\partial t}(x_i)|_{t=0} &= f_5(x_i), \end{aligned} \quad (i = 1, 2, 3).$$

Здесь x_i – координаты пространства.

Совокупность уравнений и краевых (и начальных) условий полностью определяет модель и позволяет провести ее исследование.

Стохастические модели. Точные величины и зависимости, используемые в детерминированных моделях, представляют собой лишь некоторые средние значения (математические ожидания) реальных случайных величин (зависимостей). Так, физические константы, характеризующие материалы и рабочие тела (предел прочности материала σ , теплопроводность λ , плотность ρ и т.д.) меняются в зависимости от партии материала и условий окружающей среды. Всегда имеется определенный разброс размеров деталей l , расходов топлива, сопротивления токоведущих частей электротранспорта и токоприемника с контактными проводами и т. д. Все это приводит к тому, что и результирующие функции, характеризующие процесс, также носят случайный характер. Результаты, полученные с помощью детерминированной модели, представляют собой математические ожидания этих характеристик. При этом конкретные данные для конкретной системы могут существенно отличаться от этих математических ожиданий. Например, ресурс конкретного двигателя может существенно отличаться от среднего ресурса двигателей данного типа, сроки замены ламп освещения могут значительно не совпадать с нормативным сроком эксплуатации. Для учета таких отличий вводятся всевозможные «запасы прочности», призванные гарантировать работоспособность реальных объектов при неблагоприятном стечении обстоятельств.

Значительно более полные и объективные результаты можно получить при переходе от детерминированных к *стохастическим моделям*, то есть при переходе от точно заданных величин к соответствующим случайным величинам. При этом константы (σ , λ , ρ , l , ...) заменяются случайными величинами ξ_σ , ξ_λ , ξ_ρ , ξ_l , ... , подчиненными определенным законам распределения.

Однократное исследование стохастической модели приведет к некоторой случайной величине функции отклика ξ_w . Для получения значимых результатов необходимо провести многократное исследование модели и получить распределение результирующей характеристики в интересующем исследователя диапазоне.

Такой метод исследования стохастической модели получил название метода статистических испытаний, или *метода Монте-Карло*.

Трудоемкость исследования стохастических моделей существенно выше, чем моделей детерминированных:

1 Значительно возрастает объем исходной информации: замена констант случайными величинами, введение законов распределения этих величин усложняют модель.

2 Для получения распределения результирующей функции необходимо многократное исследование модели.

С другой стороны, полученное при статистическом моделировании распределение характеристик системы дает в руки исследователя чрезвычайно ценную информацию: такое распределение позволяет оценить не только среднее значение изучаемой величины, но и разброс этих значений, вероятности появления тех или иных значений при конкретном испытании (например, вероятность выхода из строя системы электроснабжения предприятия) и их зависимость от различных факторов.

Выбор закона распределения для конкретной случайной величины, входящей в стохастическую модель, может быть обоснован экспериментально или теоретически.

Конкретные параметры распределения (μ, σ, \dots) всегда определяются на основе экспериментальных данных. Оценка параметров нормального распределения на основе выборки $\{x_i\}$ из n случайных значений величины x дается соотношениями

$$\mu = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i ; \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} .$$

При использовании метода статистических испытаний характеристики изучаемой системы оцениваются на основе некоторой ограниченной выборки реализаций. Поэтому важно определить *достоверность* этой оценки.

Вероятность p пребывания системы в некотором состоянии (например, вероятность того, что время работы элемента рулевого гидроусилителя до первого отказа составит не менее t часов), определяется частотой этого события при моделировании:

$$p \approx \frac{n_+}{n} ,$$

где n_+ – число реализаций, при которых наблюдалось изучаемое состояние системы; n – общее число реализаций.

Эта оценка является приближенной, так как определяется на основе ограниченной выборки. Отношение n_+/n называется *выборочной статистикой*.

Ошибка моделирования определяется отклонением выборочной статистики от вероятности

$$\delta = \left| \frac{n_+}{n} - p \right|.$$

Можно показать, что эта ошибка удовлетворяет неравенству

$$\delta \leq \sqrt{\frac{p(1-p)}{\alpha n}}.$$

Здесь p – вероятность рассматриваемого состояния; α – вероятность невыполнения оценки (*уровень риска*). Доверительная вероятность выполнения этой оценки равна $1 - \alpha$.

Погрешность стохастического моделирования обратно пропорциональна \sqrt{n} . То есть увеличение точности при стохастическом моделировании требует значительного увеличения числа реализаций.

Для уменьшения погрешности в 10 раз необходимо увеличить число реализаций (а значит и время счета) в 100 раз. Поэтому метод статистических испытаний не может дать решения с очень высокой степенью точности. Считается, что допустимая ошибка может составлять 1–5 % максимальной величины, полученной при моделировании.

Величина ошибки зависит также от вероятности p оцениваемого состояния и допустимого уровня риска α . Обычно α задают на одном из фиксированных уровней ($\alpha = 0,005; 0,01; 0,025; 0,05; 0,1 \dots$).

2.1.2 Эмпирические математические модели

Идентификация эмпирических математических моделей. Переход к эмпирическим моделям предполагает заведомый отказ от аналитических методов исследования. Поэтому эмпирические модели более разнообразны и включают в себя различные по форме математические зависимости.

При построении эмпирической модели обычно предполагается, что физическая теория работы объекта отсутствует или по тем или иным причинам не может быть использована.

При разработке эмпирической математической модели предполагается использование экспериментальных данных, полученных при испытаниях объектов. Результаты таких испытаний всегда представляют собой наборы величин, характеризующих работу объекта или системы при различных сочетаниях управляющих параметров.

Объект идентификации представляет собой так называемый «черный ящик» с некоторым числом регулируемых (или, по крайней мере, измеряемых) входов x и одним или несколькими наблюдаемыми (измеряемыми) выходами (рисунок 2.3).

Задачей идентификации является построение модели объекта по результатам наблюдений его реакции на возмущения внешней среды. При этом необходимо учитывать ошибки, возникающие при измерении характеристик объекта.

Требуется построить зависимость (модель) $W = f(x)$, которая описывает характеристики изучаемой системы. Это уравнение называется *уравнением регрессии*.

Идентификацию модели начинают с выбора формы модели, т.е. вида функции $f(x)$. При этом на практике может встретиться два случая:

1) форма математической модели известна заранее, а задача идентификации сводится к определению

коэффициентов этой модели. Так, описание ряда затухающих или развивающихся процессов дается зависимостями экспоненциального типа. Задачей исследования является определение коэффициентов α, β ;

2) форма математической модели заранее неизвестна. В этом случае для идентификации модели используются отрезки бесконечных рядов, а задача заключается в определении числа членов ряда и коэффициентов при этих членах. Модель может быть представлена в виде

$$W = \sum_{i=1}^k \beta_{0i} f_0(x_i) + \sum_{i=1}^k \beta_{1i} f_1(x_i) + \dots + \sum_{i=1}^k \beta_{li} f_l(x_i),$$

где $f_q(x_i)$ – некоторые заданные функции; β_{qi} – коэффициенты регрессии; $q = 0, 1, \dots, l$.

Конкретный вид модели зависит от выбора функций $f_q(x)$, по которым производится разложение W . Например, при описании колебательных процессов удобно использовать ряд Фурье $W = \alpha_0 + \sum_{k=1}^{\infty} (\alpha_k \cos kx + \beta_k \sin kx)$.

Часто в качестве функций $f_0(x), f_1(x), f_2(x), \dots, f_l(x)$ выступают степенные функции $x^0, x^1, x^2, \dots, x^l$. Если ограничиться первыми членами разложения, то уравнения сведутся к линейным, квадратичным и другим полиномиальным моделям. Однако пока остается неясным, сколько членов ряда обеспечивает наилучшее описание изучаемого процесса.

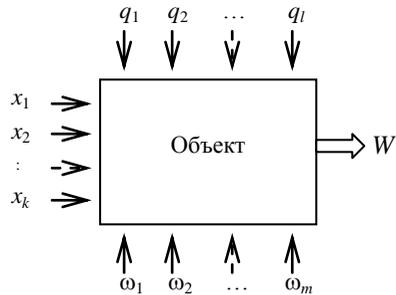


Рисунок 2.3 – Объект исследования – «черный ящик»:
 q_i – ограничения;
 x_i – управляющие переменные;
 ω_i – неопределенности (шумы);
 W – характеристическая функция

Обычно берут количество экспериментальных точек значительно больше, чем количество коэффициентов регрессии. В этом случае нельзя построить поверхность отклика, проходящую через все экспериментальные

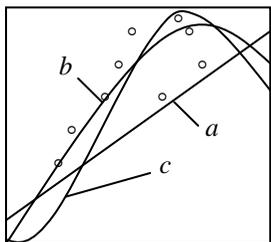


Рисунок 2.4 – Пример выбора уравнения регрессии

точки. Да этого и не требуется. При этом, однако, можно построить приближенную модель, обеспечивающую в некотором смысле наилучшее совпадение с экспериментальными данными. Например, прямая *a* построена по экспериментальным точкам, часть которых представлена на рисунке 2.4, методом наименьших квадратов; кривая *b* – квадратичная модель; *c* – полиномиальная модель 3-го порядка достаточно хорошо соответствует исходному экспериментальному материалу, хотя проходит не через все экспериментальные точки.

Таким образом, для любой экспериментальной выборки могут быть предложены различные модели идентификации. Конкретная форма модели зависит от выбора функций $f_j(x)$ и количества членов ряда.

Сама постановка задачи идентификации включает в себя элемент неопределенности, возможность множественности решений. Важно выбрать лучшее или, по крайней мере, достаточно хорошее из этих решений.

Для оценки точности модели естественно использовать величины отклонений, полученных в эксперименте величин W_j и их оценок Wm_j , предсказанных моделью

$$\varepsilon_j = W_j - Wm_j.$$

Исключительное распространение получил метод наименьших квадратов отклонений реальных значений оцениваемой величины от значений, предсказанных моделью.

Специальные *методы планирования эксперимента* позволяют существенно повысить объем получаемой информации, улучшают характеристики эмпирических моделей, а также упрощают процедуру обработки экспериментальных данных. Однако на практике очень часто приходится иметь дело с неорганизованным (пассивным) экспериментом. Связано это, по крайней мере, с тремя причинами [4]:

1) исследователь может только наблюдать входы системы, но не может их регулировать, что полностью исключает возможность планирования эксперимента (типичная ситуация: астроном, исследующий Галактику);

2) неизвестны диапазоны возможного изменения переменных (входов), что затрудняет планирование эксперимента и исключает возможность использования ряда эффективных методов планирования;

3) приходится строить модели идентификации на основе уже полученных ранее беспорядочных экспериментальных данных.

Выбор оптимальной эмпирической модели. Принцип наименьших квадратов позволяет найти наилучшую модель идентификации для исследуемой экспериментальной выборки с заданным уравнением регрессии вида

$$W = \sum_{i=1}^k \beta_{0i} f_0(x_i) + \sum_{i=1}^k \beta_{1i} f_1(x_i) + \dots + \sum_{i=1}^k \beta_{li} f_l(x_i).$$

Если имеются достаточно веские основания для выбора формы этого уравнения, никаких проблем не возникает. Однако в большинстве случаев конкретная форма модели заранее неизвестна и может, вообще говоря, быть различной.

На первый взгляд может показаться, что более сложная модель (увеличение степени полинома) всегда обеспечивает получение большей точности. На самом деле это не так. При переходе к полиномам более высокой степени можно, конечно, получить лучшее согласие регрессионной кривой с экспериментальными данными. Для $m = n$ это согласие будет абсолютным, но при этом получится худшее согласие с истинным характером процесса $W(x)$. Дело в том, что экспериментальные данные представляют собой случайные величины и содержат лишь ограниченную информацию о характере $W(x)$. Увеличение степени полинома целесообразно лишь до тех пор, пока из экспериментальной выборки извлекается надежная информация. Таким образом, возникает проблема выбора формы модели, что требует понимания сущности проблемы исследования.

Подход к решению этой проблемы основан на статистическом исследовании уравнений регрессии.

1 *Метод всех возможных регрессий* основан на последовательном изучении всех возможных моделей ($m < n$), из которых отбирается лучшая модель.

Метод представляется малоприменимым для анализа сложных систем, так как отличается высокой трудоемкостью.

2 *Метод исключения* предполагает исследование наиболее полной (в пределах разумного) модели и последовательную проверку на значимость всех ее членов. При этом для каждого из членов модели вычисляется величина критерия Фишера F . На основе полученного множества $\{F_i\}$ выбирается член уравнения регрессии, соответствующий минимальному значению критерия F_i . Если это минимальное значение меньше критического при выбранном уровне риска ($F_i < F_{кр \alpha}$), то соответствующий член исключается из регрессионного уравнения как несущественный, после чего все коэффициенты регрессии пересчитываются заново и вновь осуществляется проверка их значимости.

Если $F_i > F_{кр \alpha}$, то все члены модели существенны и уравнение регрессии остается в первоначальном виде. Однако, если это произошло уже на первом шаге исследования, стоит рассмотреть целесообразность усложнения первоначальной модели.

Трудоёмкость этого метода меньше, чем метода всех возможных регрессий.

3 *Метод включений* по существу противоположен методу исключений и предусматривает последовательное включение в модель новых членов с проверкой их статистической значимости. Трудоёмкость этого метода существенно меньше трудоёмкости рассмотренных выше методов.

Существуют и некоторые другие методы подбора оптимального уравнения регрессии.

Общим *недостатком* всех рассмотренных ранее методов является использование для оценки модели того же экспериментального материала, на основе которого эта модель построена.

4 Иной подход основан на использовании *регуляризации*. При этом подходе все экспериментальные данные разбиваются на две части: *обучающую* (n_1) и *проверочную* (n_2). Первая из них используется для определения коэффициентов регрессии модели, вторая – для оценки модели в целом. Оптимальные по этому подходу модели мало чувствительны к небольшим изменениям исходных данных.

Число точек обучающей последовательности должно быть, по крайней мере, на единицу больше числа коэффициентов регрессии ($n_1 > m + 1$). Для повышения достоверности результатов этот запас должен быть существенно увеличен ($n_1 \geq (2 \dots 3)m$). Проверочная последовательность должна включать в себя хотя бы одну точку.

В ряде случаев в качестве критерия регуляризации удобно использовать *критерий несмещенности*, обеспечивающий наименьшее изменение модели при изменении состава обучающей последовательности. При этом весь экспериментальный массив разбивается на две одинаковые по величине последовательности ($n_1 = n_2$), каждая из которых поочередно используется в качестве обучающей. В результате их использования определяются две независимые, одинаковые по форме модели $W_m^*(n_1)$ и $W_m^{**}(n_2)$. Оптимальная модель ищется *по всем точкам выборки*.

Критерий регуляризации всегда имеет четко выраженный минимум, что обеспечивает объективное выделение модели *оптимальной сложности*.

Вычислительные эксперименты, проводимые на основе математических моделей, помогают увидеть за частным общее, развить универсальные методы анализа объектов различной физической природы, познать свойства изучаемых процессов и систем.

2.2 Адекватность математических моделей

Особенностью математических моделей является то, что получение с их помощью каких-либо результатов связано с вычислениями. Так возникает необходимость понятия вычислительного эксперимента. Вычислительный эксперимент – это получение результатов с помощью математической модели для какого-либо конкретного случая исследований. Это может быть как единственный расчет одного параметра, так и комплекс расчетов целого спектра параметров модели во множестве определенным образом связанных условий. Во втором случае большое значение приобретает процедура планирования вычислительного эксперимента, целью которого является получение максимума достоверной информации при минимуме затрат. Под достоверностью результата вычислительного эксперимента понимается одновременное выполнение двух условий: во-первых, результат должен быть достаточно точен, а во вторых, не может быть опровергнут с помощью каких либо дополнительных расчетов. В математической статистике этим понятиям соответствуют понятия несмещенности и состоятельности оценок, получаемых из наблюдений. При планировании вычислительного эксперимента используются многие методы математического моделирования – от простого здравого смысла до теории катастроф и методов математической статистики.

Центральным понятием теории математического моделирования является понятие адекватности, так как аргументированная проверка адекватности обеспечивает получение практически значимых результатов.

Адекватность математической модели – это соответствие результатов вычислительного эксперимента поведению реального объекта. Это соответствие следует оценивать с точки зрения целей исследования. Поэтому возможны различные подходы к оценке адекватности различных моделей.

Для выявления этого соответствия для механических систем и процессов, характеризующихся измеримыми величинами – параметрами, необходимо провести сравнение параметров модели и оригинала в одних и тех же условиях. Очевидно, что сравнивать следует лишь соответствующие друг другу параметры между собой и только в той области функционирования объекта, в которой предполагается его исследовать.

Для адекватности математической модели поведению оригинала – механической системы – достаточно убедиться в выполнении двух свойств: точности и непротиворечивости. Однако так звучат лишь общие, образные требования к адекватности, для практического применения необходимо сформулировать математические формы этих требований.

Точность в задачах механики означает, что обобщенная характеристика рассогласования соответствующего параметра модели и оригинала ($\Delta u = u_{\text{модели}} - u_{\text{оригинала}}$) должна быть не больше, чем заранее заданное зна-

чение приемлемой погрешности $\Delta u_{\text{доп}}$. В качестве такой обобщенной характеристики может выступать наибольшее по модулю значение рассогласования, среднее значение рассогласования или статистическая оценка, как, например:

- доверительный интервал для математического ожидания рассогласования модели и оригинала;
- диапазон практически наблюдаемых значений рассогласования;
- интегральная оценка одного из следующих типов.

Хотелось бы отметить то, что критерии проверки точности не должны рассматриваться как догма, они выбираются в соответствии с целью исследований.

Непротиворечивость подразумевает идентичный характер изменения соответствующих параметров, т.е. идентичный вид основных свойств функциональных зависимостей на отдельных участках, как-то: возрастание, убывание, экстремумы, выпуклость и т.п.

Какие-либо выводы о соответствии их поведения можно сделать только на основании статистической обработки. Поэтому адекватность проверяется с помощью статистических критериев, которые могут с определенной вероятностью свидетельствовать о соответствии результатов вычислительного эксперимента поведению реального объекта в соответствующих условиях.

Для образной характеристики понятий точности и непротиворечивости можно воспользоваться рисунком 2.5. На нем изображены графики некоторой функциональной зависимости между параметрами оригинала, которую модель должна адекватно воспроизвести. Для первого знакомства с понятием адекватности нижеследующий анализ приводится в нестрогой форме – строгий математический аппарат проверки адекватности дан в виде алгоритма в отдельной главе. В случае "а" существует область, в которой выполняются некоторые заданные требования точности, т.е. погрешность модели по отношению к оригиналу меньше некоторого допустимого значения. Однако с точки зрения такого свойства рассматриваемой зависимости, как возрастание-убывание, эта модель противоречит поведению оригинала, поэтому не может быть признана адекватной. Случай "б" демонстрирует непротиворечивый ход зависимости с той же точки зрения.

Колебания на графиках "в" и "г" рисунка 2.5 связаны с возмущающими факторами, не поддающимися регистрации, а также с погрешностями записывающей аппаратуры. Тем не менее, заменять экспериментальную зависимость более "красивой" нельзя, так как истинный характер ее неизвестен. В этом случае сравнение оригинала и модели особенно сложно. В случае "в" заметна систематическая погрешность модели – постоянно присутствующее рассогласование между параметрами модели и оригинала.

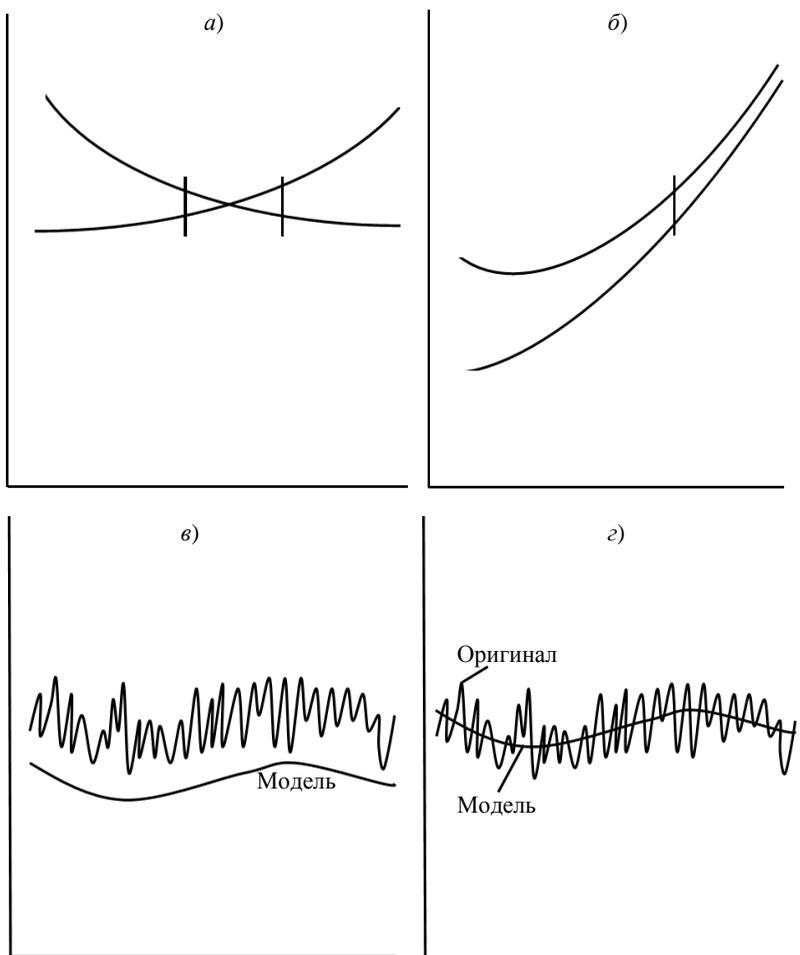


Рисунок 2.5 – Образная характеристика понятий точности и непротиворечивости

Если все наблюдаемые частные значения рассогласования *существенно* меньше допустимого значения погрешности, то модель можно считать достаточно точной. Если большое число наблюдаемых частных значений рассогласования больше допустимого значения погрешности, то модель нельзя считать достаточно точной. А в промежуточном случае необходимо руководствоваться соображениями цели исследований.

В случае "г" систематическая погрешность модели значительно меньше той случайной ее составляющей, которая обязана своим появлением возмущающим факторам. Поэтому, если большинство наблюдаемых частных зна-

чений рассогласования меньше допустимого значения погрешности, то модель можно считать достаточно точной.

Что касается свойства непротиворечивости модели в случаях "в" и "з", то этот вопрос значительно сложнее. Если по своей природе исследуемая зависимость должна быть более плавной, чем это зарегистрировано на оригинале, то это значит, что практически все высокочастотные колебания являются результатом наложения шума (неучитываемых факторов), который следует отфильтровать. Эта неформализуемая процедура должна быть построена только на одном требовании: для непротиворечивости рассогласование между оригиналом и моделью не должно подчиняться какой-либо закономерности, рассогласование должно вести себя вполне хаотически. С этой точки зрения случай "з" позволяет надеяться на непротиворечие модели поведению оригинала, а случай "в" – нет.

Таким образом, становится очевидным, что для проверки адекватности необходимо иметь:

- исчерпывающую информацию о реальном случае (что всегда трудно, а подчас бывает практически невозможно);
- результаты контрольного вычислительного эксперимента, воспроизводящего известный реальный случай;
- критерий оценки точности математической модели;
- критерий проверки непротиворечивости математической модели.

При построении критерия проверки адекватности необходимо учитывать как особенности модели, так и область ее применения:

- ограниченность допустимого диапазона изменения параметров системы,
- соответствие математического описания условий реального и вычислительного экспериментов,
- возможную неоднозначность решений вычислительного эксперимента,
- точность самого вычислительного эксперимента.

Поясним это на примере. Если предполагается исследовать начальную скорость движения транспортного средства по длине следа торможения, то нет необходимости добиваться адекватности моделирования таких явлений, как время реакции водителя и срабатывания тормозного привода, которые принимаются по справочникам.

Точность модели определяется погрешностью – рассогласованием значений рассматриваемого параметра u :

– абсолютная погрешность $\Delta u = u_{\text{модели}} - u_{\text{оригинала}}$,

– относительная погрешность $\delta u = \frac{\Delta u}{u_{\text{оригинала}}} \cdot 100\%$,

– относительная приведенная погрешность $\delta u = \frac{\Delta u}{u_{\text{меры}}}$ (где $u_{\text{меры}}$ – неко-

торое характерное значение, например $u_{\text{меры}} = |u|_{\text{max}}$).

Погрешности получили следующие эпитеты:

- грубая – недопустимая с точки зрения целей исследования;
- удовлетворительная – допустимая с точки зрения целей исследования;
- случайная – принимающая случайные значения при многократном повторении опыта в неизменных условиях (например, замер времени бега на 100 метров с помощью одного и того же секундомера);
- систематическая – принимающая неизменное значение при многократном повторении опыта в неизменных условиях (то же, что в предыдущем случае, но с испорченным секундомером, который начинает отсчет времени на 0,1 с позже старта).

При математическом моделировании возможны погрешности, обусловленные различными причинами:

- погрешности физической абстракции (неточность физических законов и закономерностей, неучет некоторых факторов);
- погрешности математического описания:

 - приближенность уравнений, приближенность данных,
 - погрешность расчетов (погрешность установок, компьютерных программ, приближенные методы расчетов);
 - погрешность обработки результатов (округление результатов, графическое изображение).

Из всех перечисленных причин в пояснении нуждается лишь погрешность расчетов, которую при моделировании всегда надо учитывать.

Выясним особенности приближенных вычислений, влияющих на погрешность расчетов с помощью математических моделей. Будем определять погрешность результатов при известных погрешностях операндов Δa и Δb .

1 Погрешности суммы:

$$(a + \Delta a) + (b + \Delta b) = (a + b) + (\Delta a + \Delta b),$$

т.е. **абсолютная погрешность суммы**

$$\Delta(a + b) \leq |\Delta a| + |\Delta b|;$$

ограничена суммой модулей абсолютных погрешностей слагаемых, а относительная

$$\delta(a + b) = \frac{\Delta a + \Delta b}{a + b} = \frac{\delta a a + \delta b b}{a + b},$$

следовательно, $\min(\delta a, \delta b) < \delta(a + b) < \max(\delta a, \delta b)$, т. е. относительная погрешность суммы принимает значение между наибольшей и наименьшей относительными погрешностями слагаемых.

2 Погрешности разности:

$$(a + \Delta a) - (b + \Delta b) = (a - b) + (\Delta a - \Delta b),$$

т.е. **абсолютная погрешность разности**

$$\Delta(a-b) \leq |\Delta a| + |\Delta b|$$

ограничена суммой модулей абсолютных погрешностей операндов, а относительная

$$\delta(a-b) = \frac{\Delta a - \Delta b}{a-b} = \frac{\delta a a - \delta b b}{a-b},$$

т.е. относительная погрешность разности принимает значения больше относительных погрешностей операндов, а при близких их значениях – не ограничена. Из этого следует, что в приближенных вычислениях необходимо **избегать** разности близких величин, что особенно важно учитывать при программировании алгоритмов.

3 Погрешности произведения:

$$(a + \Delta a) \times (b + \Delta b) = (a \times b) + (\Delta a \times b + \Delta b \times a) + (\Delta a \times \Delta b),$$

а если предполагать малость абсолютных погрешностей по сравнению со значениями самих величин, то **абсолютная погрешность произведения**

$$(a + \Delta a) \times (b + \Delta b) = \Delta a \times b + \Delta b \times a$$

приближенно равна сумме перекрестных произведений абсолютных погрешностей сомножителей на смежные сомножители, а относительная

$$\delta(a \times b) = \frac{\Delta a \times b - \Delta b \times a}{a \times b} = \delta a - \delta b,$$

т.е. относительная погрешность произведения приближенно равна сумме относительных погрешностей сомножителей.

4 Погрешности деления:

$$\frac{a + \Delta a}{b + \Delta b} = \frac{\Delta a \times b}{(b + \Delta b)^2} + \frac{\Delta a \times b + a \times \Delta b}{(b + \Delta b)^2} + \frac{\Delta a \times \Delta b}{(b + \Delta b)^2},$$

т.е. **абсолютная погрешность частного**

$$\Delta\left(\frac{a}{b}\right) \approx \frac{\Delta a \times b + a \times \Delta b}{(b + \Delta b)^2}$$

приближенно равна сумме произведений абсолютной погрешности делимого на делитель и абсолютной погрешности делителя на делимое, деленной на квадрат делителя, а относительная

$$\delta\left(\frac{a}{b}\right) \approx \frac{\Delta a \times b + \Delta b \times a}{(b + \Delta b)^2} \frac{a}{b} \approx \delta a + \delta b,$$

т.е. относительная погрешность частного приближенно равна сумме относительных погрешностей делимого и делителя.

5 Погрешности вычисления функции $y = f(x_1, x_2, \dots, x_n)$ в предположении разложимости ее в ряд Тейлора по степеням x_n и малости абсолютных погрешностей Δx_n по сравнению со значениями x_n в первом (линейном) приближении:

$$f(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) \approx f(x_1, x_2, \dots, x_n) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Delta x_i,$$

т.е. абсолютная погрешность определится величиной

$$f(x_1, x_2, \dots, x_n) \approx \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Delta x_i \leq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i} \right| \Delta x_i,$$

а относительная –

$$\delta f(x_1, x_2, \dots, x_n) \approx \frac{1}{|f(x_1, x_2, \dots, x_n)|} \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Delta x_i = \sum_{i=1}^n \left| x_i \frac{\partial \ln f}{\partial x_i} \right| \delta x_i.$$

6 Погрешность методов вычисления в более сложных случаях связана с применяемым алгоритмом. Поэтому для обеспечения возможности контроля погрешности методы должны обладать свойствами аппроксимации и устойчивости.

2.3 Алгоритм научных исследований с помощью математического моделирования

Математическое моделирование – мощное современное средство научных исследований, и его применение требует соблюдения определенной строгости во избежание получения неверных выводов. Теперь можно обоснованно дать следующий выработанный практикой *алгоритм действий*, которого рекомендуется придерживаться:

1) изучение оригинала: выявление основных факторов, особенностей, диапазонов исследуемых параметров, условий и задач исследования, постановка (формулировка) задачи исследования, оценка требуемой точности;

2) феноменологическое описание оригинала ("физическое" описание): поиск аналогий и функциональных зависимостей на основе предыдущего этапа и достижений в различных областях науки;

3) математическое описание оригинала;

4) разработка алгоритмического и программного обеспечения для реализации математического описания на компьютере;

5) проведение контрольного вычислительного эксперимента (воспроизводящего реальный известный случай поведения оригинала в конкретных условиях);

6) оценка адекватности результатов контрольного вычислительного эксперимента реальному случаю; при необходимости – повторение алгоритма с п. 3, 2 или 1;

7) *планирование* вычислительного эксперимента в целях исследования;

8) проведение вычислительного эксперимента в целях исследования, обработка его результатов;

9) анализ результатов вычислительного эксперимента, сравнение с результатами изучения оригинала (при необходимости – повторение алгоритма с п. 7 или 1);

10) формулировка выводов исследования.

Пункты 1–6 составляют процесс моделирования – построения математической модели. В нем можно выделить процесс идентификации, объединяющий пп. 3–6.

По такому алгоритму проведены многочисленные исследования, например: – вычислительные финансы (оптимизация портфеля, оценка риска, экономическое прогнозирование, рисунок 2.6) [9];

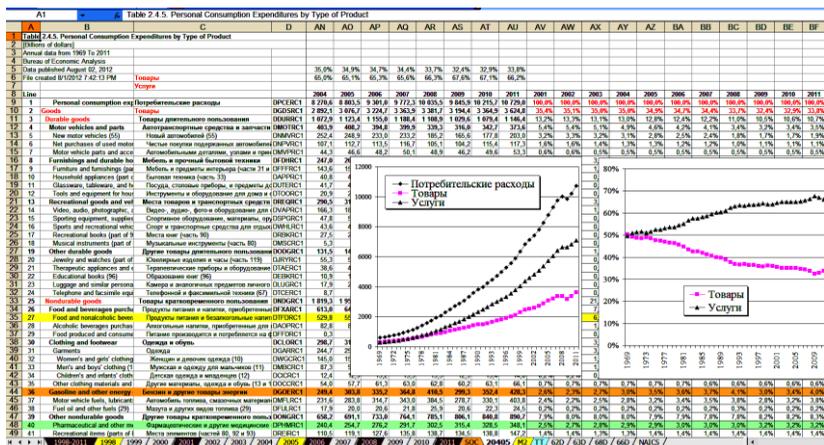


Рисунок 2.6 – Моделирование портфеля потребления домашних хозяйств США – физическое моделирование (моделирование механических, электрических, гидравлических систем (рисунок 2.7) [10];



Рисунок 2.7 – Пример моделирования в AnyLogic пассажиропотока на вокзалах к Олимпиаде "Сочи-2014"

– электрические схемы и системы электроснабжения (рисунок 2.8);

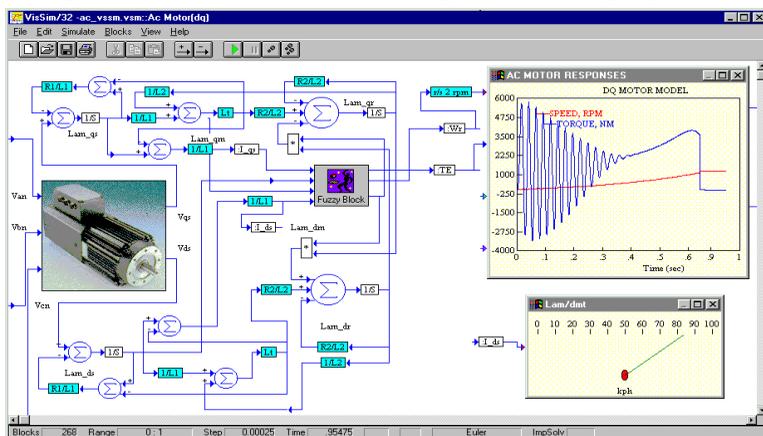


Рисунок 2.10 – Модель электрической системы

– автомобилестроение (моделирование и калибровка трансмиссии, раследование ДТП, рисунки 2.9–2.11 [24, 25]).

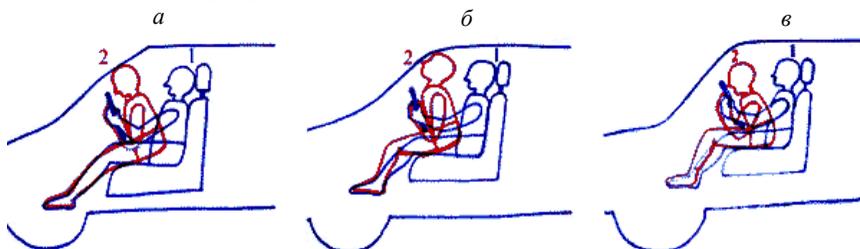


Рисунок 2.9 – Схема динамики перемещения тела водителя в салоне автомобиля при «среднем» (а), «высоком» (б) и «низком» (в) водителе (1 – до столкновения, 2 – в момент столкновения)

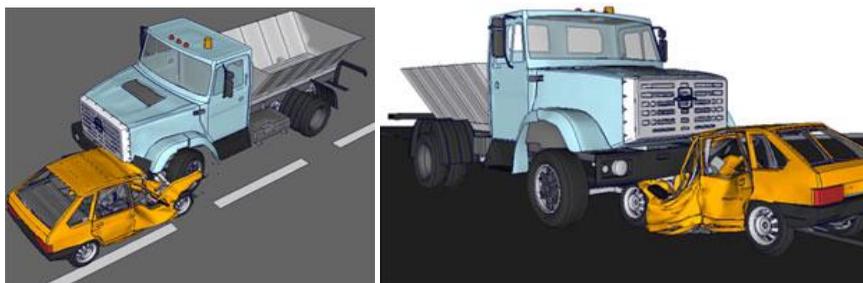


Рисунок 2.10 – Пример компьютерного моделирования лобового столкновения автомобилей ВАЗ-21093 и ЗИЛ-433362 лаборатории «Вычислительная механика»



Рисунок 2.11 – Пример моделирования в программе реконструкции и анализа дорожно-транспортных происшествий EDCRASH

Основные принципы математического моделирования. В заключение сформулируем все те "правила" строгости процесса моделирования в виде принципов математического моделирования.

1 Главным из этих принципов безусловно является обеспечение высокой степени *адекватности математической модели*. В теории математического моделирования принято называть моделью только тот объект, который успешно прошел оценку адекватности. Адекватность математической модели механических систем и процессов основывается на удовлетворительной *точности и непротиворечивости* по отношению к поведению оригинала.

2 Принцип *состоятельности результатов вычислительного эксперимента* трактуется, как обеспечение результатов, безусловно приближающихся к истине. Состоятельность здесь следует понимать как статистический термин, обозначающий стремление по вероятности при увеличении объема информации результатов вычислительного эксперимента к истинным значениям параметров исследуемого явления. Этот принцип требует предельной математической строгости, то есть использования в программном обеспечении вычислительных методов, проявляющих при их применении одновременно устойчивость, сходимую и однозначность.

3 Принцип *удобства исследователя* – простота обращения с программным обеспечением, компоновки вариантов расчета, обработки и представления результатов вычислительного эксперимента – все это достигается развитым диалоговым режимом работы, сервисным программным обеспечением (таблицы, графики и т.п.) и унификацией всего программного обеспечения. Обычно выделяемые принципы математического моделирования: гибкость, инвариантность и динамичность – сводятся в основном к полной унификации всего программного обеспечения.

4 Принцип *планирования вычислительного эксперимента* обеспечивается применением методов и приемов планирования эксперимента.

5 Приоритет *физически обоснованных закономерностей* перед статистическим моделированием и приемами упрощения моделей.

3 МЕТОДЫ РАЗРАБОТКИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

Можно построить очень сложную математическую модель, учитывающую все видимые и предполагаемые факторы и явления, но получение результата с ее помощью может оказаться не менее сложным, чем на оригинале. Можно построить очень простую модель, отображающую минимум очевидных свойств объекта, но тогда нельзя с ее помощью исследовать тонкие свойства. Задача построения математической модели – это отыскание оптимального компромисса между простотой модели и степенью ее адекватности изучаемому оригиналу.

Построение математической модели (синтез математической модели) требует решения достаточно сложных проблем, среди которых:

- множественность критериев оценки качества функционирования моделируемой системы (многокритериальность);
- большая размерность описания сложных систем ("проклятие размерности");
- адекватность.

Под многокритериальностью понимается наличие подчас противоречивых требований к различным элементам сложной системы или к системе в целом (например, экономичность и безопасность пассажирских перевозок, быстрота и качество обслуживания). Для решения этой проблемы применяют различные приемы ранжирования, в том числе и основанные на результатах применения методов экспертных оценок.

С "проклятием размерности" борются тоже ранжированием, а также агрегированием, что позволяет решать задачу поблочно (поагрегатно). Наиболее сложной при этом остается задача выявления факторов, способных описать изучаемое явление, а также взаимосвязи различных факторов, входных и выходных данных системы. Для этого помимо глубокого изучения физических особенностей системы подчас бывает необходимо проводить многомерный статистический анализ результатов экспериментов (вычислительных или натуральных).

Приведем краткую характеристику методов математического моделирования.

Ранжирование – неформализуемый анализ, в результате которого можно произвести распределение параметров по важности (рангу); наиболее важные необходимо учитывать, наименее важными иногда можно пренебречь, промежуточные по важности можно учесть в виде поправок, каждому из них можно приписать весовые коэффициенты. Например из рисунка 3.1 можно сделать вывод, что с 2009 года добыча золота является абсолютным лидером по темпам роста энергопотребления в Российской Федерации.

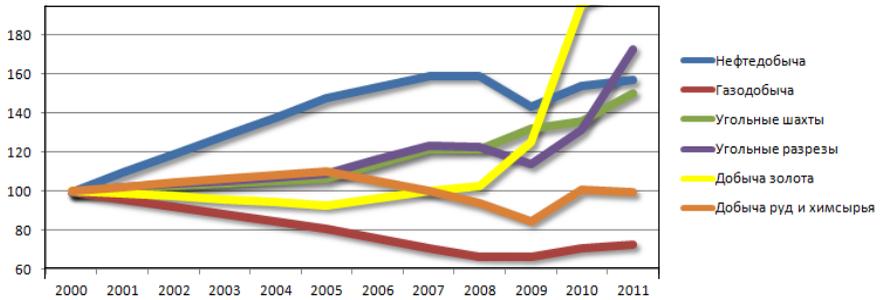


Рисунок 3.1 – Динамика потребления всех видов энергии по типам добычи минеральных ресурсов с 2000 по 2011 гг.

Агрегирование (декомпозиция) – разбиение большого числа факторов (параметров) задачи на небольшое число групп, блоков (агрегатов) по определенному принципу; предполагает, с одной стороны, вполне конкретные связи между блоками, которые нетрудно формализовать и учесть, а с другой стороны, возможность решения необходимых вопросов внутри агрегата.

На рисунке 3.2 приведен пример декомпозиции технологической схемы ремонта, включающий: разборку (RAZB), дефектовку узлов и деталей на рабочих местах (PM), выполнение рабочими (R) технологических операций (MTXO), сборку (SBOR).

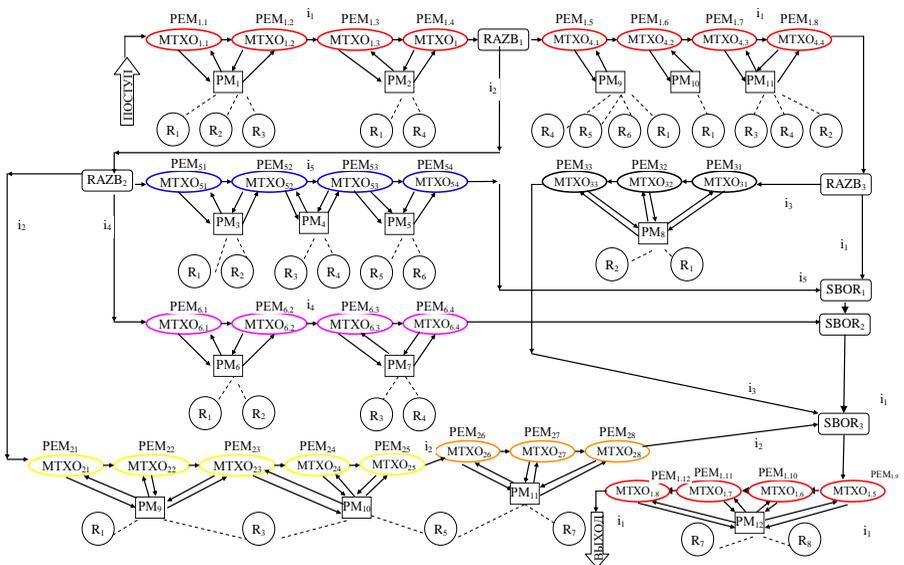


Рисунок 3.2 – Структура рабочих мест ремонтных операций изделия

Теория катастроф – часть математической логики, которая позволяет в области изменения основных параметров (факторов), связанных аналитически, выявить точки, линии, плоскости и т.п. границы (бифуркации), на которых происходят резкие изменения качественного поведения рассматриваемой системы – "катастрофы" той или иной интерпретации поведения системы. Так, предположение, что изменения флоры и фауны на границе мезозоя и кайнозоя (вымирание динозавров, массовое расселение птиц и млекопитающих, вытеснение голосеменных растений покрытосеменными) вызваны катастрофой – падением астероида на Землю. Применение теории катастроф можно найти в исследованиях устойчивости сложных систем, колебаниях и разрушениях в строительной механике и даже возникновении бунтов в тюрьмах.

Метод последовательных приближений – общее название группы математических методов, в которых на каждом очередном цикле однообразных вычислений определяются новые значения параметров, более точные, которые в свою очередь используются на следующем цикле. Например, метод последовательных приближений используется для определения силы тока в катушке электромагнита, необходимый для создания магнитного потока (рисунок 3.3).

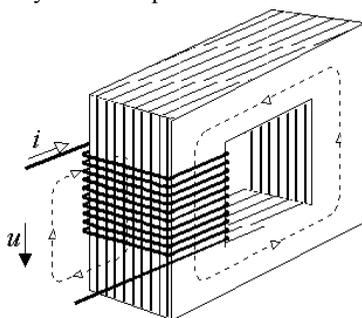


Рисунок 3.3 – Магнитная цепь

Метод проб и ошибок – по результатам одного или нескольких (отличающихся подбираемыми значениями параметров) расчетов делается вывод о направлении дальнейшего подбора искомых значений для минимизации ошибки. Также этот метод называют методом перебора вариантов. Этот прием для непрерывно распределенных факторов, принимающих бесконечное множество значений на любом отрезке своего изменения, не может считаться методом, поскольку не гарантирует получение решения.

Достоинства метода: этому методу не надо учиться; удовлетворительно решаются простые задачи (не более 10 проб и ошибок).

Недостатки метода: плохо решаются задачи средней сложности (более 20–30 проб и ошибок) и практически не решаются сложные задачи (более 1000 проб и ошибок); нет приёмов решения; нет алгоритма мышления; неизвестно, когда будет решение и будет ли вообще; требуются большие затраты времени; иногда этот метод не подходит (не будет же человек, например, резать на бомбе провода наугад).

Метод проверки гипотез – процесс выдвижения, анализа и проверки разнообразных предположений о причинах появления определенного результата. Этот метод имеет смысл применять там, где требуется найти скорее качественное, чем количественное объяснение сложного и неординарного явления.

Приведем пример гипотез потепления, в которых, как правило, учитывается изменение газового состава атмосферы (изменение концентрации парниковых газов), вариации светимости Солнца, вулканические извержения, изменения в орбитальном движении Земли вокруг Солнца (рисунок 3.4).

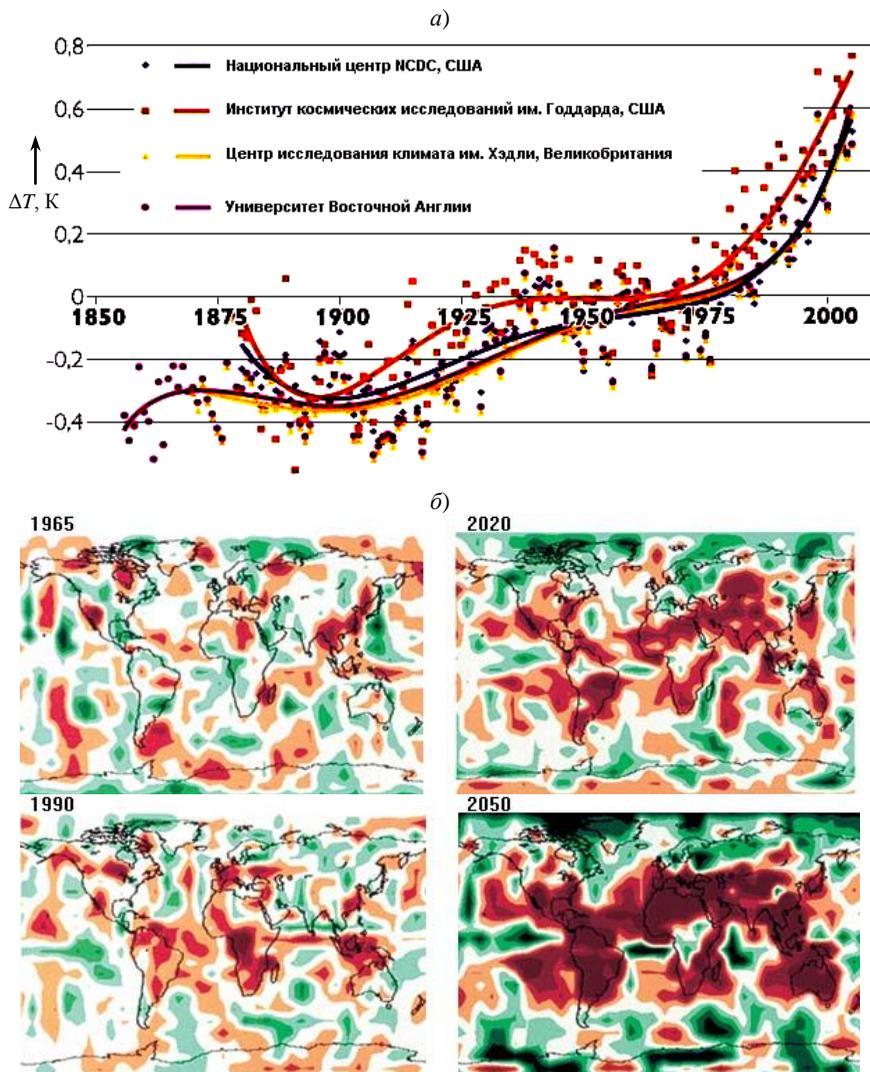


Рисунок 3.4 – Моделирование динамики глобального потепления (а) и пример моделирования роста засушливых областей в Институте космических исследований им. Годдарда (б)

Обзор методов экспертных оценок и многомерный статистический анализ приведены в отдельных разделах данного пособия.

Подобие и анализ размерностей. Наиболее распространенным частным случаем математических моделей является случай подобных моделей. Два объекта подобны, если выполнены одновременно два условия:

- 1) они имеют *одинаковые математические описания*;
- 2) их соответствующие переменные связаны *коэффициентами подобия* (масштабами, константами подобия, коэффициентами пропорциональности).

Еще древние заметили, что величины ведут себя по-разному по отношению к арифметическим действиям. Некоторые из них можно складывать, вычитать, умножать и делить, а результат арифметических действий с другими величинами не имеет смысла. Так, например, не имеет смысла сумма длины и времени, зато результат деления длины на время имеет вполне конкретный физический смысл скорости. Сумма длины и ширины прямоугольника, наоборот, имеет смысл полупериметра. Величины, сумма или разность которых имеет физический смысл, назвали *однородными*.

Величины, численное значение которых зависит от принятых единиц измерения, называются размерными, а величины, численное значение которых не зависит от принятых единиц измерения, – безразмерными. Все известные законы природы описываются с помощью функциональных связей между размерными величинами, поэтому для расчетов необходимо подставлять значения величин вместе с их размерностями.

Порядок и правила применения размерностей устанавливают системы единиц измерения. В 1832 г. Гаусс предложил в качестве основных единиц измерения выбирать независимые единицы (в совокупности не связанные между собой законами природы), на которых строится вся система. В СИ основными единицами приняты:

- метр [м] в качестве меры длины;
- килограмм [кг] в качестве меры массы;
- секунда [с] в качестве меры времени;
- ампер [А] в качестве меры силы электрического тока;
- кельвин [К] в качестве меры термодинамической температуры;
- моль [моль] в качестве меры количества вещества;
- кандела [кд] в качестве меры силы света.

Кроме этого вводятся дополнительные единицы измерения плоских углов – радиан [рад] и телесных углов – стерадиан [ср], по сути, являющиеся безразмерными.

Исходным положением теории размерностей является то, что все основные законы природы в любой системе единиц измерения описываются степенными комплексами:

$$z = x_1^{y_1} x_2^{y_2} \dots x_n^{y_n}.$$

Этот факт нельзя доказать, но он легко проверяется: действительно, законы Ньютона, Кулона, Фарадея и т.п. описываются именно степенными комплексами. Следует оговориться, что функциональные связи, содержащие знак "+" или "-", не являются основными законами природы, а представляют суперпозицию нескольких независимых природных явлений, каждое из которых в свою очередь выражается степенным комплексом.

Недостающий в критерии подобия безразмерный числовой коэффициент можно определить эмпирически в процессе идентификации модели. Таким образом, можно составить недостающие элементы математической модели сложного явления.

Замечание 1. Получить с помощью теории размерностей принципиально новые законы природы невозможно. Действительно, для получения критериев подобия необходимо знать размерности всех основных определяющих параметров. Так, например, заслуга Джорджа Стокса, получившего выражение для силы трения, действующей на сферические объекты в жидкости, состоит именно в том, что он на основании многолетних экспериментальных исследований умозрительно определил физический смысл и размерность коэффициента динамической вязкости μ .

Замечание 2. Безразмерный коэффициент k не зависит от размерных параметров критерия подобия. Поэтому для применения найденного таким образом соотношения в математическом описании необходимо определить коэффициент k эмпирически в процессе сбора информации для идентификации модели.

Теория графов. Идея использования компактных и наглядных схем легла в основу теории графов. Начавшись с задачи о коммивояжере, современная теория графов позволяет решать задачи сложных систем, менеджмента и программирования.

В задаче о коммивояжере ей требуется найти кратчайший замкнутый маршрут (цикл), проходящий через все назначенные пункты по одному разу. Следует отметить, что к категории транспортных задач относятся задачи менеджмента, например, об оптимальном назначении сотрудников для получения наибольшего эффекта или для минимизации возможных допустимых просчетов, а также задачи распределения ресурсов.

Успехи применения теории графов объясняются тем, что большинство задач в ней доведено до строго обоснованных алгоритмов. Однако решить конкретную практическую задачу значительно проще, чем подобрать пригодное готовое решение обобщенной.

Теория массового обслуживания. Она изучает модели систем массового обслуживания (СМО), представляющие собой системы, которые по одному или многим каналам обслуживают поступающие в них заявки. Примерами СМО могут служить АТС, кассы, диспетчер и т.п.

Заявки в систему массового обслуживания поступают не одновременно, а как-то случайно – распределённо во времени – т.е. случайным потоком. Каждая заявка может быть выполнена за свой собственный интервал времени. В некоторых СМО заявка может попадать в очередь и дожидаться, когда освободится какой-либо канал обслуживания. Таким образом, СМО представляют собой модели случайных процессов поступления и обработки заявок. Поток заявок, время их выполнения (обслуживания), условия существования очереди – эти параметры СМО имеют характеристики, описываемые законами распределения.

Теория массового обслуживания различает некоторое число **состояний системы** (например, 1 заявка находится на обслуживании в одноканальной системе, а 4 ожидают в очереди). Каждое из них характеризуется вероятностью нахождения системы в этом состоянии. Кроме того, рассматриваются вероятности перехода системы из одного состояния в другое.

Для наглядности представления состояний СМО применяют графы. Например, телефонный номер может быть в одном из двух состояний: свободен или занят. Граф состояний такой СМО изображен на рисунке 3.5.

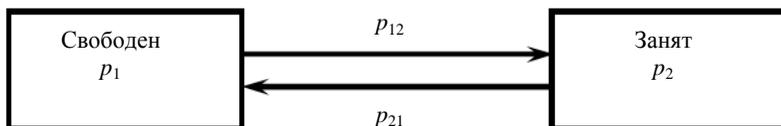


Рисунок 3.5 – Граф состояний СМО

В этом примере p_1 и p_2 задают вероятности того, что номер находится в свободном или занятом состоянии, соответственно. Вероятность перехода телефонного номера из свободного состояния в занятое задается величиной p_{12} , а в обратном направлении – p_{21} . Очевидно, что $\sum_{i=1}^n p_i = 1$, где n – число возможных состояний системы. Это уравнение носит название условия нормировки.

Вероятности переходов определяются *потоком заявок* на обслуживание, поэтому чрезвычайно важно, какие характеристики имеет этот поток.

Потоком событий называется последовательность однородных событий, следующих одно за другим в случайные моменты времени. Поток событий называется стационарным, если он однороден во времени (не зависит от календарного времени). Поток событий называется ординарным, если события в потоке происходят по одиночке, а совместного появления двух и более событий не происходит. В потоке отсутствует последствие, если события в потоке появляются независимо друг от друга (момент появления следующего события не зависит от момента появления предыдущего).

Поток событий называется простейшим (стационарным пуассоновским), если он *стационарен, однороден* и не имеет *последствия*.

Интенсивностью потока $\lambda(t)$ событий называется среднее число событий в единицу времени ($\lambda = \text{const}$ – для стационарного потока).

Для *простейшего* потока вероятность того, что за промежуток времени Δt не появится ни одного события определяется формулой:

$$p(0, \Delta t) = \lambda \Delta t e^{-\lambda \Delta t},$$

где $\lambda \Delta t$ может трактоваться как среднее число событий на интервале Δt .

Для СМО с пуассоновскими потоками заявок и их выполнения применяются математический аппарат марковских случайных процессов.

СМО делятся на два типа: системы с отказами и системы с ожиданием (с очередями). В *системах с отказами* заявка, поступившая в момент, когда все каналы обслуживания заняты, получает отказ и пропадает. В *системах с ожиданием* в таком случае заявка становится в очередь и ждет, когда освободится какой-нибудь канал, и сразу поступает в него на обслуживание.

На примере рисунка 3.5 рассмотрим вероятность появления за Δt одного телефонного вызова: $p(1, \Delta t) = \mu \Delta t e^{-\mu \Delta t}$. Здесь μ – интенсивность потока "завершения" разговоров по телефону". Для СМО типа телефона в установленном режиме работы можно определить следующие характеристики:

- вероятность соединения с абонентом $p_1 = \frac{\mu}{\lambda + \mu}$;
- вероятность получить отказ ("занято") $p_2 = \frac{\lambda}{\lambda + \mu}$;
- абсолютная пропускная способность – среднее число обслуженных заявок (разговоров) за единицу времени: $A = \lambda p_0 \frac{\lambda \mu}{\lambda + \mu}$.

Классическим примером СМО с отказами является так называемый процесс гибели и размножения, характеризующийся последовательной цепочкой состояний и возможностью перехода только в *соседние* состояния (рисунок 3.6).

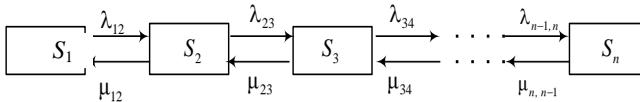


Рисунок 3.6 – Граф состояний СМО с переходом только в соседние состояния

Здесь S_i – состояния системы, $\lambda_{i,i+1}$ – интенсивности переходов из низшего состояния в очередное высшее, $\mu_{i+1,i}$ – интенсивности обратных переходов из высшего состояния в предыдущее низшее.

Предельные вероятности состояний (при $t \rightarrow \infty$ в установившемся случае) определяются следующими формулами:

$$P_1 = \frac{1}{1 + \frac{\lambda_{12}}{\mu_{21}} + \frac{\lambda_{12}\lambda_{23}}{\mu_{21}\mu_{32}} + \dots + \frac{\lambda_{12}\lambda_{23}\dots\lambda_{n-1,n}}{\mu_{21}\mu_{32}\dots\mu_{n,n-1}}}; \quad P_2 = \frac{\lambda_{12}}{\mu_{21}} P_1; \dots; \quad P_n = \frac{\lambda_{12}\lambda_{23}\dots\lambda_{n-1,n}}{\mu_{21}\mu_{32}\dots\mu_{n,n-1}} P_1.$$

В классической теории массового обслуживания вероятностные характеристики (λ , μ) и законы распределения состояний и переходов между ними полагаются пуассоновскими. Если это не так, то их можно определить статистически – наблюдая работу оригинальной системы. В этом случае приходится пользоваться более сложным математическим аппаратом, описывающим процессы перехода системы из состояния в состояние.

Приведем некоторые примеры теории массового обслуживания.

1 В компьютер, управляющий технологическим процессом, поступают сигналы от датчиков, связанных с управляемым объектом. Каждый сигнал требует обработки в течение некоторого случайного времени (зависящего от содержания сигнала). Требуется решить задачу: способна ли машина с данным объемом памяти и быстродействием справиться с обработкой всех поступающих сигналов.

2 Автоматическая телефонная станция (АТС) обслуживает некоторое число абонентов, выходящих на связь по случайному закону. Абонент, пытающийся позвонить (послать вызов) в момент занятости всех линий, получает отказ ("частые гудки").

3 В сборочный цех поступают для сборки детали различных видов. При нехватке хотя бы одного вида деталей производство останавливается; избыточные детали поступают в бункеры определенной вместимости. На процесс поступления деталей, как и на время сборки изделия, влияют случайные факторы. Требуется ответить на вопросы: какова вероятность простоя производственной линии? Чему равна вероятность переполнения бункеров? Элементарной операцией в данном случае является сборка одного изделия из готового комплекта деталей.

4 В морской порт прибывают суда не строго по графику, а со случайными отклонениями. Имеется несколько погрузочно-разгрузочных площадок с соответствующим оборудованием. Спрашивается: чему равно среднее время от момента прибытия судна до окончания его разгрузки и погрузки? Элементарной операцией здесь можно считать процесс разгрузки и погрузки одного судна.

Метод Монте-Карло. В тех случаях, когда решение уравнений перехода в СМО затруднено, используется метод статистических испытаний. Этот универсальный метод *стохастического (имитационного)* моделирования позволяет не только определять параметры системы, но и имитировать ее

работу. *Метод статистических испытаний* (метод Монте-Карло) сводится к *розыгрышу* случайных событий в СМО. Элементарным примером такого розыгрыша может служить выбор одного из двух исходов с помощью подбрасывания монетки.

Метод Монте-Карло включает в себя три этапа: получение случайного числа R , отождествление его с вероятностью и розыгрыш единичного жребия.

Случайное число R – значение случайной величины, равномерно распределенной на интервале $[0, 1]$. Такое случайное число можно получить с помощью рулетки, размеченной, например, простыми десятичными дробями – отсюда и название метода Монте-Карло. Сейчас существуют специальные программы – "датчики случайных чисел", которые позволяют при каждом обращении к программе получить "псевдослучайное число" (случайную величину, распределенную *почти* равномерно на интервале $[0, 1]$ и принимающую конечное множество значений).

Случайное число ставят в *соответствие* вероятности рассматриваемого события, так как и случайное число, и вероятность принимают значения на интервале $[0, 1]$.

В теории вероятностей условились называть *единичным жребием* любой опыт со случайным исходом, который отвечает на один из следующих вопросов:

- "произошло" или "не произошло" определенное событие A ;
- какое событие из полной группы несовместных событий $\{A, B, \dots, C\}$ "произошло";
- какое значение "приобрела" случайная величина.

Для ответа на *первый вопрос* единичного жребия необходимо знать вероятность события A : p . Тогда, если разыгранное случайное число $R < p$, то считают, что событие A "произошло", если $R > p$, то "не произошло" (рисунок 3.7).

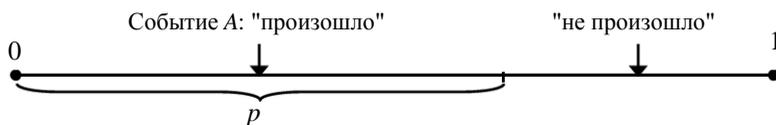


Рисунок 3.7 – Геометрическая интерпретация розыгрыша события A

Полная группа событий – это такая группа событий, кроме которых никаких других событий произойти не может. *Несовместные события* не происходят одновременно. Таким образом, полная группа несовместных событий $\{A, B, \dots, C\}$ имеет сумму вероятностей, равную единице. Иначе говоря, на интервале $[0, 1]$ можно выделить последовательность непересекающихся подинтервалов длиной, равной вероятностям этих событий $p(A)$,

$p(B)$, ..., $p(C)$. Тогда ответ на *второй вопрос* единичного жребия о том, какое из событий "произошло", делают по тому факту, на какой из подинтервалов попало случайное число R (рисунок 3.8).

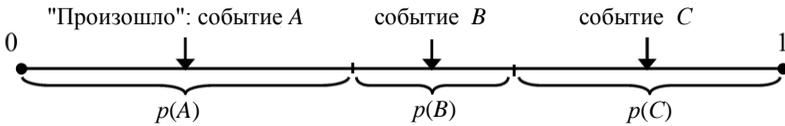


Рисунок 3.8 – Геометрическая интерпретация розыгрыша событий A, B, C

В третьем случае, если случайная величина дискретна, то процедура сводится к предыдущей. Непрерывная случайная величина, как известно, задается законом распределения в виде интегральной функции распределения $F(x) = P(\xi < x)$, т.е. вероятности того, что случайная величина ξ , примет значение, не превосходящее x . Эта функция по своему смысловому содержанию является монотонно возрастающей от 0 до 1 при изменении ξ от нижней до верхней границы области возможных значений. В силу монотонности эту функцию можно обратить однозначно: по заданному значению $F(x)$ определить единственное значение x , ему соответствующее. Тогда случайное число R рассматривают как значение вероятности $F(x)$ и по нему находят значение x (рисунок 3.9).

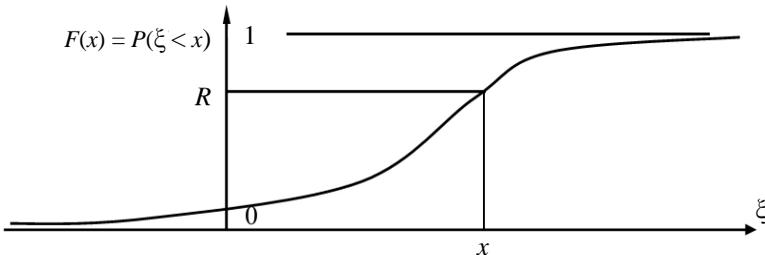


Рисунок 3.9 – Применение интегральной функции распределения вероятности того, что случайная величина ξ , примет значение, не превосходящее x

Имея такой "аппарат" розыгрыша единичного жребия, нетрудно построить цепочку жребиев для реализации случайного явления любой сложности, в том числе и для имитации процесса функционирования СМО.

Пример применения метода Монте-Карло с помощью вычисления площади фигуры на плоскости.

Пусть дана некоторая плоская фигура F (рисунок 3.10), для которой требуется найти площадь. Введем следующие предположения: предположим, что эта фигура целиком расположена внутри единичного квадрата; периметр фигуры может быть устроен совершенно произвольно; фигура может не быть связной, т.е. может состоять из нескольких областей; фигура может быть задана аналитически или графически.

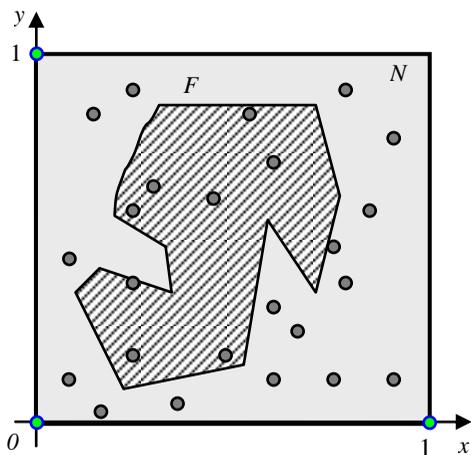


Рисунок 3.10 – Вычисление площади фигуры с помощью метода Монте-Карло

Сгенерируем в квадрате N случайных точек. Пусть N^* – количество точек, попавших внутрь рассматриваемой фигуры. Тогда при достаточно больших значениях N площадь фигуры F может быть оценена как

$$\hat{S} = \frac{N^*}{N}.$$

Конечно, в задаче вычисления площади существуют и более точные алгоритмы нахождения площади, но данный пример демонстрирует простейший случай применения метода Монте-Карло.

Особенностью имитационных моделей с законами распределения случайных величин, получаемых из наблюдения за оригиналом, является то, что вопрос о проверке адекватности отпадает. Имеет смысл проверка адекватности только в том случае, когда используются априорные законы распределения, не полученные статистической обработкой наблюдений.

Применение метода Монте-Карло, например, в экономической сфере связано с реализацией численных экспериментов для оценки степени влияния на результат ряда факторов в виде цены, объема выпуска продукции или переменных расходов. При этом должны быть жестко закреплены взаимосвязи между выходными и исходными показателями в форме математического неравенства или уравнения. В виде результата выступает любой из критериев эффективности (рисунок 3.11).

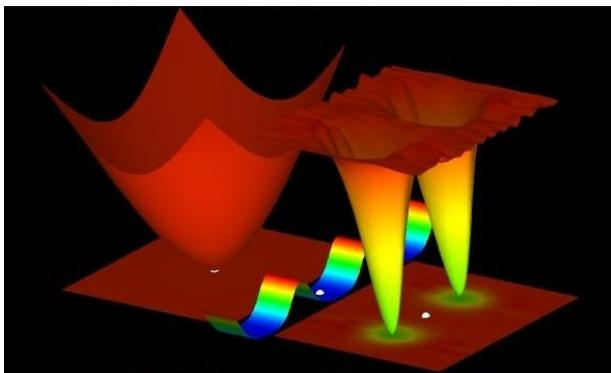


Рисунок 3.11 – Пример трехмерной экономической модели анализа проектных рисков методом Монте-Карло

4 ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ И ПРИЕМЫ

4.1 Вычислительные методы алгебры

В данном разделе представлены наиболее распространенные вычислительные методы, используемые для численного решения отдельных задач, встречающихся при математическом моделировании [1].

а) **Методы решения систем линейных алгебраических уравнений** достаточно подробно изучаются в курсе высшей математики. Здесь следует упомянуть о необходимости анализа условий *применимости* каждого метода к решению конкретной задачи. Так, например, если в процессе вычисления коэффициентов матрицы системы нельзя гарантировать априори существенно отличное от нуля значение главного определителя системы, то применение правила Крамера или матричного метода невозможно. Наиболее универсальными (однако тоже с оговорками) являются *методы исключения неизвестных*: различные варианты схем Гаусса, Жордана; а также *итерационные методы*: простой итерации, метод Зейделя и т. п.

б) **Методы решения нелинейных алгебраических уравнений** вида $f(x) = 0$ (или систем нелинейных алгебраических уравнений) обычно строятся на основе итераций – многократных *последовательных приближений*. Общая идея итерационных методов заключается в преобразовании исходной задачи отыскания корня функции $f(x)$ к итерационному виду: $x = \varphi(x)$. Далее строится *итерационный процесс* ("пошаговое уточнение" искомого значения x) по формуле $x^{[i+1]} = \varphi(x^{[i]})$, где $[i]$ обозначает номер шага итераций. Такого рода формулы, позволяющие вычислять каждое следующее приближение, исходя из предыдущего, называются рекуррентными формулами. Возможны различные способы приведения к итерационному уравнению, но для всех итерационных методов формулируются *условия сходимости* и оценка *погрешности*. Метод можно применять, только убедившись в выполнении для исходной функции условий сходимости – условий того, что итерационный процесс последовательного приближения сходится именно к решению этого уравнения. Последнее бывает не всегда: неудачно построенный процесс последовательных приближений (не считающийся в математике методом) может сходиться к решению совсем не исходной задачи, а другой, может вообще расходиться или не сходиться ни к какому решению.

Наиболее распространенными итерационными методами решения одного нелинейного уравнения $f(x) = 0$ являются *методы деления отрезка пополам, секущих, золотого сечения, касательных (Ньютона)*.

Все эти методы применяются только в той области изменения аргумента x , где безусловно *существует единственный корень искомого уравнения*. Если корня на этом отрезке нет, то и искать его там бессмысленно – решения нет. Если на отрезке несколько корней, то необходимо разбить его на такие части, которые содержат только по одному корню. Поэтому необхо-

дим заранее убедиться в выполнении этого требования, проанализировав функцию $f(x)$ на предполагаемом исходном отрезке.

Выбор начального интервала, на котором безусловно *существует единственный корень искомого уравнения*, называется отделением корней. Указанные условия можно выполнить, опираясь на теорему о монотонной на отрезке функции: всякая монотонная на отрезке функция принимает любое свое промежуточное значение в одной единственной точке внутри отрезка. В этом случае необходимо лишь показать выполнение *одновременно двух свойств*:

- *монотонности* на этом интервале функции $f(x)$, что проверяется по условию $f'(x) \neq 0$ или из физических соображений;

- на концах этого отрезка $x^{[0]}$ и $x^{[1]}$ функция принимает значения *разных знаков* (т.е. на одном конце $f(x^{[0]}) > 0$, а на другом – $f(x^{[1]}) < 0$).

Этим будет гарантировано существование одной единственной точки внутри отрезка, в которой $f(x) = 0$.

В итоге процедуры отделения корней получается, что положение корня уравнения известно с точностью до длины выбранного отрезка. Остается построить итерационный процесс таким образом, чтобы на каждой итерации *уменьшать отрезок*, на котором находится корень.

1) Метод деления отрезка пополам для решения нелинейного алгебраического уравнения применяется на отрезке, для которого проведена процедура отделения корней, и использует итерационное уравнение в виде

$$x^{[i+1]} = 0,5 (x^{[i]} + x^{[i-1]}).$$

Идея этого метода заключается в простейшей процедуре разбиения отрезка на две равные части и исследования, на какой из них находится искомый корень уравнения. Такие деления и исследования *повторяются на каждой итерации* (рисунок 4.1).

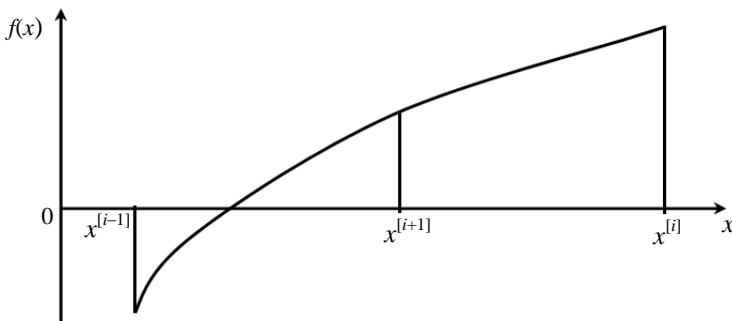


Рисунок 4.1 – Геометрическая интерпретация метода деления отрезка пополам

Поскольку функция монотонна на всем отрезке, то она монотонна и на его части, поэтому на каждом шаге итерации достаточно выбрать тот (вдвое меньший) отрезок, на концах которого выполняются условия $f(x^{[i+1]}) > 0$ и

$f(x^{[i]}) < 0$. Так как после каждой итерации новый отрезок всегда меньше старого, то область возможного расположения корня постепенно сужается – стягивается в точку, а именно к решению исходного уравнения.

Итерации завершают, когда будет выполнено условие *заданной точности*. Это условие, в зависимости от постановки задачи исследований может быть сформулировано одним из двух способов: *по аргументу* $|x^{[i+1]} - x^{[i]}| < \varepsilon$ (размер интервала стал меньше требуемой погрешности определения корня) или *по функции* $f'(x^{[i+1]}) < \delta$ (значение функции пренебрежимо мало отличается от нуля).

Этот экономный метод, как видно из формулы, не использует значения функции для определения очередного приближения; и даже при выборе части интервала для следующего шага использует не столько значения функции, сколько лишь ее *знаки*. Алгоритм этого метода предельно прост.

2) Метод секущих (метод хорд) для решения нелинейного алгебраического уравнения применяется, проводится и завершается аналогично методу деления отрезка пополам (рисунок 4.2).

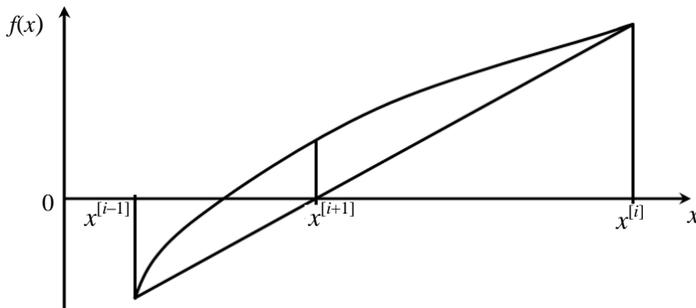


Рисунок 4.2 – Геометрическая интерпретация метода секущих

Единственное его отличие заключается в итерационной формуле для отыскания очередного приближения, которая основана на пропорции для подобных треугольников:

$$x^{[i+1]} = x^{[i]} - \frac{x^{[i]} - x^{[i-1]}}{f(x^{[i]}) - f(x^{[i-1]})} f(x^{[i]}).$$

Этот метод, как видно из формулы, использует для определения очередного приближения больше информации о значениях функции, поэтому от него следует априорно ожидать более быстрой сходимости к решению.

3) Метод золотого сечения для решения нелинейного алгебраического уравнения применяется, начинается и завершается так же, как и пре-

дыдущие методы. Его отличие от них заключается в применении не одной, а двух точек внутри отрезка, используемых для следующего шага итерации.

Золотым сечением отрезка $[a, b]$ называются две точки:

$$u_1 = a + \frac{3 - \sqrt{5}}{2}(b - a) \approx a + 0,3811966911(b - a),$$

$$u_2 = a + \frac{\sqrt{5} - 1}{2}(b - a) \approx a + 0,618033989(b - a),$$

расположенные симметрично относительно середины отрезка. Каждая из этих точек делит исходный отрезок на две неравные части таким образом, что *отношение длины всего отрезка к длине большей части равняется отношению длины большей части к длине меньшей части*:

$$\frac{b - a}{b - u_1} = \frac{b - u_1}{u_1 - a}, \quad \frac{b - a}{u_2 - a} = \frac{u_2 - a}{b - u_2}.$$

На каждом очередном шаге итераций при известных $x^{[i]}$ и $x^{[i+1]}$ определяются точки u_1 и u_2 золотого сечения отрезка между точками $x^{[i]}$ и $x^{[i+1]}$ и знаки функции в точках золотого сечения: $f(u_1)$ и $f(u_2)$ (рисунок 4.3).

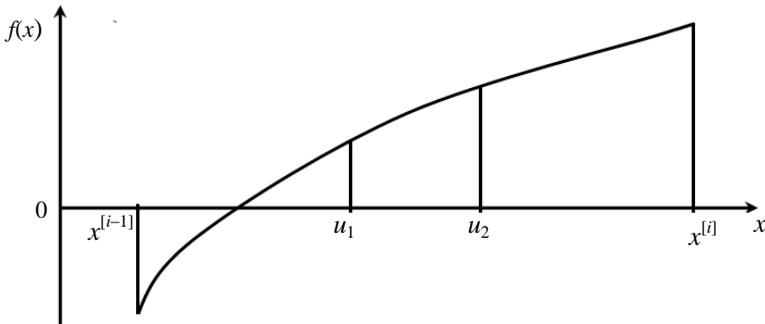


Рисунок 4.3 – Геометрическая интерпретация метода золотого сечения

Для перехода к следующему шагу итерации выбираются те *две ближайшие друг к другу точки* из четырех: $x^{[i]}$, $x^{[i+1]}$, u_1 , u_2 , в которых значения функции различаются знаком. Эти точки образуют новый отрезок, на котором находится решение, и следует проводить очередное золотое сечение.

Заметим, что метод золотого сечения, как и метод деления отрезка пополам, не использует значений функции: на каждой итерации для выбора отрезка нужны только знаки функции. Поэтому этот метод следует считать, вообще говоря, более экономным, чем метод секущих. С другой стороны, он быстрее сходится, чем метод деления отрезка пополам, так как на каждой итерации отрезок уменьшается почти втрое.

Предыдущие сравнения скорости сходимости методов весьма условны, так как этот процесс существенно зависит не только от вида функции, но и от выбора исходного приближения.

Метод секущих, метод деления отрезка пополам и метод золотого сечения, а также их модификации удобны тогда, когда функция $f(x)$ вычисляется относительно просто, а искомый корень – единственный на известном отрезке. В случае, когда еще и производная $f'(x)$ вычисляется достаточно просто, можно использовать более быстро сходящиеся методы, основанные на информации о производной, например, метод касательных (Ньютона).

4) Для применения *метода касательных (метода Ньютона)* требуется соблюдение не только прежних условий единственности решения на исходном отрезке, но и дополнительного условия сохранения своего знака второй производной $f''(x) \neq 0$ (функция не только строго *монотонна* и имеет на концах отрезка значения *разных знаков*, но и выпукла, т.е. метод нельзя применять на интервале, где возможны несколько корней или точки перегиба).

Перед началом метода проводится процедура отделения корня с помощью проверки указанных свойств.

В качестве итерационной формулы используется выражение

$$x^{[i+1]} = x^{[i]} - \frac{f(x^{[i]})}{f'(x^{[i]})},$$

которое следует применять к тому концу отрезка, на котором знаки $f(x)$ и $f''(x)$ *совпадают* (рисунок 4.4). Если применить ее неверно, то можно получить следующее "приближение" вне исходного отрезка и метод начнет расходиться.

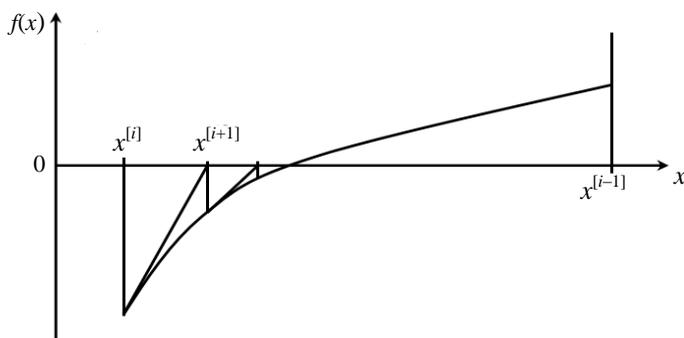


Рисунок 4.4 – Геометрическая интерпретация метода золотого сечения

Завершается метод касательных (метод Ньютона) так же, как и предыдущие методы.

Метод касательных (метод Ньютона), в отличие от остальных рассмотренных, может применяться и в многомерном случае, т.е. для решения невырожденных систем нелинейных алгебраических уравнений с числом уравнений, равным числу неизвестных. В этом случае формула метода трактуется в матричном представлении.

в) **Методы интерполяции таблично заданных функций** применяются для вычисления значений функции в точках между соседними узлами x_k и x_{k+1} , в которых значения функции $f(x_k)$ и $f(x_{k+1})$ заданы. Таким образом, интерполяция служит для *доопределения* функции в промежутках между заданными ее значениями в узлах.

Необходимость в такой процедуре возникает при использовании в математических моделях таких характеристик объекта, которые получены экспериментальным или сложным расчетным способом. При этом в зависимости от поставленной задачи могут предъявляться различные специфические требования к свойствам такой интерполяции. Различают следующие методы интерполяции.

1) Кусочно-постоянная интерполяция используется нами повседневно, когда мы говорим, сколько сейчас времени: в течение, например, минуты время считается постоянным (12 часов 27 минут). Графическое представление такой интерполяции приведено на рисунок 4.5.

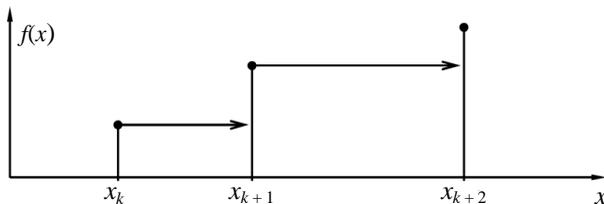


Рисунок 4.5 – Геометрическая интерпретация кусочно-постоянной интерполяции

Кусочно-постоянная интерполяция самая простая, но и обладает самыми примитивными качествами с точки зрения применения в моделировании. Действительно, в каждом узле полученная *интерполяционная функция* терпит разрыв, а разрывная функция применима далеко не во всех задачах.

2) Самая распространенная в расчетах – линейная интерполяция – для нахождения значения функции в точке x , расположенной между соседними узлами (рисунок 4.6), предполагает **линейный** характер изменения функции:

$$f(x) = f(x_k) + \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} (x - x_k).$$

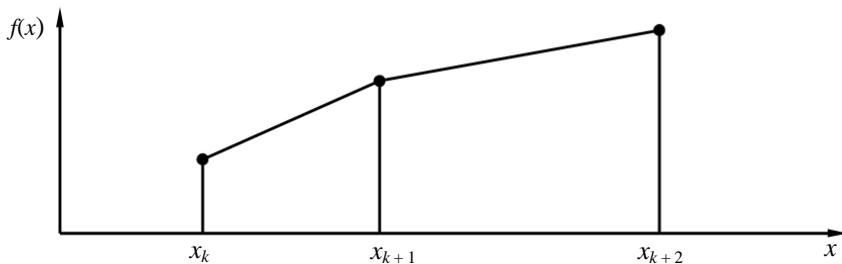


Рисунок 4.6 – Геометрическая интерпретация линейной интерполяции

Линейная интерполяционная функция непрерывна, однако имеет *разрывы производной* в узлах (представляет собой ломанную, связывающую отрезками прямых все заданные узлы). Поэтому, например, в задачах оптимизации она неприемлема. Однако для простых расчетных процедур она самая употребительная, ее изучают в средней школе при работе с тригонометрическими функциями по таблицам В. М. Брадиса.

3) Квадратичная интерполяция развивает идею линейной для поиска "удобной" функции. Если одна-единственная точка задает лишь одно свое значение (постоянная), а две точки – отрезок прямой (линейная функция), то квадратичная функция, как известно, проходит через три заданные точки. Поэтому для построения квадратичной интерполяционной функции используются три соседних узла таблично заданной функции (рисунок 4.7).

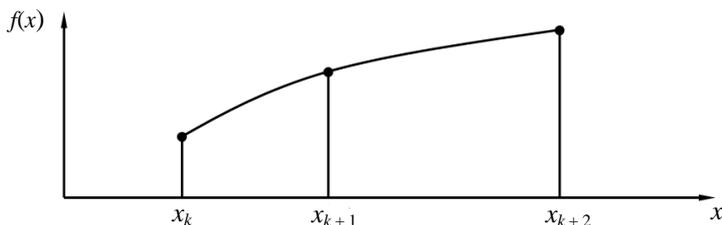


Рисунок 4.7 – Геометрическая интерпретация квадратичной интерполяции

Квадратичная интерполяционная формула получается именно из системы уравнений, описывающих прохождение многочлена второй степени через три точки, заданные соседними узлами:

$$f(x) = f(x_k) + \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} (x - x_k) + \frac{f(x_{k+2}) - f(x_{k+1})}{x_{k+2} - x_{k+1}} - \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} (x - x_{k+1})$$

При малых изменениях x – между x_k и x_{k+2} – производная такой интерполяционной функции остается непрерывной даже в среднем узле x_{k+1} . Однако при смене "троек" узлов разрыва производной избежать не удастся.

4) Полиномиальная интерполяция развивает идею использования многочленов (полиномов) до *необходимого числа узлов*. В общем случае через $n + 1$ точку проходит единственный многочлен степени n , так как для определения всех его коэффициентов (от свободного члена a_0 до старшего a_n) необходимо $n + 1$ уравнение. Однако обычно вместо процедуры вычисления коэффициентов интерполяционного многочлена используются готовые интерполяционные формулы, подобные приведенным выше, с помощью которых непосредственно вычисляется значение интерполяционной функции в любой точке между крайними узлами x_0 и x_n . Наиболее известной из них является *интерполяционная формула Лагранжа*:

$$f(x) = \frac{(x-x_1)(x-x_2)\cdots(x-x_n)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)} f(x_0) + \frac{(x-x_0)(x-x_2)\cdots(x-x_n)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)} f(x_1) + \dots + \frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_1)(x_n-x_2)\cdots(x_n-x_{n-1})} f(x_n) = \sum_{i=0}^n \frac{\prod_{\substack{k=0 \\ k \neq i}}^n (x-x_k)}{\prod_{\substack{k=0 \\ k \neq i}}^n (x_i-x_k)} f(x_i).$$

Нетрудно видеть, что полиномиальная интерполяция, хотя и достаточно громоздка, но обеспечивает сколь угодно гладкую функцию – непрерывную вместе со всеми производными.

5) Сплайновая интерполяция – интерполяция с помощью таких многочленов (сплайнов) на каждом участке интерполяции между соседними узлами, которые не только совпадают в определенном числе узлов со значениями заданной функции, но и дают необходимое число непрерывных производных при переходе от одного участка интерполяции к соседнему. Для этого при определении очередного сплайна используют значения не только заданной в узлах функции, но и производных предыдущего (например, левого) сплайна в точках сопряжения. Сплайновая интерполяция позволяет достаточно экономным образом получить интерполяционную функцию с заданными свойствами гладкости, что бывает необходимо, например, в задачах оптимизации.

г) **Методы аппроксимации функций** – методы приближенной замены заданной сложной функциональной зависимости более простой функцией (алгебраическим полиномом, тригонометрическим полиномом и другими функциями, которую можно построить с помощью метода наименьших квадратов).

Следует четко различать задачи интерполяции и аппроксимации. Если интерполяционная функция обязательно совпадает в узлах с заданной, то аппроксимирующая – не обязательно! Последняя чаще всего не проходит вообще ни через одну заданную узловую точку. Аппроксимация нужна для простого вычисления сложных функций или для сглаживания (построения гладкой заменяющей функции) таблично заданных функций, чаще всего экспериментальных.

Простейшей аппроксимационной формулой является известная формула Тейлора, приближенно отражающая поведение известной функции в окрестности единственной точки с учетом необходимого числа производных.

Если зависимость имеет явно выраженный характер ограниченной на отрезке или периодической функции, то может быть использована *аппроксимация тригонометрическими функциями* (конечной частью ряда Фурье).

4.2. Вычислительные методы решения дифференциальных уравнений

Численные *методы интегрирования* обыкновенных дифференциальных уравнений и их систем вида $y' = f(x, y)$ можно построить только для случая известных начальных значений всех интегрируемых переменных (для задачи Коши). Общее решение дифференциальных уравнений содержит произвольные постоянные, которые недопустимы в математических моделях, поэтому в качестве искомой функции используется определенное начальными условиями частное решение.

Эти вычислительные методы основаны на замене дифференциальных уравнений алгебраическими. Операцию взятия производной невозможно представить в компьютерных программах, поэтому производная заменяется *разностным* выражением того или иного вида. В зависимости от этого вида различаются *разностные схемы* численного представления дифференциальных уравнений и соответствующие им методы.

1 Методы Эйлера. Простейший метод Эйлера основан на аппроксимации производной простейшей разностной схемой вида

$$y' = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \approx \frac{y_{k+1} - y_k}{x_{k+1} - x_k}.$$

Отсюда в силу решаемого дифференциального уравнения $y' = f(x, y)$ выводится разностное уравнение метода:

$$y_{k+1} = y_k + f(x_k, y_k) \Delta x.$$

Рисунок 4.8 дает геометрическое представление об этом методе: в силу вида исходного дифференциального уравнения функция $f(x_k, y_k)$ представляет собой значение производной в левом конце интервала Δx , называемого шагом интегрирования. Тогда разностное уравнение метода Эйлера просто описывает выходящую из левого конца шага интегрирования касательную к неизвестной искомой интегральной кривой (изображенной пунктиром).

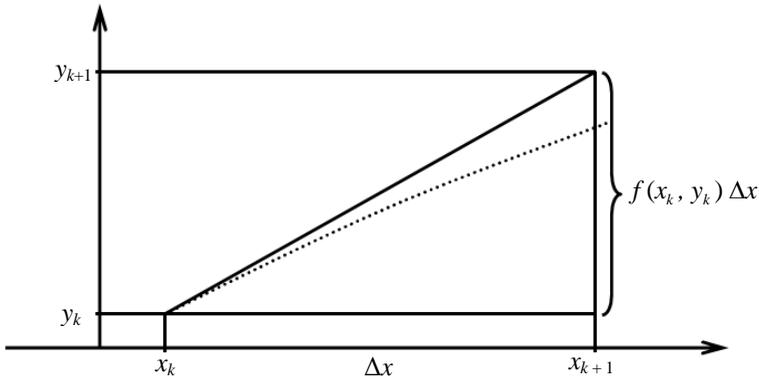


Рисунок 4.8 – Геометрическая интерпретация метода Эйлера

Очевидно, что небольшую неизбежную погрешность при такой аппроксимации можно обеспечить только малым шагом интегрирования Δx . Поэтому численное решение задачи Коши на достаточно большом промежутке изменения аргумента – очень кропотливая процедура, немислимая без вычислительной техники.

Простейший метод Эйлера относится к методам I порядка, поскольку использует в разностной формуле значение функции в *одной* точке.

Заметим попутно, что приведенная выше разностная схема аппроксимации производной не единственно возможная. Например, схема I порядка $y' \approx \frac{y_k - y_{k-1}}{x_k - x_{k-1}}$ и схема II порядка $y' \approx \frac{y_{k+1} - y_{k-1}}{x_{k+1} - x_{k-1}}$ имеют такое же право на аппроксимацию производной, однако интегрирование с их помощью обладает рядом особенностей.

Простейший метод Эйлера на практике почти не используется. Наибольшее распространение получили модифицированные методы Эйлера II порядка. Идея первой модификации заключается в выполнении шага интегрирования за *два полушага* и дает уравнение:

$$y_{k+1} = y_k + f\left(x_k + \frac{\Delta x}{2}, y_k + f(x_k, y_k) \frac{\Delta x}{2}\right) \Delta x.$$

Идея второй модификации заключается в выполнении *предварительного шага* интегрирования и поправки на касательную в конце шага:

$$y_{k+1} = y_k + 0,5 \Delta x [f(x_k, y_k) + f(x_k + \Delta x, y_k + f(x_k, y_k) \Delta x)].$$

Общими недостатками методов Эйлера I порядка являются *невысокая точность* и *слабая устойчивость* (погрешность одного шага интегрирования не только не компенсируется на последующих шагах, а растет). Например, для интегрирования уравнений динамики полета транспортных само-

летов в условиях, близких к установившимся, с практической точки зрения допустимо пользоваться простейшим методом Эйлера. Но при исследовании неустановившихся режимов полета этого недостаточно – следует применять модифицированные методы Эйлера, а для моделирования движения самонаводящихся ракет использование методов Эйлера практически недопустимо.

2 Методы Адамса используют значения функции в нескольких предыдущих точках (учитывают предысторию поведения функции: y_{k-1}, \dots) для исправления направления касательной. Формула метода Адамса I порядка совпадает с формулой простейшего метода Эйлера:

$$y_{k+1} = y_k + f(x_k, y_k) \Delta x,$$

а формулы более высокого порядка строятся наращиванием формул меньшего порядка:

$$y_{k+1} = y_k + f(x_k, y_k) \Delta x + [f(x_k, y_k) - f(x_{k-1}, y_{k-1})] 0,5 \Delta x \quad \text{– II порядка};$$

$$y_{k+1} = y_k + f(x_k, y_k) \Delta x + [f(x_k, y_k) - f(x_{k-1}, y_{k-1})] 0,5 \Delta x + \\ + [f(x_k, y_k) - 2f(x_{k-1}, y_{k-1}) + f(x_{k-2}, y_{k-2})] \frac{5}{12} \Delta x \quad \text{– III порядка}.$$

Методы Адамса более *устойчивы*, чем методы Эйлера, а точность их растет с увеличением порядка. Трудоемкость расчетов по сравнению с другими методами такого же порядка (см. ниже) значительно меньше, так как используются значения функции, вычисленные ранее на предыдущих шагах интегрирования. Существенным неудобством методов Адамса является необходимость на первых шагах интегрирования использовать другие методы, поскольку значения функции в "предыдущих" точках не определены.

3 Методы "прогноз-коррекция" осуществляют расчет в два шага: предварительный расчет y_{k+1}^{II} – "прогноз" ("предсказание") и последующее уточнение – "коррекцию" y_{k+1}^{K} . Для построения формул метода "прогноз-коррекция" определенного порядка используются формулы *метода Адамса* того же порядка; например, для простейшего метода I порядка:

$$y_{k+1}^{\text{II}} = y_k + f(x_k, y_k) \Delta x \quad \text{– "предсказание"},$$

$$y_{k+1}^{\text{K}} = y_{k+1}^{\text{II}} = y_k + f(x_k, y_{k+1}^{\text{II}}) \Delta x \quad \text{– "коррекция"}.$$

Геометрическая интерпретация этого метода I порядка показана на рисунке 4.9.

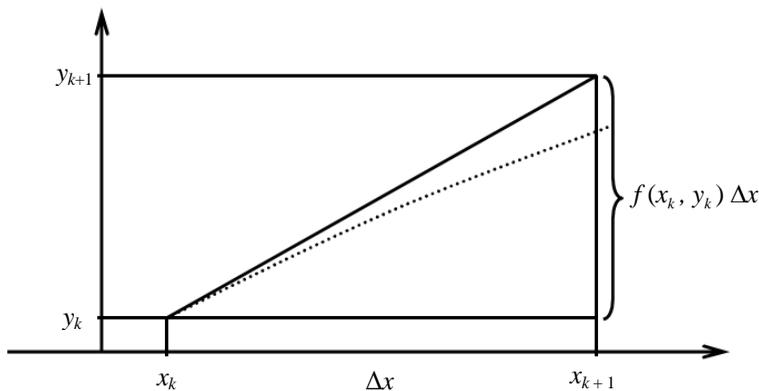


Рисунок 4.9 – Геометрическая интерпретация метода "прогноз-коррекция"

При предсказании, как и при методе Эйлера (рисунок 4.8), решение разностного уравнения на шаге отклоняется от точного решения в сторону *выпуклости* функции, так как строится с помощью касательной *в начале шага*. В свою очередь коррекция приводит к отклонению в сторону *вогнутости*, так как строится с помощью прямой, проведенной из той же исходной точки шага, но с наклоном, соответствующим наклону касательной *в конце шага*. Таким образом, разность между $y_{k+1}^{\text{п}}$ и $y_{k+1}^{\text{к}}$ может служить мерой погрешности численного интегрирования на одном шаге. То есть методы "прогноз-коррекция" выгодно отличаются от ранее описанных методов тем, что допускают *контроль* величины погрешности на каждом шаге интегрирования. Это можно использовать для повышения точности расчетов с помощью уменьшения шага или для экономии времени расчетов с помощью увеличения шага Δx .

Для всех разностных методов справедливо утверждение: чем меньше Δx , тем меньше погрешность на шаге, тем выше точность интегрирования дифференциальных уравнений. Однако нельзя заранее сказать, какова должна быть величина Δx для обеспечения заданной точности. Поэтому расчеты с неприемлемой погрешностью просто идут "в корзину". В отношении этого выгодно отличаются разностные методы, которые позволяют не только контролировать погрешность, но и *изменять шаг в процессе* интегрирования. Этим последним удобством обладают все разностные методы I порядка, но из них только метод "прогноз-коррекция" дает возможность проконтролировать погрешность и подсказать, когда возникает необходимость изменения шага. Из методов более высокого порядка предоставляют возможность изменения шага интегрирования методы Эйлера и Рунге – Кутта.

4 Методы Рунге – Кутта m -го порядка используют m внутренних точек шага интегрирования, которые задаются характерным для определенной

модификации этого метода способом и в которых *последовательно* вычисляются m значений функции:

$$k_1 = f(x_k^{(1)}, y_k); \quad k_2 = f\left(x_k^{(2)}, y_k + k_1(x_k^{(1)} - x_k^{(2)})\right); \dots$$

$$k_m = f\left(x_k^{(m)}, y_k + k_{m-1}(x_k^{(m)} - x_k^{(1)})\right),$$

а затем производится непосредственно сам шаг интегрирования:

$$y_{k+1} = y_k + \left(\sum_{i=1}^m \alpha_i k_i\right) \Delta x.$$

Простейший метод Рунге – Кутта I порядка ($m = 1$) – это метод Эйлера. Наиболее *распространенный* в программном обеспечении алгоритмических языков – "стандартный" метод Рунге – Кутта IV порядка использует четыре значения функции, вычисленные для двух промежуточных точек на шаге (в середине) и обеих крайних, и соответствующий набор коэффициентов α_i :

$$k_1 = f(x_k, y_k); \quad k_2 = f\left(x_k + \frac{1}{2} \Delta x, y_k + k_1 \frac{1}{2} \Delta x\right);$$

$$k_3 = f\left(x_k + \frac{1}{2} \Delta x, y_k + k_2 \frac{1}{2} \Delta x\right); \quad k_4 = f\left(x_k + \Delta x, y_k + k_3 \Delta x\right);$$

$$y_{k+1} = y_k + (k_1 + 2k_2 + 2k_3 + k_4) \frac{1}{6} \Delta x.$$

Наиболее *экономичным* из методов Рунге – Кутта является метод II порядка следующего вида:

$$k_1 = f(x_k, y_k); \quad k_2 = f(x_k + \Delta x, y_k + k_1 \Delta x); \quad y_{k+1} = y_k + (k_1 + k_2) \frac{1}{2} \Delta x,$$

который по форме совпадает со вторым из приведенных выше модифицированных методов Эйлера.

Все методы Рунге – Кутта отличаются *устойчивостью* и возможностью *контроля погрешности* и *изменения шага* интегрирования. Однако по сравнению с методами Адамса того же порядка данные методы менее экономны, поскольку вычисленные для одного шага интегрирования значения функции нигде больше не используются. Поэтому применение методов Рунге – Кутта высоких порядков оправдано только тогда, когда необходима высокая точность или когда значения функции вычисляются сравнительно просто.

4.3 Приемы контроля математических моделей

Разработка математических моделей – трудоемкий процесс, сопряженный с подбором частных согласованных моделей, адекватных в своих областях, с идентификацией по результатам эксперимента. Поэтому такой дорогостоящий

продукт нуждается в постоянном контроле на всех стадиях разработки. К основным приемам контроля математических моделей можно отнести следующие.

1 *Контроль размерностей*, который позволяет избежать несогласованностей в формулах основных законов природы и закономерностей объекта и подготовить их к применению в алгоритмах для вычислительной техники. Для контроля размерностей следует соблюдать три правила:

– знаки $+$, $-$, $<$, $>$, \leq , \geq , $=$ могут связывать величины только одной размерности;

– аргументами трансцендентных функций должны быть безразмерные величины;

– во всех расчетных формулах следует применять одну систему единиц измерения.

Так, например, в выражении $e^{\alpha t}$ показатель степени должен быть безразмерным, т. е. α и t безразмерны или имеют взаимно обратные размерности. В эмпирических формулах коэффициенты должны иметь размерность. Внесистемные единицы измерения следует перевести в применяемую систему.

2 *Контроль основных законов природы*, прежде всего законов сохранения, необходимый в моделях, не претендующих на всеобъемлющее описание оригинала, или в моделях, использующих численные методы вычисления. Так, например, если в модели используется только дифференциальное уравнение движения (2-й закон Ньютона), то разностная схема для его интегрирования должна строиться так, чтобы это уравнение, продифференцированное численным образом по времени, давало бы закон сохранения энергии с учетом особенностей явления.

3 *Контроль качественного поведения зависимостей*. Его необходимо проводить во всех тех случаях, когда о промежуточных результатах можно что-либо сказать. Проверку качественного поведения проводят *на контрольных примерах*, как правило, для всей модели или для ее законченных частей, имеющих самостоятельное значение и смысл. В любом случае о поведении оригинала должна быть достоверная информация, как для оценки адекватности, хотя, может быть, и неполная. Используются три вида контрольных примеров: простейшие случаи, случаи особого поведения (например, резонанс) и наиболее общие случаи, исследованные в специальных экспериментах. В отличие от задачи идентификации проверка на контрольных примерах дает лишь общий вывод о качественной правильности модели.

4 *Контроль математической замкнутости* задачи, т. е. соответствие количества уравнений количеству неизвестных. Действительно, без этого просто невозможно "решить задачу".

5 ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

5.1 Основы теории вероятностей

Приступая к изучению реально существующего объекта, мы сталкиваемся с вопросом, какая информация и в каком виде нам необходима. Любой изучаемый объект многообразен. Число факторов, влияющих на него, бесконечно. Проявления этих факторов могут быть неоднозначны или случайны. Как выбрать основное? Общие соображения для ответа на этот вопрос может дать четко сформулированная цель исследований. Такие соображения можно назвать *постановкой задачи отбора информации*.

Более конкретное представление о наборе рассматриваемых факторов, обоих количественных характеристиках может быть получено только в результате нескольких *последовательных* приближений эксперимента, сопровождающихся постепенным "созреванием" модели объекта. При этом для достижения даже четко сформулированной цели исследований не всегда ясно, когда можно остановиться, а когда необходимо углублять свое представление об изучаемом объекте, с какой точностью и как необходимо регистрировать данные наблюдений.

Было бы очень кстати иметь аппарат для *строгого обоснования отбора информации*.

Таким образом, прослеживается аналогия между объектом наблюдения и зарегистрированными результатами, с одной стороны, и понятиями генеральной совокупности и выборки, с другой. То есть можно почерпнуть недостающий исследовательский аппарат из математической статистики.

Удобное представление результатов расчетов и экспериментов подразумевает не только компактность, но и информативность, а также возможность получения *достоверных* выводов по результатам анализа, т.е. *адекватных моделей*. Это понятие достоверности опять приводит нас к необходимости применения *статистических методов* для отбора и обработки информации.

5.1.1 Необходимые сведения из теории случайных событий

Для количественной оценки различных показателей исследуемого объекта используют понятия случайного события, случайной величины и случайного процесса.

Элементарным исходом ω называется любой мысленно возможный неразложимый результат вероятностного эксперимента E .

Множество всех мыслимых взаимоисключающих результатов вероятностного эксперимента E образуют пространство элементарных событий Ω .

Случайным называется такое событие, о котором нельзя заведомо точно сказать, произойдет оно или нет. Например, занос автомобиля на скользкой

дороге; совпадение плохой освещенности дороги и дождливой погоды, отказ выключателя при возникновении короткого замыкания в электропроводке и т.д. Такие события обладают какой-то степенью возможности: одни – большей, другие – меньшей. Чтобы качественно сравнить между собой события по *степени их возможности*, нужно с каждым событием связать определенное число, которое тем больше, чем более возможно событие (его вероятность).

Вероятностью случайного события (СС) A называется функция $P(A)$, характеризующая меру объективной (не зависящей от воли исследователя) возможности наступления этого события и удовлетворяющая для каждого СС трём аксиомам (Колмогорова):

- 1) вероятность удовлетворяет неравенству $0 \leq P(A)$;
- 2) вероятность появления достоверного события равна 1, т.е. $P(\Omega) = 1$;
- 3) вероятность суммы счётного множества попарно несовместных событий A_i , $i = 1, \infty$ равна сумме вероятностей этих событий, т.е. $P(\cup A_i) = \sum P(A_i)$.

По классическому способу вычисления вероятности вероятность СС A [$P(A)$] равна отношению числа элементарных исходов m , благоприятных событию A , к общему количеству элементарных исходов n пространства Ω , т.е. $P(A) = m / n$, а учитывая, что A и Ω – множества (элементарных исходов), можно записать: $P(A) = |A| / |\Omega|$.

Ограничения классического способа вычисления вероятности: все элементарные исходы вероятностного эксперимента E должны быть равновероятными, т. е. $P(\omega_i) = P(\omega_j)$, $\forall i, j$; количество элементарных исходов (размерность пространства Ω) ограничено или счётно и чтобы $|A| / |\Omega| \neq 0 / 0 \neq \infty / \infty$.

При одновременном изучении двух или нескольких событий различают события совместные и несовместные. Несовместными называются события, произведение которых является пустым множеством, т.е. невозможным событием. На практике это означает, что несовместные события вместе произойти не могут. События называются совместными, если они могут произойти одновременно. Пример совместного события – одновременный отказ двух и более элементов в один и тот же момент времени в относительно простой последовательной схеме.

Несколько событий образуют *полную группу событий*, если в результате опыта обязательно должно произойти хотя бы одно из них. На основании введенных понятий формулируются следующие основные теоремы теории вероятностей, которые применяются при решении задач надежности электроснабжения.

Теорема сложения вероятностей. Пусть A и B – произвольные случайные события, принадлежащие Ω , тогда вероятность суммы этих событий равна сумме вероятностей этих событий без вероятности их совместного появления (рисунок 5.1), т.е.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Вероятность суммы n несовместных событий равна сумме вероятностей этих событий:

$$P(A) = \sum_{i=1}^n P(A_i),$$

где $A = \sum_{i=1}^n A_i$.

Вероятность суммы любого числа событий

$$P\left(\sum_{i=1}^n A_i\right) = \sum_i P(A_i) - \sum_{i,j} P(A_i A_j A_k) - \dots (-1)^{n-1} \sum P(A_i A_j \dots A_n).$$

Событие B называется зависимым от события A , если появление (или неоявление) события A изменяет вероятность появления события B . Если происхождение A не изменяет вероятности появления B , событие B называется независимым от события A .

Теорема умножения вероятностей. Вероятность произведения двух событий равна произведению вероятности одного из этих событий на условную вероятность другого, в предположении, что первое уже произошло: $P(A \cap B) = P(A)P(B|A) = P(B \cap A) = P(B)P(A|B)$. При этом неважно, какое из событий считать первым, а какое вторым.

В общем виде

$$P(A_1 A_2 \dots A_n) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 A_2) \cdot \dots \cdot P(A_n/A_1 A_2 \dots A_{n-1}).$$

Пример. Вероятность выхода из строя измерительного прибора равна p . Для повышения надёжности в прибор поставлены m дублирующих ветвей. Определить, во сколько раз (k) увеличится надёжность прибора, если под надёжностью понимать вероятность безотказной работы.

Решение. Вероятность того, что откажут все параллельные ветви (событие A),

$$P(A) = \prod_{i=1}^m p_i = p^m.$$

Вероятность того, что не откажет хотя бы одна из параллельных ветвей (событие B) $P(B) = 1 - P(A) = 1 - p^m$.

Надёжность одной ветви $r = 1 - p$.

Искомое увеличение надёжности прибора

$$k = \frac{P(B)}{r} = \frac{1 - p^m}{1 - p}.$$

Дополнение. При относительно малых вероятностях повреждений, которые характерны для элементов ЭЭС, например при $p = 0,01$,

$$k_2 = \frac{1 - (0,01)^2}{1 - 0,01} = 1,01, \quad k_3 = \frac{1 - (0,01)^3}{1 - 0,01} = 1,0101.$$

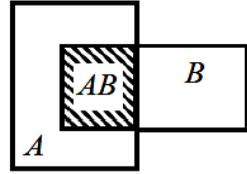


Рисунок 5.1 – Иллюстрация совместности двух событий (диаграмма Венна)

Формула полной вероятности. Пусть требуется определить вероятность некоторого СС A , которое может произойти или не произойти вместе с некоторой из гипотез H_1, H_2, \dots, H_n . Тогда вероятность указанного события можно вычислить по формуле $P(A) = \sum_{i=1}^n P(H_i) \cdot P(A|H_i)$, т.е. как сумму произведений вероятности каждой гипотезы на условную вероятность СС A при этой гипотезе.

Пример. Светодиодные светофоры закупаются на трех заводах. Предположим, что вероятность выпуска на первом заводе равна 0,6, вторым – 0,3, третьим – 0,1.

Вероятности того, что при определённых условиях работы изделие сохранит работоспособное состояние, для первого, второго и третьего заводов соответственно равны 0,99; 0,98; 0,97. Чему равна вероятность того, что поступивший светофор сохранит работоспособное состояние?

Решение. Обозначим событие A – поступивший светофор сохранит работоспособное состояние. Это изделие может оказаться с первого завода (событие H_1), второго (H_2) или третьего (H_3). Интересующее нас событие A имеет вероятность

$$P(A) = \sum_{i=1}^3 P(H_i) \cdot P(A|H_i) = 0,6 \cdot 0,99 + 0,3 \cdot 0,98 + 0,1 \cdot 0,97 = 0,985.$$

Теорема гипотез (формула Байеса). Произведён опыт, в результате которого наблюдалось событие A , которое может произойти или не произойти вместе с одной из гипотез H_1, H_2, \dots, H_n . Причём известны априорные (т.е. доопытные) вероятности этих гипотез – $P(H_1), P(H_2), \dots, P(H_n)$. Пусть известно, что событие A произошло. Требуется определить вероятности событий H_1, H_2, \dots, H_n после опыта. На основании теоремы умножения и формулы полной вероятности имеем:

$$P(H_i|A) = (P(H_i) P(A|H_i)) / (\sum [P(H_i) P(A|H_i)]).$$

5.1.2 Случайные величины и законы их распределения

Если случайные события характеризуют результат вероятностного эксперимента качественно (событие может произойти или нет), то случайные величины – количественно. Например, безотказная работа некоторого устройства в течение гарантийного срока – случайное событие, а время наработки устройства на отказ – случайная величина.

Случайной величиной называется функция $\xi = \xi(\omega)$, которая каждому элементарному исходу ω пространства элементарных событий Ω вероятностного эксперимента E ставит в соответствие некоторое действительное число x .

В зависимости от размерности пространства элементарных исходов Ω вероятностного эксперимента случайные величины могут быть дискретными и непрерывными.

Непрерывной называется случайная величина ξ , которая в результате эксперимента может принимать все значения из некоторого промежутка или всей числовой оси. Множество значений непрерывной случайной величины, соответствующее пространству Ω , несчетно. Примерами непрерывных случайных величин являются: время безотказной работы элементов, устройств, агрегатов, систем; время вынужденного простоя оборудования из-за отказов; уровень того или иного технического параметра и т.д.

Дискретной называется случайная величина ξ , которая в результате эксперимента может принимать только определенные, изолированные друг от друга, значения. Множество значений дискретной случайной величины конечно или счетно. Примерами дискретных случайных величин являются: число неисправных элементов, устройств, агрегатов из общего числа находящихся в эксплуатации; число дефектных изделий в какой-либо партии продукции; количество повреждений элементов какого-либо оборудования в единицу времени и т.д.

Случайная величина будет полностью определена с вероятностной точки зрения, если мы укажем, с какой вероятностью наступает каждое из случайных событий $\{\xi = x_i\}$. Этим мы установим так называемый закон распределения случайной величины ξ .

Законом распределения случайной величины ξ называется правило, которое каждому возможному значению x величины ξ ставит в соответствие вероятность появления данного значения. Закон распределения полностью характеризует случайную величину ξ с вероятностной точки зрения, т.е. определяет множество значений, которое может принимать величина, и то, с какими вероятностями величина ξ принимает значения x_1, x_2, \dots

Закон распределения случайной величины ξ может быть задан таблично, графически и аналитически. Непосредственное перечисление значений возможно только для дискретных случайных величин, множество которых конечно или счётно.

Рядом распределения называется таблица, в которой непосредственно указаны возможные значения случайной величины ξ (x_1, x_2, \dots, x_n) и соответствующие им вероятности (таблица 5.1).

Таблица 5.1 – Ряд распределения дискретных случайных величин

| | | | | | |
|----------------|-------|-------|-----|-------|-----|
| x_i | x_1 | x_2 | ... | x_n | ... |
| $P(\xi = x_i)$ | p_1 | p_2 | ... | p_n | ... |

Для наглядности закон распределения дискретных величин можно представить графически. По оси абсцисс откладывают возможные значения случайной величины, а по оси ординат – вероятности данных значений (рисунок 5.2).

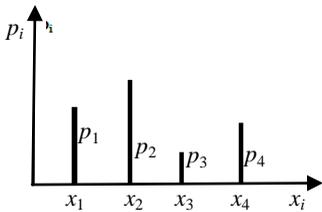


Рисунок 5.2 – Графический способ задания закона распределения

Табличный и графические способы задания законов распределения применимы только к дискретным случайным величинам. Универсальным способом задания закона распределения является функция распределения.

Функцией *распределения* случайной величины ξ называется функция $F(x)$ действительного аргумента x , которая каждому значению аргумента x ставит в соответствие вероятность того, что величина ξ примет значение

меньше x , т.е. вычисляет вероятность события $\{\xi < x\}$.

Таким образом, $F(x) = P(\xi < x)$.

Функция *плотности распределения* непрерывной случайной величины – первая производная от функции распределения:

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

Вероятностный смысл функции $f(x)$ заключается в том, что она указывает как часто возможные значения непрерывной случайной величины ξ группируются в окрестности точки x .

Случайные величины в вероятностном смысле полностью характеризуются законами распределения. Однако на практике знание закона распределения случайной величины часто оказывается излишним. Иногда бывает достаточно знать лишь отдельные числовые параметры, характеризующие существенные черты закона распределения исследуемой случайной величины или некоторые её характерные значения.

Характеристики, выражающие в сжатой форме наиболее существенные особенности распределения случайной величины, называются *числовыми характеристиками* случайной величины. Все числовые характеристики случайных величин разделяют на характеристики положения, рассеяния, характеристики асимметрии и эксцесса (таблица 5.2) [11].

Таблица 5.2 – Числовые характеристики случайных величин

| Числовые характеристики случайных величин | | | |
|---|---------------------------------|------------------------|----------------------|
| положения | рассеяния | асимметрии | эксцесса |
| Математическое ожидание | Среднеквадратическое отклонение | Коэффициент асимметрии | Коэффициент эксцесса |
| Мода | Дисперсия | | |
| Медиана | | | |

Математическим ожиданием случайной величины ξ называется число $M[\xi]$, характеризующее среднее значение случайной величины с учётом вероятностей её значений.

Математическое ожидание дискретной случайной величины ξ вычисляется по формуле $M[\xi] = \sum_{i=1}^n x_i P(\xi = x_i)$, а непрерывной случайной величины – по формуле $M[\xi] = \int_{-\infty}^{+\infty} x f(x) dx$.

Мода – наиболее вероятное значение случайной величины ξ , обозначаемое $\text{Mod}[\xi]$.

Медианой называется значение случайной величины ξ , обозначаемое $\text{Med}[\xi]$, для которого равновероятно случайная величина ξ примет большее или меньшее значение, т.е. $P(\xi < \text{Med}[\xi]) = P(\xi > \text{Med}[\xi]) = 0,5$. Поэтому медиану называют средневероятным значением случайной величины.

Дисперсией случайной величины ξ называется число $D[\xi]$, характеризующее меру рассеяния значений случайной величины вокруг ее математического ожидания. Дисперсия дискретной случайной величины

$$D[\xi] = \sum_{i=1}^n x_i^2 P(\xi = x_i) - M^2[\xi];$$

для непрерывной случайной величины

$$D[\xi] = \int_{-\infty}^{+\infty} x^2 f(x) dx - M^2[\xi].$$

Среднеквадратическим отклонением случайной величины ξ называется число $\sigma[\xi]$, равное положительному значению квадратного корня из дисперсии, т.е. $\sigma[\xi] = +\sqrt{D[\xi]}$. Таким образом, среднеквадратическое отклонение случайной величины имеет размерность, равную размерности самой случайной величины. Среднеквадратическое отклонение $\sigma[\xi]$, как и дисперсия $D[\xi]$, характеризует степень разброса значений случайной величины ξ вокруг ее математического ожидания $M[\xi]$.

Коэффициент вариации – число, характеризующее, насколько хорошо математическое ожидание представляет ряд её возможных значений. Вычисляется по формуле $V[\xi] = \sigma[\xi] / M[\xi]$.

Коэффициентом асимметрии (скошенности) распределения называется число, вычисляемое по формуле

$$\beta_1[\xi] = \frac{M[(\xi - M[\xi])^3]}{\sigma^3}.$$

Коэффициентом эксцесса случайной величины ξ называется число $\beta_2[\xi]$, характеризующее островершинность распределения случайной величины ξ по сравнению с нормальным распределением и определяемое по формуле

$$\beta_2[\xi] = \frac{M[(\xi - M[\xi])^4]}{\sigma^4} - 3.$$

Некоторые распределения случайных величин. Биномиальное распределение. Говорят, что случайная величина ξ имеет биномиальный закон распределения, т. е. $\xi \sim Bi(n, p)$, если данная величина дискретна и определяет число успехов k в схеме n испытаний Бернулли. Очевидно, что случайная величина ξ , имеющая биномиальное распределение, принимает только целые значения на отрезке $[0, n]$ с вероятностями, определяемыми формулой Бернулли: $P_k = P\{\xi = k\} = C_n^k p^k (1-p)^{n-k}$.

Биномиальное распределение характеризуется двумя параметрами: числом проводимых экспериментов n и вероятностью успеха p в каждом испытании. Примерами СВ, имеющей биномиальное распределение, является число бракованных сварных швов автомобиля.

Основные числовые характеристики случайных величин, распределенных по биномиальному закону, определяются следующими выражениями:

$$M[\xi] = np; \quad D[\xi] = np(1-p).$$

Пример. На диспетчерском пункте работает четыре однотипных компьютера. Вероятность отказа в работе программного обеспечения каждого из них $p = 0,03$. Составить ряд распределения вероятного числа отказов программного обеспечения компьютеров.

Решение. Число повреждённых компьютеров является дискретной случайной величиной. Пользуясь формулой Бернулли для биномиального распределения $P_k = P\{\xi = k\} = C_n^k p^k (1-p)^{n-k}$, приведем результаты расчетов в таблице 5.3.

Таблица 5.3 – Ряд распределения дискретных случайных величин

| | | | | | |
|--------------------------------|--------|--------|--------|---------|------------|
| Число повреждённых компьютеров | 0 | 1 | 2 | 3 | 4 |
| Вероятность $P(\xi = x_i)$ | 0,9606 | 0,1164 | 0,0053 | 0,00011 | 0,00000081 |

Геометрическое распределение. На практике геометрическое распределение возникает при следующих условиях: пусть производится серия независимых опытов, в каждом из которых может произойти событие A с одной и той же вероятностью p . Опыты продолжаютя до первого появления события A . Тогда случайная величина X , определяющая *число неудач, предшествующих успеху*, распределена по геометрическому закону. Возможные значения этой случайной величины: $0, 1, 2, \dots, n, \dots$, а вероятность каждого из этих значений определяется по формуле

$$P(X = m) = q^m p,$$

где $0 \leq p \leq 1$; $q = 1 - p$; $m = 0, 1, 2, \dots, n, \dots$.

Для случайной величины, распределенной по геометрическому закону,

$$M[X] = \frac{1}{1-p}, \quad D[X] = \frac{1-p}{p^2}, \quad \sigma[X] = \sqrt{\frac{1-p}{p^2}}, \quad \text{mod}[X] = 0.$$

Пример. Вероятность изготовления нестандартного изделия при некотором технологическом процессе равна 0,06. Контролер из отдела технического контроля берет из партии изделие и сразу проверяет его качество. Если оно оказывается нестандартным, то дальнейшие испытания прекращаются, а партия задерживается. Если же изделие оказывается стандартным, то контролер проверяет следующее изделие и т. д. Записать закон распределения случайной величины X – числа стандартных изделий, проверенных до выявления брака.

Решение. Условие задачи соответствует проведению независимых испытаний, в каждом из которых с вероятностью $p = 0,06$ может осуществиться событие $A = \{\text{обнаружено нестандартное изделие}\}$. В этом случае *неудача* – обнаружение стандартного изделия, *успех* – обнаружение нестандартного изделия. Случайная величина X – число стандартных изделий, проверенных до выявления брака, распределена по геометрическому закону. Возможные значения этой случайной величины: 0, 1, 2, 3, ..., m , По условию $p = 0,06$, $q = 1 - 0,06 = 0,94$.

Вероятности значений определяются по формулам:

$P(X = 0) = q^0 p = 0,94^0 \cdot 0,06 = 0,06$ (то есть нестандартное изделие будет обнаружено сразу же при проверке первого изделия, при этом число стандартных изделий, проверенных до появления брака, будет равно 0);

$P(X = 1) = q^1 p = 0,94 \cdot 0,06 = 0,0564$ (то есть нестандартное изделие будет обнаружено при проверке второго изделия, при этом число стандартных изделий, проверенных до появления брака, будет равно 1) и т. д.

Закон распределения случайной величины X : $P(X = m) = 0,94^m \cdot 0,06$.

Распределение Пуассона. Если в схеме Бернулли число испытаний бесконечно велико ($n \rightarrow \infty$), а вероятность успеха в каждом испытании стремится к нулю ($p \rightarrow 0$) таким образом, что $np = \lambda = \text{const}$, то вероятность появления в схеме n испытаний Бернулли ровно k успехов определяется предельной теоремой Пуассона $\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}$. В данном случае говорят,

что случайная величина ξ , определяющая число успехов k в схеме n испытаний Бернулли, имеет Пуассоновский закон распределения, т.е. $\xi \sim P(\lambda)$.

Таким образом, закон распределения Пуассона является предельным случаем биномиального закона распределения при $n \rightarrow \infty$, $p \rightarrow 0$, $np = \lambda = \text{const}$.

Очевидно, что распределение Пуассона характеризуется единственным параметром $\lambda = np$; а случайная величина ξ , имеющая Пуассоновское распределение, принимает только целые значения на полуинтервале $[0, \infty)$ с вероятностями, определяемыми предельной теоремой Пуассона:

$$P_k = P\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Примерами случайных величин, имеющих Пуассоновское распределение, являются число сигналов, поступающих на диспетчерский пульт управления в течение часа и др. [12].

Основные числовые характеристики случайных величин, которые имеют Пуассоновский закон распределения, определяются следующими выражениями: $M[\xi]=\lambda$; $D[\xi]=\lambda$.

Пример 1. Выпущена партия шин 1000 штук. Вероятность того, что шина имеет брак, $p = 0,003$. Найти вероятность того, что в партии ровно пять бракованных шин.

Решение. Вероятность числа бракованных шин рассчитывается по формуле Пуассона с параметром $\lambda = np = 1000 \cdot 0,003 = 3$. Значит,

$$P(5) = \frac{3^5}{5!} e^{-3} = 0,1008.$$

Пример 2. Определить вероятность того, что за 1000 ч работы произойдет два отказа в сложном изделии, если известно, что интенсивность отказов $a = 3 \cdot 10^{-3} \text{ ч}^{-1}$.

Решение. Находим параметр Пуассоновского распределения и определяем вероятность двух отказов в сложном изделии по формуле Пуассона:

$$\lambda = a p = 3 \cdot 10^{-3} \cdot 1000 = 3; \quad P(2) = \frac{3^2}{2!} e^{-3} = 0,224.$$

Экспоненциальное распределение. Говорят, что СВ имеет показательное распределение, если она непрерывна и имеет функцию плотности распределения (рисунок 5.3) $f(x) = \begin{cases} 0, & \text{если } x \leq 0, \\ \lambda e^{-\lambda x}, & \text{если } x > 0. \end{cases}$

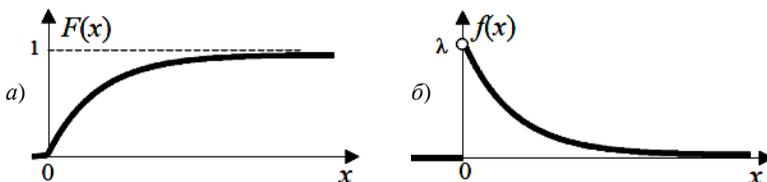


Рисунок 5.3 – Показательный закон распределения
 а – функция распределения $F(x)$ случайной величины ξ ;
 б – функция плотности распределения $f(x)$ случайной величины ξ

Основные числовые характеристики случайной величины ξ , имеющей показательный закон распределения, определяются следующими выражениями:

$$M[\xi] = \sigma[\xi] = 1 / \lambda; \quad D[\xi] = 1 / \lambda^2.$$

Считается, что время наработки различных технических устройств на отказ имеет показательное распределение (наиболее вероятно, что устройство откажет сразу после включения, менее вероятно, что оно проработает долго). Функция надежности $R(t)$ используется в качестве модели вероятности безотказной работы за то же время:

$$R(t) = 1 - F(t) = e^{-\lambda t},$$

где λ – интенсивность отказов.

Плотность вероятности отказов $f(t) = F'(t) = \lambda e^{-\lambda t}$. Среднее время до возникновения отказа или среднее время безотказной работы является математическим ожиданием экспоненциального распределения, т.е. величиной, обратной интенсивности отказов.

Таким образом, зная среднее время безотказной работы T_{cp} или постоянную интенсивность отказов λ , можно в случае экспоненциального распределения найти вероятность безотказной работы для интервала времени от момента включения элемента, устройства или агрегата до любого заданного момента t .

Равенство среднеквадратического отклонения среднему времени работы – характерный признак экспоненциального распределения.

Пример. Время безотказной работы амортизатора при перегрузке распределено по экспоненциальному закону $f(t) = 0,02e^{-0,02t}$, где t – время, ч. Найти вероятность того, что амортизатор в перегрузочном режиме проработает безотказно в течение 100 ч.

Решение. В качестве модели вероятности безотказной работы амортизатора воспользуемся формулой экспоненциального закона распределения. Значит, $R(t) = e^{-\lambda t} = e^{-0,02 \cdot 100} = 0,14$.

Среднее время безотказной работы $T_{cp} = 1/\lambda = 1/0,02 = 50$ ч.

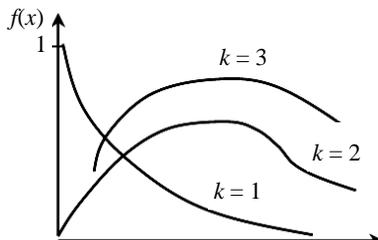
Гамма-распределение. Применяется для моделирования параметров надёжности в первый период эксплуатации и в период износа. Если отказ устройства или системы возникает тогда, когда происходит не менее k отказов его элементов, а отказы элементов подчинены экспоненциальному закону с параметром λ_0 , плотность вероятности отказа определяется как

$$f(t) = \frac{\lambda_0^k t^{k-1}}{\Gamma(k)} e^{-\lambda_0 t},$$

где λ_0 – исходная интенсивность отказов элементов устройства, отказ которого вызывается отказом k элементов $\Gamma(k) = \int_0^{\infty} x^{k-1} e^{-x} dx$.

Этому распределению подчиняется время работы резервированных устройств, отказ которых вызывается отказом k их элементов.

При $k = 1$ гамма-распределение совпадает с экспоненциальным распределением. При увеличении k гамма-распределение приближается к симметричному распределению, что и показывает рисунок 5.4.



Распределение Вейбулла. Применяется в следующей форме:

Рисунок 5.4 – Изменение формы функции плотности вероятности

$$f(t) = \frac{a}{t_0} t^{a-1} e^{-\frac{t}{t_0}}$$

где t_0 – приведённое значение среднего времени безотказной работы; $a > 0$ – параметр формы распределения.

Вероятность отсутствия отказов за время t $P(t) = \lambda e^{-\lambda_0 t^a}$.

Интенсивность отказов $\lambda(t) = a \lambda_0 e^{-a-1}$.

Распределение Вейбулла имеет большое значение в моделях надёжности. Считается, что время безотказной работы устройств в периоде приработки, когда интенсивность отказов устройства со временем уменьшается (отказы устройства на данном этапе вызваны конструкторскими ошибками, недостаточным качеством монтажа и комплектующих системы), имеет распределение Вейбулла с параметром $\alpha < 1$ (рисунок 5.5), соответствующее

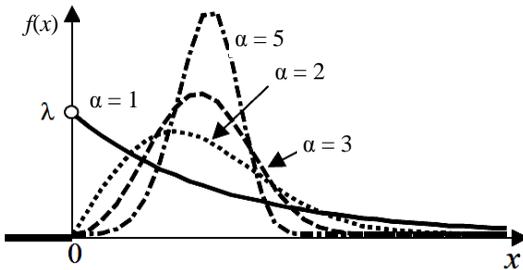


Рисунок 5.5 – Изменение формы функции плотности вероятности распределения Вейбулла

убывающей функции плотности вероятности. В периоде нормальной эксплуатации, когда интенсивность отказов устройства постоянна, время наработки устройства на отказ совпадает с показательным распределением (распределение Вейбулла с параметром $\alpha = 1$). На этапе старения, когда интенсивность отказов устройств возрастает из-за выработки ресурса некоторых компонентов, время наработки устройства на отказ имеет распределение Вейбулла с параметром $\alpha > 1$ [12].

Равномерный закон распределения. Непрерывная случайная величина, которая принимает значения, только принадлежащие отрезку $[a, b]$ с постоянной плотностью распределения, называется распределенной по равномерному закону.

Функция плотности распределения вероятностей (рисунок 5.6) определяется соотношением

$$f(x) = \begin{cases} 0 & \text{при } x < a; \\ \frac{1}{b-a} & \text{при } a \leq x \leq b; \\ 0 & \text{при } x > b. \end{cases}$$

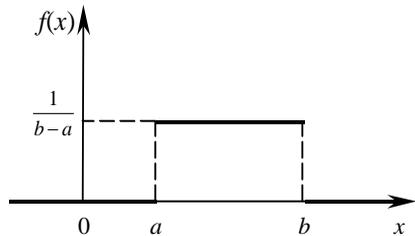


Рисунок 5.6 – График функции $f(x)$ равномерного распределения

Математическое ожидание случайной величины X , распределенной по равномерному закону на участке $[a, b]$, как следует из механической интерпретации (центр массы), равно абсциссе середины участка: $M[X] = (a + b) / 2$. Дисперсию случайной величины X также можно найти, исходя из механической интерпретации (момент инерции распределения относительно центра массы): $D[X] = (b - a)^2 / 12$.

Примером случайной величины, которая имеет равномерный закон распределения, является время ожидания регулярных событий, например, время ожидания троллейбуса, ошибки округления величин с помощью прибора.

Пример. Поезда метрополитена идут с интервалом в 4 минуты. Пассажир приходит на платформу поезда в произвольный момент времени. Найти вероятность того, что он будет ожидать прихода поезда не более одной минуты. Найти среднее время ожидания поезда пассажиром, вычислить дисперсию и среднее квадратическое отклонение времени ожидания поезда пассажиром.

Решение. Рассмотрим случайную величину X – время ожидания пассажиром поезда. Все возможные значения данной случайной величины принадлежат отрезку $[0; 4]$, и, согласно условию, все эти значения равновозможны. Следовательно, случайная величина распределена по равномерному закону с параметрами $a = 0$ и $b = 4$. Функция плотности распределения вероятностей данной случайной величины:

$$f(x) = \begin{cases} 0 & \text{при } x \in (-\infty; 0) \text{ и } x \in (4; \infty); \\ 0,25 & \text{при } x \in [0; 4]. \end{cases}$$

Найдем вероятность того, что пассажир будет ожидать поезд не более одной минуты:

$$P(0 \leq X \leq 1) = \int_0^1 f(x) dx = \int_0^1 0,25 dx = 0,25.$$

Среднее время ожидания прихода поезда пассажиром

$$M[X] = (a + b) / 2 = (0 + 4) / 2 = 2 \text{ [мин]}.$$

$$\text{Дисперсия } D[X] = \frac{(b - a)^2}{12} = \frac{(4 - 0)^2}{12} = 1,333 \text{ [мин}^2\text{]}.$$

$$\text{Среднее квадратическое отклонение } \sigma[X] = \frac{b - a}{2\sqrt{3}} = \frac{4 - 0}{2\sqrt{3}} = 1,155 \text{ [мин]}.$$

Нормальное (гауссовское) распределение. Фундаментальное значение в теории вероятностей имеет *центральная предельная теорема*, доказывающая, что сумма произвольно распределенных независимых случайных величин при условии *одинакового* их влияния распределена по *нормальному закону*.

Случайная величина X называется распределенной по нормальному закону, если ее плотность распределения вероятностей задается формулой

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

где m – математическое ожидание;

σ – среднее квадратическое отклонение.

Кривая функции плотности вероятности нормального распределения имеет колоколообразную форму, симметричную относительно прямой $x = m$, и при $x \rightarrow -\infty$ и $x \rightarrow \infty$ эта кривая асимптотически приближается к оси абсцисс. Общий вид графика функции плотности распределения вероятностей $f(x)$ для произвольных значений m и σ изображен на рисунке 5.7.

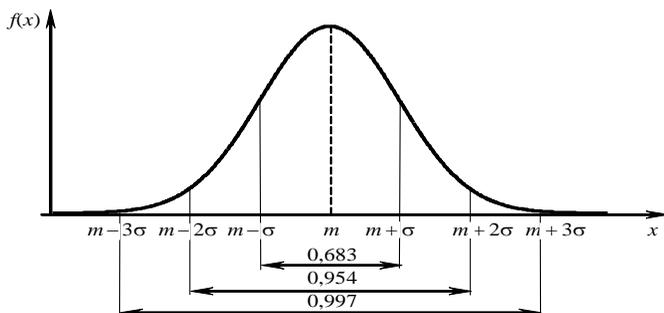


Рисунок 5.7 – График функции плотности распределения вероятностей нормального распределения

Главная особенность этого закона состоит в том, что он является *предельным*; к нему при определенных условиях приближаются другие законы распределения; в теории надёжности применяется для оценки постепенных отказов. Результаты измерения длины, времени, ошибки измерения и многие другие случайные величины имеют нормальное или близкое к нормальному распределение.

С помощью функции Лапласа (табулированная функция) вероятность попадания нормально распределенной случайной величины X на участок от α до β выражается формулой

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right).$$

Например, пусть производятся измерения некоторой физической величины. Любое измерение дает лишь приближенное значение измеряемой величины, так как на результат испытания оказывают влияние очень многие, не зависящие друг от друга случайные факторы: температурные колебания в помещении, воздействия окружающей среды, неточность измерительной шкалы, смена контрольного персонала и т. д.

В зависимости от источников появления ошибок различают системати-

ческие и случайные ошибки. К *систематическим ошибкам* относятся, например, односторонние отклонения, вызванные, скажем, изменением настройки измерительного прибора или сменой контрольного персонала. Эти ошибки можно устранить путем систематического изучения причины их возникновения.

Случайные ошибки вызваны влиянием множества различных неконтролируемых причин: температурных колебаний, влажности, вибраций в окружающей среде и т. п. Каждый из этих факторов порождает ничтожную «частную ошибку». Но поскольку число этих факторов очень велико, совокупное их действие порождает уже заметную «суммарную ошибку». Полностью исключить воздействие этих факторов невозможно, так как нельзя заранее предусмотреть степень их влияния на результат конкретного измерения. Подобные случайные ошибки вызывают при измерениях отклонения в обе стороны от истинного значения.

Рассматривая общую ошибку как сумму очень большого числа взаимно независимых частных ошибок, мы вправе заключить, что суммарная ошибка имеет распределение, близкое к нормальному. Опыт подтверждает справедливость такого заключения.

Пример. Случайная величина X распределена по нормальному закону с параметрами $M[X] = m$, $\sigma[X] = \sigma$. Найти вероятность того, что случайная величина X будет принимать значения, удаленные от математического ожидания не более чем на σ .

Решение. Для вычисления искомых вероятностей воспользуемся формулой на основе функции Лапласа:

$$P(m - \sigma < X < m + \sigma) = \Phi\left(\frac{m + \sigma - m}{\sigma}\right) - \Phi\left(\frac{m - \sigma - m}{\sigma}\right) = \Phi(1) - \Phi(-1) = 0,34134 - (-0,34134) = 0,68268.$$

Распределение χ^2 . Случайная величина χ^2 определяется единственным параметром ν , имеет функцию плотности распределения (рисунок 5.8).

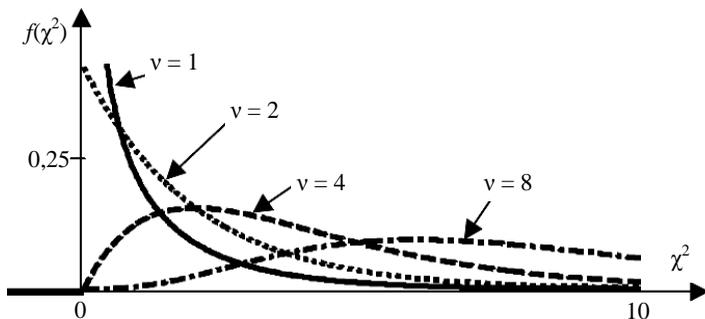


Рисунок 5.8 – Функция плотности распределения χ^2 для различных значений параметра ν

$$f(\chi^2) = \begin{cases} 0, & \text{если } \chi^2 < 0, \\ \frac{1}{2^{\frac{\nu}{2}} \Gamma(\nu/2)} (\chi^2)^{\frac{\nu}{2}-1} \cdot e^{-\frac{\chi^2}{2}}, & \text{если } \chi^2 \geq 0, \end{cases}$$

где $\Gamma(\nu/2)$ – гамма-функция (интеграл Эйлера второго рода).

Распределение χ^2 применяется при решении задач, связанных с оценкой параметров надежности, определяемых при испытаниях или эксплуатации оборудования.

Изменение интенсивности отказов элементов во времени. Типичная функция интенсивности отказов во времени (в течение срока службы объекта) имеет U-образный характер (рисунок 5.9). В начальный период I преобладают приработочные отказы. После него наступает наиболее продолжительный период нормальной эксплуатации II, в котором на объект воздействуют случайные факторы. Последние вызывают внезапные отказы, интенсивность которых в период нормальной эксплуатации практически не зависит от времени.



Рисунок 5.9 – Кривая жизни изделия

В период старения и износа III в основном имеют место постепенные отказы, возникающие вследствие накопления ухудшений физико-химических свойств объекта.

5.1.3 Предельные теоремы теории вероятностей

Предельные теоремы теории вероятностей условно разделяют на две группы: закон больших чисел и центральная предельная теорема.

Закон больших чисел во всех его формах утверждает факт сходимости по вероятности тех или иных случайных величин к определенным постоянным. Говорят, что величина (или функция) $\xi(n)$ *сходится по вероятности* к

величине α , если при сколь угодно малом ε вероятность выполнения неравенства $|\xi(n) - \alpha| < \varepsilon$ с увеличением n неограниченно приближается к единице. Другими словами, с увеличением n величина $\xi(n)$ вероятнее приближается к значению α , чем отдалается от него.

Наиболее распространенными формами закона больших чисел являются теорема Бернулли, а также теорема и неравенство Чебышева.

Теорема Бернулли. При неограниченном увеличении числа опытов n относительная частота наступления случайного события A сходится по вероятности к вероятности события A , т.е.

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - P(A) \right| < \varepsilon \right\} > 1 - \delta,$$

где n – число повторных испытаний; n_A – число испытаний, в которых событие A наступило; $P(A)$ – вероятность события A ; ε и δ – произвольные сколь угодно малые положительные числа.

Обобщением теоремы Бернулли для случайных величин является теорема Чебышева. Она определяет свойство устойчивости среднего арифметического последовательности значений случайной величины.

Теорема Чебышева. При неограниченном увеличении числа независимых опытов среднее арифметическое наблюдаемых значений случайной величины ξ сходится по вероятности к ее математическому ожиданию, т.е.

$$\lim_{n \rightarrow \infty} P \left\{ |\bar{x} - M[\xi]| < \varepsilon \right\} > 1 - \delta,$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ – среднее арифметическое наблюдаемых значений величины ξ ;

$M[\xi]$ – математическое ожидание случайной величины ξ .

Неравенство Чебышева. Для произвольной случайной величины ξ с математическим ожиданием $M[\xi]$ и дисперсией $D[\xi]$ справедливо неравенство

$$P\left(\xi - M[\xi] \geq \alpha\right) \leq \frac{D[\xi]}{\alpha^2},$$

где α – произвольное положительное число.

Неравенство Чебышева ограничивает вероятность значительных отклонений случайной величины ξ от своего математического ожидания.

Пример. Пусть случайная величина ξ , характеризующая ошибку округления результата измерения напряжения на зажимах аккумулятора цифровым вольтметром, имеет равномерный закон распределения на интервале $(-2 \text{ мВ}; 2 \text{ мВ})$, т.е. $\xi \sim R(-2, 2)$. Определить наибольшее значение вероятности, с которой ошибка округления вольтметра отклонится от нуля более чем на 1,75 мВ.

Решение. Согласно формулам для математического ожидания и дисперсии равномерного закона распределения случайной величины ξ $M[\xi] = (-2 + 2) / 2 = 0$ мВ; $D[\xi] = (2 - (-2))^2 / 12 = 1,333$ мВ². В соответствии с неравенством Чебышева получим $P(|\xi - 0| \geq 1,75) \leq \frac{1,333}{1,75^2} = 0,4354$, т.е. вероятность того, что ошибка округления ξ превысит (по абсолютной величине) 1,75 мВ не превышает 0,4354.

В действительности эта вероятность определяется выражением

$$P(|\xi| > 1,75) = P(\{-2 < \xi < -1,75\} \cup \{1,75 < \xi < 2\}) = 1 - P(-1,75 < \xi < 1,75).$$

Используя формулу расчета вероятности равномерного распределения, получим

$$1 - P(-1,75 < \xi < 1,75) = 1 - \frac{1,75 - (-2)}{2 - (-2)} + \frac{-1,75 - (-2)}{2 - (-2)} = 1 - \frac{3,75}{4} + \frac{0,25}{4} = 0,125.$$

Следовательно, $P(|\xi| > 1,75) = 0,125$.

Центральная предельная теорема (теорема Ляпунова). Пусть ξ_i , $i = 1, \infty$ – последовательность независимых случайных величин, для каждой из которых существует математическое ожидание $M[\xi_i]$, дисперсия $D[\xi_i]$ и третий центральный абсолютный момент $M\left[|\xi_i - M[\xi_i]|^3\right]$. Пусть также вы-

полняется условие Ляпунова $\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n M|\xi_i - M[\xi_i]|^3}{\left(\sum_{i=1}^n D[\xi_i]\right)^{3/2}} = 0$, из которого, в частно-

сти, вытекает, что дисперсия каждой из величин ξ_i составляет малую долю суммарной дисперсии, $\sum_{i=1}^n D[\xi_i]$. Тогда при бесконечном увеличении числа n

распределение суммы величин $\eta = \sum_{i=1}^n \xi_i$ неограниченно приближается к нормальному закону распределения с математическим ожиданием $M[\eta] = \sum_{i=1}^n M[\xi_i]$ и дисперсией $D[\eta] = \sum_{i=1}^n D[\xi_i]$.

Все формы центральной предельной теоремы посвящены установлению условий, при которых возникает нормальный закон или другие законы распределения случайных величин.

На основании центральной предельной теоремы можно утверждать, что во всех случаях, когда случайная величина образуется в результате суммирования большого числа независимых или слабо зависимых случайных величин, дисперсия каждой из которых мала по сравнению с дисперсией сум-

мы, распределение этой величины оказывается практически нормальным. Поскольку перечисленные выше условия выполняются достаточно часто, то нормальный закон распределения случайных величин достаточно распространен [11].

Замечание 1. Многими нормативными документами устанавливается, что ошибки физических измерений, которые связаны с погрешностями измерительных приборов, ошибками округлений, в соответствии с центральной предельной теоремой, имеют нормальный закон распределения.

Замечание 2. В теории вероятностей доказывается теорема о том, что закон распределения случайной величины $\eta = \min\{\xi_1, \xi_2, \dots, \xi_n\}$ при $n \rightarrow \infty$ неограниченно приближается к показательному закону распределения, даже если случайные величины $\xi_1, \xi_2, \dots, \xi_n$ имеют произвольные законы распределения, но принимают лишь положительные значения, т.е. если $\xi = \min\{\xi_1, \xi_2, \dots, \xi_n\}$, где $\xi_i > 0 \forall i = 1, \dots, n$, то $\lim_{n \rightarrow \infty} \xi = \min\{\xi_1, \xi_2, \dots, \xi_n\} \sim E(\lambda)$.

5.2 Основы математической статистики

Математическая статистика – наука для разработки методов регистрации, описания и анализа экспериментальных данных наблюдения массовых случайных явлений. Центральное место в математической статистике занимают *теория оценок* и *теория проверки гипотез*. Основное правило математической статистики гласит: каждое выдвинутое предложение должно быть *оценено* и *проверено* на правдоподобие.

Всякое наблюдение дает какое-то ограниченное представление о явлении в целом, в том числе и об определенной случайной величине. Итогом такого наблюдения, т.е. отбора информации, становится *выборка* – некоторая совокупность результатов наблюдения случайной величины (*генеральной совокупности*), отобранная для исследования по определенному правилу. Таким образом, результаты наблюдения случайного явления дают лишь ограниченную информацию о случайной величине в целом. Поэтому переход от зафиксированного экспериментального факта к выводу и прогнозу – далеко не очевиден и нуждается в обосновании. Математическая статистика позволяет по результатам наблюдения частного (*выборки*) сделать некоторые обоснованные выводы о характеристиках общего (*генеральной совокупности*).

Пример. Вся продукция электролампового завода – генеральная совокупность, проверенная часть ламп – выборка. По сроку службы в среднем, вычисленному по выборке, можно судить о сроке службы ламп в среднем для всей генеральной совокупности. Но лишь "судить", "оценивать", ибо нельзя знать заранее срок службы какой-либо конкретной лампы.

Общая последовательность применения методов математической статистики (в скобках дается комментарий в современных терминах):

- 1) планирование исследований (*планирование эксперимента*, определение способа отбора информации);
- 2) конкретизация математико-статистического описания (выбор *дисперсионной* или *регрессионной модели*);
- 3) оценка параметров модели (получение *точечных* и *интервальных оценок*) и составление их выборочных (эмпирических) распределений;
- 4) изучение согласия между моделью и наблюдениями (*адекватность модели* оригиналу и проверка критериев согласия в обоснование модели);
- 5) реальное решение задачи посредством оценок параметров и критериев значимости (*статистический анализ* результатов и разработка выводов).

Первый и третий этапы представляют собой так называемую процедуру *первичной обработки информации*, которая начинается с отбора информации, включает построение гистограмм и полигонов частот (если это необходимо для визуальной оценки вида эмпирического распределения) и завершается расчетом точечных и интервальных оценок. Эти результаты служат исходным материалом для четвертого и пятого этапов – *статистического анализа*, целью которого является установление статистических закономерностей.

5.2.1 Отбор информации

Исследователь, озабоченный составлением некоторой модели явления, стремится получить из эксперимента лишь ограниченный круг параметров, ее описывающих. Под эту задачу и собирается информация. Таким образом формулируются первичные требования к отбору информации.

Однако статистический материал может содержать в себе намного больше информации, чем ставилось целью собрать. Можно не только получить значения наблюдаемых параметров с контролируемой погрешностью и с заданной вероятностью, что позволяет сделать первичная обработка информации. Умелое извлечение информации позволяет еще и быстрее достичь требуемого результата, и оценить его добротность, и даже оценить адекватность разрабатываемой модели. Но не стоит думать, что математическая статистика отвечает на все вопросы. Во-первых, она способна обрабатывать только ту информацию, которая собрана. Во-вторых, ее методы дают лишь более или менее вероятные (и эту вероятность можно оценить) результаты. Поэтому нет и не может быть математической обработки информации *вообще*, есть только аппарат для целевых исследований, невозможных без предварительных предположений о модели. То есть за человеком в любом случае остается творческий подход к выбору модели, способам отбора информации и формулировке выводов.

Для анализа информации неважно, получена она из натурального эксперимента или из вычислительного эксперимента на детерминированной либо стохастической модели. В любом случае основными *проблемами сбора и*

обработки информации являются:

- выбор существенных факторов;
- выбор процедуры отбора информации;
- обеспечение достоверности выводов по результатам анализа.

Все эти проблемы разрешимы с помощью математического аппарата статистического анализа, излагаемого в последующих параграфах.

Следующий пример иллюстрирует необходимость строгого научного подхода к сбору и обработке информации.

Пример. Социологический опрос. Пусть результаты ответов 50 человек на некоторый вопрос представлены рядом знаков "+" – "да" и "-" – "нет":

+ + - + + + + - + - + - - - + + - + - - - + - + + + - - - - + + + - + + - + - - + + +

Подсчитаем количество тех и других знаков: "+" встречается 29 раз, "-" – 21 раз. То есть ответов "да" на 38 % больше, чем "нет". Но подождем делать соответствующие практические выводы. Подсчитаем баланс знаков, стоящих на четных местах: "+" – 18, "-" – 7; и на нечетных: "+" – 11, "-" – 14. Насколько можно доверять общим итогам опроса в таком случае?

В каждой практической задаче можно выдвинуть множество гипотез о происхождении таких разногласий. Например, в рассматриваемом примере можно предположить и психологические особенности разных групп респондентов, и недоброкачественный (непредставительный) отбор респондентов, и недобросовестность сборщиков информации и т.д. Можно ли оценить степень их влияния на итоговый результат? И как отобрать "хорошую" информацию?

Математическая статистика может дать ответы на все эти вопросы! Конечно, она не сформулирует причины обнаруженного разногласия и не даст непосредственного рецепта для отбора информации. Но она может оценить вероятность правильности выдвигаемых гипотез, определить число факторов, которые оказали решающее влияние на результаты, и оценить вклад каждого из них. С помощью математической статистики можно оценить и добротность самого статистического материала.

Отбор информации – важнейшая операция, от правильного проведения которой существенно зависит результат статистического анализа, а следовательно, и выводы исследований. Не стоит заблуждаться относительно якобы объективности данных наблюдения или эксперимента. Во-первых, результаты таких наблюдений всегда имеют свойства случайной выборки из генеральной совокупности. Так обстоит дело и при контроле технологических процессов, когда проверяются не все характеристики не всех изделий; и при изучении природных явлений, когда не все факторы учитываются и контролируются; и при экспертизе, в которой принимает участие ограниченный круг экспертов. Во-вторых, информация собирается для определенных целей. Для проверки различных моделей необходима различная информация,

подчас одна в другую непреобразуемая. В-третьих, результаты наблюдений фиксируются всегда с какой-то погрешностью: из-за методики измерения, измерительной аппаратуры, округлений и т.п. Опыт человечества, давно заметившего эти обстоятельства, привел к простейшему приему отбора информации: чем ее больше, тем лучше. В теории измерений это нашло свое выражение в методике многократного повторения опытов в идентичных условиях.

Однако не всегда есть возможность повторять опыты: это может быть слишком дорогим удовольствием или в принципе невозможно. В связи с этим приходится подробнее рассматривать различные виды отбора информации в эксперименте – получения выборки.

Приведем пример фиксации информации, соответствующий нарушению работоспособности (функционирования):

- общую наработку и наработку от предыдущего отказа;
- внешние признаки и степень влияния отказа на работоспособность объекта или системы;
- условия среды в момент отказа;
- причину отказа (предполагаемую причину);
- «адресную» информацию;
- данные об оперативности переключения резерва;
- способ и время устранения отказа [12].

Отбор информации, происходящий помимо воли исследователя, называется *естественным* в противоположность *искусственному*. Здесь не следует путать ситуацию с *пассивным* и *активным экспериментом* – естественный отбор предполагает получение информации в виде констатации определенных событий, процесс которой (констатации) не зависит от исследователя. Далее рассматриваются различные виды *искусственного отбора*.

Пристрастный отбор осуществляется по заранее намеченному признаку. Наука до XX века пользовалась именно пристрастным отбором информации: для выявления какой-либо зависимости в изучаемом природном явлении необходимо было избавиться от влияния "посторонних" факторов (например, от притяжения Земли, от проходящего трамвая и т.п.). Поэтому каждый отдельный опыт в эксперименте ставился в одних и тех же специальных условиях, имевших немаловажное значение и получивших наименование "чистоты эксперимента".

Случайный отбор производится с помощью случайных чисел по любой методике.

Механический отбор – отбор данных из всей совокупности по какому-либо правилу (например, каждый пятый).

Типический отбор – отбор из слов (частей) всей имеющейся совокупности

сти. Так делается отбор материала из отдельных партий продукции для технического контроля.

Пропорциональный отбор – частный случай *типического* отбора, когда из каждого слоя отбирается часть, пропорциональная объему слоя.

При *репрезентативном отборе* получается представительная выборка, достаточно полно характеризующая всю совокупность с точки зрения влияния важных и существенных факторов. Безусловно, к такому отбору следует стремиться, однако для оценки степени репрезентативности необходимы именно те характеристики, которые являются результатом конечного анализа отобранной информации. Поэтому в таких областях, как, например, политическая социология или экология, где цена принимаемого решения чрезвычайно высока, проводится специальный статистический эксперимент для оценки репрезентативности различных выборок и построения оптимальной из них по определенному критерию.

Для получения наиболее представительных оценок необходимо объединять данные целого ряда различных объектов. Допустимость такого объединения основывается на проверке гипотезы однородности статистического материала. Для оценок доверительных интервалов и проверки гипотез широко используется распределение χ^2 .

С помощью математической статистики определяется также необходимый объем испытаний для оценки надёжности с заданной точностью и проверяется влияние различных факторов (*дисперсионный анализ*) и взаимное влияние случайных параметров друг на друга (*корреляционный анализ*). Наконец, математическая статистика позволяет планировать испытания и исследования так, чтобы получить максимальную информацию при минимальных временных, материальных, организационных и других затратах (*теория экстремальных экспериментов*) [13].

Порядок обработки информации. С точки зрения объёма априорных сведений, *задачи статистической обработки* сводятся к двум вариантам:

1 Вид функции распределения наблюдаемой случайной величины известен априори. В этом случае задача статистической обработки – получение оценок показателей надёжности с учётом вида функции распределения и характера имеющегося статистического материала.

2 Вид функции распределения наблюдаемой случайной величины неизвестен или известен лишь предположительно. В этом случае принимается некоторая гипотеза о виде функции распределения и оцениваются ее параметры. Затем проводится проверка, не противоречат ли данные наблюдений принятой гипотезе. При положительных результатах этой проверки приступают к заключительному этапу обработки статистических данных наблюдений.

В такой постановке процесс статистической обработки более трудоёмок и состоит из следующих *этапов*: построение вариационного ряда; построение гистограммы; принятие гипотезы о виде закона распределения; оценка точечных значений параметров для принятия функции распределения; проверка непротиворечивости результатов наблюдений принятой гипотезе; оценка интервальных значений параметров функций распределения и показателей надёжности (при положительных результатах предыдущего этапа).

В случае отрицательного результата процедуры проверки гипотезы процесс статистической обработки повторяется, начиная с этапа принятия гипотезы; при другом предположении о виде функции распределения в некоторых случаях приходится возвращаться к этапу дальнейшего накопления данных.

Одной из особенностей информации в энергетике является то, что результаты наблюдений представляют собой, как правило, выборки, многократно усеченные, т.е. содержат кроме наработок до отказа и безотказные наработки (от начала наблюдений или от последнего восстановления до прекращения наблюдения).

Характер статистического материала определяется в основном порядком: начала и прекращения наблюдений; ремонта (замены) отказавшего элемента; контроля функционирования объекта или системы в условиях эксплуатации. В зависимости от назначения объекта и конкретных условий его эксплуатации перечисленные факторы могут реализовываться в различных вариантах. В частности, контроль функционирования может быть непрерывным, периодическим либо только перед началом и по окончании выполнения некоторого фиксированного задания (объема работы).

В результате наблюдений (испытаний) получаются лишь *три типа реализаций*:

1) полные – наработки до отказа (между отказами или до предельного состояния), или отрезки времени от начала до окончания восстановления;

2) неполные – безотказные наработки как результат прекращения наблюдений до наступления отказа (цензурирование);

3) условные – наработки на момент контроля, при котором обнаружен отказ (получаются в условиях, когда нет непрерывного контроля функционирования и момент отказа неизвестен).

Перечисленным типам реализаций присвоены следующие графические обозначения: полная – \times ; неполная – $-$; условная – \bullet . При помощи обозначений результаты наблюдений могут быть представлены графически соответствующим набором реализаций – диаграммой реализаций (рисунок 5.10).

Диаграмма реализаций наглядно и однозначно отражает специфику полученного статистического материала и позволяет выбрать адекватные его характеру правила обработки, т.е. правила получения точечных и интервальных оценок параметров функций распределения и показателей надёжности.

Вопросы организации наблюдений, инженерной и статистической обработки информации при оценках показателей надежности являются универсальными. Они одинаково применимы к объектам разных отраслей промышленности, в том числе и энергетики, составным частям и элементам оборудования, приборам и комплектующим изделиям. Основными специфическими особенностями являются: четкое определение объекта исследования; формулировка критериев отказов (критического, существенного, несущественного, полного, частичного и т.п.). При определении объекта исследования и формулировке критериев отказа иногда требуется учитывать влияние человека-оператора [8].

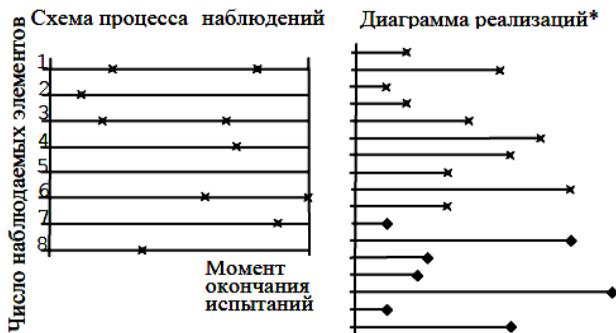


Рисунок 5.10 – Графическая иллюстрация результатов наблюдений

5.2.2 Требования к точечным оценкам и объему выборок

В процессе любого эксперимента, в том числе и вычислительного, приходится иметь дело с теми или иными значениями наблюдаемых параметров. Определение этих значений с достаточной точностью невозможно без многократных повторений опытов и специальной процедуры их обработки. Простейшим примером этого является определение средней величины результатов однотипных измерений. В более сложных случаях приходится вычислять значения *ненаблюдаемых* параметров по значениям наблюдаемых. Определение значения некоторого параметра наблюдаемого объекта по экспериментальным данным носит название **статистической точечной оценки**.

В первую очередь для расчетов и анализа любой случайной величины ξ необходимо получить точечные оценки *параметров закона* ее распределения, основными из которых являются математическое ожидание M и дисперсия σ^2 . Поэтому изучение статистических методов точечных оценок сосредоточено на получении оценок именно этих величин.

Наиболее простым методом нахождения точечных оценок является метод моментов, предложенный К. Пирсоном. Он заключается в приравнивании начальных или центральных моментов генеральной совокупности соответствующим моментам выборки.

Но если в нашем распоряжении k выборок (например, полученных k исследователями), то неясно, какую из выборочных средних надо принять за

оценку математического ожидания. Более того, эти выборочные средние сами по себе случайные величины, так как и выборки, и условия экспериментов могут случайно меняться. А может быть вообще следует взять за искомую оценку какую-то другую величину, например медиану или моду, или среднее от средних? Тогда возникает вопрос, как это надо считать, ведь объем информации в разных выборках различен!

Способ определения *точечной оценки* λ^* истинного значения параметра λ определяется теми ее свойствами, которые необходимо обеспечить в конкретном случае. Математическая статистика рассматривает следующие *свойства точечных оценок*.

Несмещенность – свойство точечной оценки λ^* , при котором ее математическое ожидание равно истинному значению λ оцениваемого параметра. То есть несмещенная точечная оценка определяет искомый параметр *без систематической ошибки*: отклонения, например, от генеральной средней (математического ожидания) распределены симметрично.

Состоятельность – свойство точечной оценки λ^* , при котором с возрастанием объема выборки N она стремится по вероятности к истинному значению λ оцениваемого параметра:

$$\lim_{N \rightarrow \infty} P(|\lambda^* - \lambda| < \varepsilon) = 1,$$

где P – вероятность, ε – произвольное сколь угодно малое число.

В специальном курсе математической статистики доказывается, что выборочная средняя λ^* является состоятельной оценкой математического ожидания.

Эффективной называется точечная оценка λ^* , имеющая при заданном объеме выборки N наименьшую дисперсию. Естественно, что из нескольких *состоятельных* оценок следует выбирать наиболее эффективную. Препятствием выявления этого свойства может стать необходимость знания закона распределения изучаемой случайной величины. Поэтому далеко не всегда удается обеспечить эффективность найденной точечной оценки. Выборочное среднее является эффективной оценкой математического ожидания.

Необходимость получения состоятельных, эффективных и достаточных оценок в случаях *неизвестных* законов распределения привела к разработке приемов получения *робастных* оценок и критериев, не зависящих от вида закона распределения или, по крайней мере, устойчивых к его нарушениям. Однако обольщаться этими приемами не следует, так как они в основном имеют эмпирическое происхождение, связаны с удачными решениями определенных задач, а в общем случае теоретически *не обоснованы*, т.е. не гарантируют получения результата.

Перечисленные выше требования к оценкам наталкивают на вопрос о приемлемом для тех или иных оценок объеме информации. Ограничения снизу на объем статистического материала для уверенных результатов: $N \geq 30$. Это ограничение касается не столько обеспечения состоятельности и

эффективности оценки, сколько достаточности степеней свободы выборки. Число степеней свободы для системы n случайных величин называется число n этих величин минус число линейных связей между ними. Поэтому при определении выборочного среднего – первой числовой характеристики (связи) – в качестве числа степеней свободы используется объем выборки N . При вычислении второй характеристики – дисперсии – число степеней свободы необходимо уменьшить на 1 за счет использования связи в виде выборочного среднего.

В таблице 5.4 приведены формулы для вычисления требуемых объемов выборок из слоев, отобранных различными способами, и выборочного среднего всей совокупности с его дисперсией.

Таблица 5.4 – Формулы для вычисления требуемых объемов выборок

| Выборка | Объем | Математическое ожидание | Дисперсия |
|--|--|---|---|
| Генеральная совокупность | | $\sum_{j=1}^k \mu_j \alpha_j$ | $\sum_{j=1}^k \mu_j \sigma_j^2 + \sum_{j=1}^k \mu_j (\alpha_j - \sum_{i=1}^k \mu_i \alpha_i)^2$ |
| Расслоенный отбор | Требуемый N_i | Выборочное среднее | Дисперсия выборочного среднего |
| Произвольная | | $\sum_{j=1}^k \mu_j \bar{x}_j$ | $\sum_{j=1}^k \mu_j^2 \frac{\sigma_j^2}{N_j} (1 - \alpha_j)$ |
| Пропорциональная | $\mu_j N$ | $\sum_{j=1}^k \mu_j \bar{x}_j$ | $\frac{1 - \alpha}{N} \sum_{j=1}^k \mu_j \sigma_j^2$ |
| Оптимальная (минимальная дисперсия D) | $\frac{\mu_i \sigma_i}{\sum_{j=1}^k \mu_j \sigma_j} N$ | $\sum_{j=1}^k \mu_j \sigma_j \sum_{j=1}^k \frac{\mu_j \bar{x}_j}{\sigma_j}$ | $\frac{\left(\sum_{j=1}^k \mu_j \sigma_j \right)^2}{N} - \frac{\sum_{j=1}^k \mu_j \sigma_j^2}{M}$ |
| Оптимальная (минимальная стоимость C) | $\frac{\mu_i \sigma_i}{\sum_{j=1}^k \frac{\mu_j \sigma_j}{\sqrt{c_j}}} \frac{N}{\sqrt{c_i}}$ | $\sum_{j=1}^k \frac{\mu_j \sigma_j}{\sqrt{c_j}} \sum_{j=1}^k \frac{\mu_j \bar{x}_j}{\sigma_j} \sqrt{c_j}$ | $\frac{\left(\sum_{j=1}^k \frac{\mu_j \sigma_j}{\sqrt{c_j}} \right)^2}{N} \sum_{j=1}^k \mu_j c_j (1 - \alpha_j)$ |

В таблице использованы следующие обозначения: μ_j – центральный момент j -го слоя; \bar{x} – выборочное среднее; k – количество выборок; N – объем выборки; $\alpha = \frac{N}{M}$, $\mu_i = \frac{M_j}{M}$, $\alpha_j = \frac{N_j}{M}$. Стоимость эксперимента

$$C = C_0 + \sum_{j=1}^k c_j N_j,$$

где C_0 – накладные расходы, c_j – стоимость одного наблюдения в j -м слое.

5.2.3 Проверка статистических гипотез

В многообразной практической деятельности человека процедура выдвигаения и проверки гипотез имеет самые различные формы – от простейших бытовых обсуждений (насколько разбавлена сметана недобросовестным продавцом) до расчетов стоимости эффекта и потерь (какова средняя стоимость ремонта аварийного автомобиля – для страховой компании). Наиболее обоснованной из этих форм является математическая теория проверки статистических гипотез.

В математической статистике существует жесткое правило: любое предположение о свойствах распределения некоторой величины, основанное на *выборочных* данных, должно быть проверено. Статистическая проверка не может доказать истинность (чего не может сделать вообще никто), но может указать с некоторой долей уверенности на *наличие или отсутствие признаков опровержения* данного суждения. Дальше уже дело человека – принять или отвергнуть предлагаемую гипотезу на основании такого статистического вывода.

Проверка однородности статистического материала. Для более точного определения показателей надёжности и установления видов законов распределения наработки на отказ и других случайных величин необходимо объединять статистические данные, собранные на различных объектах промышленных предприятий и энергосистем. В связи с этим возникает задача проверки однородности статистического материала.

Помимо выборки x_1, x_2, \dots, x_n имеются также взаимно независимые величины x_1', x_2', \dots, x_m' , распределённые одинаково и непрерывно, но принадлежащие другой выборке. Объединим эти совокупности, расположив их в порядке возрастания значений $x_1'', x_2'', \dots, x_{n+m}$. Обозначим функцию эмпирического распределения как $G_m(x)$, соответствующую выборке x_1', x_2', \dots, x_m' .

Основная гипотеза H_0 , подлежащая проверке, заключается в предположении того, что обе выборки извлечены из одной и той же совокупности, а значения функции распределения величин x и x' одинаковы.

Для проверки нулевой гипотезы используется критерий Вилкоксона, основанный на числе инверсий, под которыми понимается следующее: если какому-либо значению x предшествует некоторый x' , то говорят, что эта пара дает инверсию.

Гипотеза H_0 отвергается, если сумма инверсий u превосходит выбранную в соответствии с уровнем значимости границу, определяемую из расчёта, что при объёмах $n > 10$ и $m > 10$ выборочное число инверсий распределено нормально с центром

$$M(u) = 0,5mn$$

и дисперсией

$$D(u) = (1/12)mn(m + n + 1).$$

Пример. Собраны статистические сведения о повреждаемости контактной сети городских маршрутных предприятий электротранспорта двух городов (таблица 5.5).

Таблица 5.5 – Исходные данные двух энергосистем

| Обозначение данных энергосистем | Месяц | | | | | | | | | | | |
|---------------------------------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | я | ф | м | а | м | и | и | а | с | о | н | д |
| x | 0,8 | 1,9 | 3 | 3,5 | 3,8 | 2,5 | 1,7 | 0,9 | 1 | 2,3 | 3,3 | 3,4 |
| x' | 1,4 | 2,1 | 3,1 | 3,6 | 2,7 | 1,8 | 1,1 | 0,2 | 1,6 | 2,8 | 4 | 4,7 |

Требуется определить, можно ли считать, что между данными о повреждаемости контактной сети в разных городах нет систематических расхождений и что они имеют одинаковые систематические погрешности, т. е. нужно проверить нулевую гипотезу H_0 .

Решение. Располагаем данные в общую последовательность в порядке возрастания повреждаемости (таблица 5.6).

Таблица 5.6 – Обобщенный вариационный ряд

| | | | | | | | | | | | |
|------|------|------|-----|------|------|------|-----|------|-----|------|------|
| x' | x | x | x | x' | x' | x' | x | x' | x | x' | x' |
| 0,2 | 0,8 | 0,9 | 1 | 1,1 | 1,4 | 1,6 | 1,7 | 1,8 | 1,9 | 2,1 | 2,3 |
| x | x' | x' | x | x' | x | x | x | x' | x | x' | x' |
| 2,5 | 2,7 | 2,8 | 3 | 3,1 | 3,3 | 3,4 | 3,5 | 3,6 | 3,8 | 4 | 4,7 |

Число инверсий для x

$$u = 1 + 1 + 1 + 4 + 5 + 6 + 6 + 8 + 9 + 9 + 9 + 10 = 69.$$

Находим математическое ожидание и дисперсию:

$$M(u) = 0,5 \cdot 12 \cdot 12 = 72; \quad D(u) = (1/12) \cdot 12 \cdot 12 (12 + 12 + 1) = 300;$$

$$\sigma = \sqrt{D(u)} = 17,3.$$

Задавшись уровнем значимости 5 % и учитывая, что оба энергохозяйства городских транспортных предприятий являются равноправными, строим критическую область больших по абсолютной величине отклонений, используя значение $t_{0,05} = 1,96$.

Критическая область для гипотезы H_0

$$u \leq 72 - 1,96 \cdot 17,3 \approx 38; \quad u \geq 72 + 1,96 \cdot 17,3 \approx 106.$$

Полученное значение инверсии $u = 69$ не лежит в критической области, поэтому гипотеза H_0 не опровергается и нет оснований считать контактные сети двух городских маршрутных предприятий электротранспорта существенно различающимися между собой.

Проверка гипотез о законе распределения случайной величины. В качестве выдвигаемых *гипотез* обычно выступают предположения о свойствах закона распределения генеральной совокупности. Таковыми могут быть предположения о значениях параметров λ закона распределения – для их проверки применяются *параметрические критерии*, а также предположения о непараметризуемых свойствах распределения – для них служат *непараметрические критерии*. Первые требуют знания общего вида закона

распределения, зато вторые оказываются менее эффективными. В качестве критерия соответствия понимается достижение определенного значения *функции правдоподобия* полученной выборки. То есть если выборка попадает в область малого правдоподобия, то присутствуют признаки опровержения гипотезы – *есть основания отвергнуть* эту гипотезу. В противоположном случае *нет оснований отвергнуть* гипотезу.

Нулевой (H_0) называют выдвинутую гипотезу. Альтернативной (H_a) называется гипотеза, конкурирующая с нулевой гипотезой в том смысле, что если отвергается нулевая гипотеза, то принимается альтернативная.

Однако, поскольку наши суждения о реальности весьма ограничены, при проверке статистических гипотез по выборочным данным всегда существует возможность принятия ложного решения. Это объясняется тем, что объем выборки конечен, и поэтому нельзя точно определить ни вид функции распределения, ни значения параметров. Возможная ошибка в наших выводах предполагает не два, а четыре исхода:

- 1) гипотеза *верна* и *не отвергается* согласно критерию (правильный вывод);
- 2) гипотеза *неверна* и *отвергается* согласно критерию (правильный вывод);
- 3) гипотеза *верна*, но *отвергается* согласно критерию (ошибка I рода);
- 4) гипотеза *неверна*, но *не отвергается* согласно критерию (ошибка II рода).

Вероятность ошибки I рода, т.е. вероятность ошибки *при условии* верности гипотезы, принято обозначать α и называть уровнем значимости критерия. Для оценки гипотезы необходимо назначать *уровень значимости* – максимальное значение вероятности, которое принимается за *практическую невозможность* получения конкретной выборки с гипотетическими свойствами. Тогда вероятность правильного неотвергания проверяемой гипотезы (1-го исхода) равна $1-\alpha$.

Вероятность ошибки II рода, т.е. вероятность ошибки при условии неверности гипотезы, может быть оценена только после конкретизации *альтернативной (конкурирующей) гипотезы*, поскольку в этом случае решающее значение имеет "отдаленность" исходной гипотезы от ее альтернативы. Эту вероятность принято обозначать β . При этом вероятность правильного отвергания проверяемой гипотезы (2-й исход) равна $1-\beta$, эта величина называется мощностью критерия.

Будем рассматривать только один вид статистических критериев – статистические критерии значимости. Это значит, что будет заранее фиксироваться вероятность совершения ошибки первого рода, и тогда нет необходимости в нахождении ошибки второго рода β . Например, если при отклонении нулевой гипотезы H_0 на уровне значимости $\alpha = 0,05$ мы совершаем ошибку первого рода (считаем правильную нулевую гипотезу H_0 ложной), то в среднем ошибаемся в 5 из 100 случаев применения данного статистического критерия значимости.

Критерий χ^2 Пирсона. Если изучается непрерывная случайная величина, то вид гистограммы относительных частот обычно значительно облегчает задачу выдвижения гипотезы H_0 . Например, по виду гистограммы, изображённой на рисунке 5.11, *а*, логично выдвинуть гипотезу о равномерном законе распределения исследуемой случайной величины. Вид гистограмм, приведённых на рисунках 5.11, *б* и 5.11, *в*, напоминает соответственно кривые экспоненциального и нормального законов распределения. Большое внимание при выдвижении гипотез уделяется и вычисленным значениям числовых характеристик.

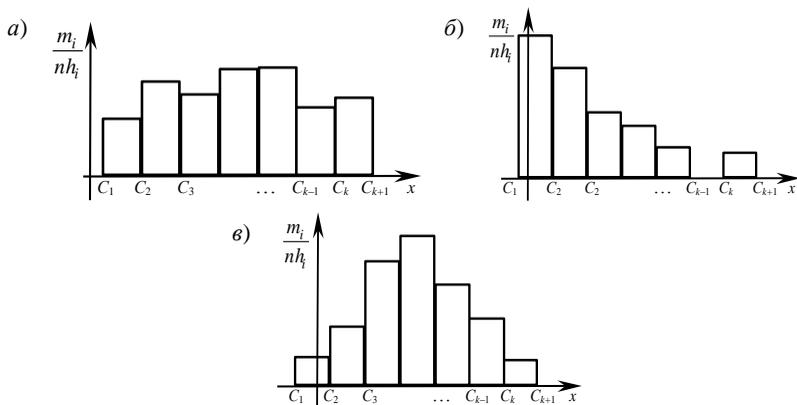


Рисунок 5.11– Примеры гистограмм относительных частот

Проверка гипотезы о предполагаемом распределении производится с помощью непараметрических критериев значимости. Одной из групп таких критериев значимости являются критерии согласия, с помощью которых проверяются нулевые гипотезы о виде функции распределения случайной величины. Одним из наиболее широко используемых на практике критериев согласия является критерий χ^2 Пирсона. Он может использоваться для проверки гипотез о виде закона распределения как дискретных, так и непрерывных случайных величин [14].

Применение критерия χ^2 основано на сопоставлении эмпирических m_i и теоретических np_i (вычисленных в предположении справедливости проверяемой гипотезы) частот попадания значений исследуемой случайной величины в рассматриваемые частичные разряды. В качестве меры расхождения эмпирического и теоретического распределений используется статистика

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i},$$

которая при $n \rightarrow \infty$ независимо от вида предполагаемого распределения

стремится к распределению χ^2 с $\nu = k - r - 1$ степенями свободы (здесь k – число разрядов разбиения, r – число параметров теоретического распределения, оцениваемых по выборке).

Легко заметить, что при незначительных отклонениях значений m_i от np_i значение критерия χ^2 будет близким к нулю. И наоборот, большое значение критерия χ^2 свидетельствует о существенном отклонении значений m_i и np_i .

Необходимыми условиями применения критерия χ^2 являются достаточно большой объем выборки ($n \geq 30$) и отсутствие в теоретическом распределении разрядов с небольшим ($np_i < 5$) числом наблюдений. Для обеспечения последнего условия интервалы, для которых $np_i < 5$, необходимо объединить с соседними. На хвостах распределения допускается значение $np_i \geq 1$.

Алгоритм применения критерия χ^2 для проверки гипотезы о виде закона распределения исследуемой случайной величины:

- 1 Строится статистический закон распределения случайной величины.
- 2 Выбирается уровень значимости α .
- 3 Формулируется гипотеза о виде закона распределения исследуемой случайной величины.

4 Вычисляются теоретические вероятности p_i попадания значений случайной величины X в рассматриваемые разряды разбиения:

$p_i = P(C_i \leq X < C_{i+1}) = F(C_{i+1}) - F(C_i)$, ($i = 1, 2, \dots, k$), где $F(x)$ – гипотетическая функция распределения случайной величины X .

5 Определяются значения теоретических частот np_i ($i = 1, 2, \dots, k$). Для обеспечения условия $np_i \geq 5$ объединяются соседние разряды разбиения.

6 Вычисляется выборочное значение критерия χ^2 .

7 По таблице критических точек распределения χ^2 определяется критическое значение $\chi_{\alpha, \nu}^2$, соответствующее заданному уровню значимости α и числу степеней свободы $\nu = k - r - 1$.

Если расчётное значение критерия попадает в критическую область, т. е. $\hat{\chi}^2 > \chi_{\alpha, \nu}^2$, то проверяемая гипотеза отвергается (при этом вероятность отклонения верной нулевой гипотезы равна α).

В случаях, когда наблюдаемое значение $\hat{\chi}^2$ не превышает критического $\chi_{\alpha, \nu}^2$, считают, что выдвинутая гипотеза не противоречит опытным данным. Подчеркнем, что полученный результат свидетельствует лишь о приемлемом согласовании проверяемой гипотезы с имеющимися выборочными данными и в общем случае не является доказательством истинности этой гипотезы.

Пример. На основании приведенных опытных данных подобрать закон распределения непрерывной случайной величины X , характеризующей продолжительность работы 50 светодиодных ламп (в тысячах часов):

6,421; 5,034; 0,599; 10,687; 26,294; 7,852; 14,040; 8,933; 4,062; 1,573;
 5,455; 2,810; 15,658; 3,692; 1,825; 17,760; 8,030; 3,218; 2,872; 8,247;
 0,417; 1,995; 0,611; 12,059; 0,665; 21,434; 22,102; 10,709; 2,283; 5,649;
 6,773; 30,034; 3,702; 12,834; 2,723; 5,255; 12,595; 3,533; 34,540; 9,238;
 1,673; 24,919; 15,511; 7,154; 1,816; 8,401; 22,239; 2,902; 3,083; 7,176.

Уровень значимости α принять равным 0,05.

Решение. Используя результаты первичной обработки выборочных данных [вид полученной гистограммы (рисунок 5.12) и значения оценок числовых характеристик $\hat{M}[X] = 8,98$, $\hat{\sigma}[X] = 8,34$], а также учитывая сведения о физическом смысле полученных значений, выдвигаем гипотезу о том, что случайная величина X распределена по экспоненциальному закону.

Вычислим оценку параметра экспоненциального закона распределения:

$$\hat{\lambda} = \frac{1}{\hat{M}[X]} = \frac{1}{8,98} \approx 0,1113.$$

При вычислении вероятностей $p_i = P(C_i \leq X < C_{i+1})$ изменим границы первого и последнего интервалов разбиения (таблица 5.7).

Таблица 5.7 – Результаты расчета теоретических вероятностей

| C_i | $Z_i = e^{-0,1113C_i}$ | $P_i = Z_i - Z_{i+1}$ |
|------------------|------------------------|-----------------------|
| 0 | 1 | 0,2828 |
| 2,987 | 0,7172 | 0,3124 |
| 8,127 | 0,4047 | 0,1763 |
| 13,267 | 0,2284 | 0,0995 |
| 18,407 | 0,1289 | 0,0562 |
| 23,547 | 0,0727 | 0,0317 |
| 28,687 | 0,0410 | 0,0179 |
| 33,827 | 0,0232 | 0,0232 |
| ∞ | 0 | — |
| $\sum_i P_i = 1$ | | |

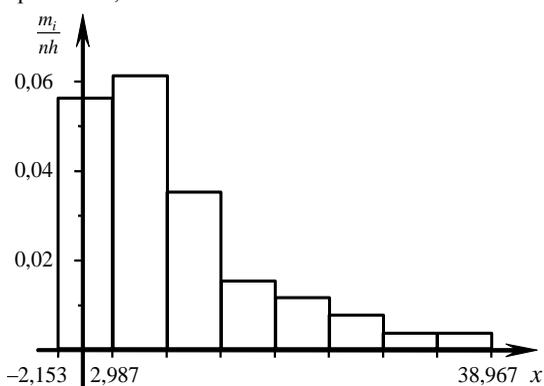


Рисунок 5.12 – Гистограмма относительных частот

Определим значения теоретических частот и занесём их в таблицу 5.8.

Таблица 5.8 – Результаты расчета значений теоретических частот

| $[C_i; C_{i+1})$ | [0; 2,987) | [2,987; 8,127) | [8,127; 13,267) | [13,267; 18,407) | [18,407; 23,547) | [23,547; 28,687) | [28,687; 33,827) | [33,827; ∞) |
|------------------|------------|----------------|-----------------|------------------|------------------|------------------|------------------|-------------|
| m_i | 14 | 16 | 9 | 4 | 3 | 2 | 1 | 1 |
| p_i | 0,2828 | 0,3124 | 0,1763 | 0,0995 | 0,0562 | 0,0317 | 0,0179 | 0,0232 |
| np_i | 14,140 | 15,620 | 8,815 | 4,975 | 2,810 | 1,585 | 0,895 | 1,160 |

Поскольку значения np_i , соответствующие четырем последним интервалам разбиения, не превышают пяти единиц, объединим эти интервалы в один.

$$\chi^2 = \frac{(14 - 14,140)^2}{14,140} + \frac{(16 - 15,620)^2}{15,620} + \frac{(9 - 8,815)^2}{8,815} + \frac{(4 - 4,975)^2}{4,975} + \frac{(7 - 6,450)^2}{6,450} = 0,252.$$

Критическое значение критерия, соответствующее значениям $\alpha = 0,05$ и $\nu = k - r - 1 = 5 - 1 - 1 = 3$: $\chi_{\alpha, \nu}^2 = \chi_{0,05;3}^2 = 7,82$.

Поскольку $\chi^2 < \chi_{\alpha, \nu}^2$, можно сделать вывод о том, что проверяемая гипотеза об экспоненциальном законе распределения изучаемой случайной величины X противоречит экспериментальным данным и нет основания для отклонения нулевой гипотезы.

Критерий «согласия» А. Н. Колмогорова. Он применяется для проверки гипотез $H_0: F_{\xi}(x) = F^*(x)$ о виде функции распределения только непрерывных случайных величин, когда параметры гипотетической функции распределения $F^*(x)$ точно известны [11].

В отличие от критерия χ^2 Пирсона, где сравниваются гипотетические и эмпирические частоты (np_i и m_i) попадания значений величины в интервалы статистического ряда распределения, критерий «согласия» λ Колмогорова сравнивает гипотетические и эмпирические функции распределения $F^*(x)$ и $\hat{F}(x)$ (рисунок 5.13).

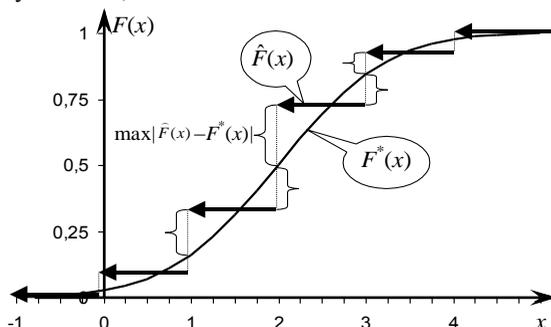


Рисунок 5.13 – Определение максимального отклонения гипотетической и эмпирической функций распределения величины ξ

Для проверки гипотезы используется статистический критерий значимости λ , равный максимальному отклонению эмпирической и гипотетической функций распределения, т.е.

$$\lambda = \max_x |\hat{F}(x) - F^*(x)|\sqrt{n},$$

где $\hat{F}(x)$, $F^*(x)$ – значения эмпирической и гипотетической функций распределения, соответственно; n – объем выборки.

В случае идеального согласования гипотезы с выборочными данными критерий λ примет значение, равное нулю. В противном случае критерий λ принимает большие положительные значения.

Если наблюдаемое значение $\hat{\lambda}$ меньше критического $\lambda_{\text{крит}}$, определяемого для заданного уровня значимости α и объема выборки n [11, 12], то гипотеза H_0 согласуется с результатами опытов, и оснований для ее отклонения нет. В противном случае (когда $\hat{\lambda} > \lambda_{\text{крит}}$) гипотеза H_0 отклоняется, и с вероятностью ошибки α можно утверждать, что верна гипотеза H_a : $F_{\xi}(x) \neq F^*(x)$.

Несмотря на вышеупомянутые ограничения, критерий λ Колмогорова, по сравнению с критерием χ^2 Пирсона, обладает рядом преимуществ. Во-первых, критерий Колмогорова построен на индивидуальных (а не сгруппированных) значениях случайной величины, поэтому результаты анализа не зависят от способа группировки данных. При этом из исследования исключается субъективный фактор. Во-вторых, критерий λ можно применять для проверки гипотезы о виде распределения случайной величины на основе меньшего объема выборки (от 20–30 элементов).

Пример. Проверим соответствие гипотезы об экспоненциальном распределении данных о поврежденных линиях электропитания. Исходные данные: t_i – время безотказной работы, m_i – количество наблюдений и необходимые расчеты сведены в таблицу 5.9.

Таблица 5.9 – Исходные данные и расчеты для проверки гипотезы об экспоненциальном распределении

| i | t_i | m_i | $q_i = m_i/n$ | $\sum q_i$ | $\frac{t_i}{T_{\text{cp}}}$ | $F = 1 - e^{-\frac{t_i}{T_{\text{cp}}}}$ | $D(t) = \sum q_i - F $ |
|-----|-------|-------|---------------|------------|-----------------------------|--|-------------------------|
| 1 | 0,091 | 1 | 0,143 | 0,143 | 0,265 | 0,233 | 0,09 |
| 2 | 0,143 | 1 | 0,143 | 0,286 | 0,385 | 0,320 | 0,034 |
| 3 | 0,167 | 1 | 0,143 | 0,429 | 0,435 | 0,353 | 0,076 |
| 4 | 0,2 | 2 | 0,286 | 0,715 | 0,495 | 0,390 | 0,325 |
| 5 | 0,25 | 1 | 0,143 | 0,858 | 0,575 | 0,437 | 0,421 |
| 6 | 1 | 1 | 0,143 | 1 | 0,965 | 0,619 | 0,381 |
| 7 | 1,1 | 1 | 0,143 | 1 | 0,97 | 0,621 | 0,379 |

Максимальному отклонению $D_n = 0,421$ при $n = 7$ соответствует
 $\lambda_n = 0,421 \cdot 7^{0,5} = 1,11$.

По таблице значений $P(\lambda)$ критерия Колмогорова при $\lambda_n = 1,11$ $P(0,75) = 0,175$ (с аппроксимацией). Эта вероятность достаточно велика, поэтому гипотеза о экспоненциальном законе распределении имеющихся выборочных данных отвергается.

5.2.4 Точность оценки. Доверительный интервал

Интервальной называют оценку, которая определяется двумя числами – концами интервала. Пусть для параметра a случайной величины получена оценка a^* . Необходимо оценить возможную ошибку; δ – точность оценки.

Надежностью (доверительной вероятностью) называют вероятность, с которой выполняется неравенство $|a - a^*| < \delta$:

$$P(|a - a^*| < \delta) = \beta.$$

Тогда $P(a^* - \delta < a < a^* + \delta) = \beta$.

Ширина доверительного интервала характеризует точность выборочной оценки, а доверительная вероятность β – достоверность оценки. Чем меньше β , тем шире доверительный интервал. Обычно β принимается в пределах 0,8–0,99 [12].

Пример. Одним из методов уменьшения случайной составляющей погрешности электрических измерений являются многократные измерения. Предполагая, что результат измерения амперметром силы тока I имеет нормальный закон распределения с математическим ожиданием, равным действительному значению силы тока I , и среднеквадратическим отклонением 3 мА (определяемым классом точности прибора), найти доверительный интервал для оцениваемого значения силы тока I , если в 10 повторных измерениях были получены следующие результаты (мА): 154,64; 158,25; 161,04; 155,93; 164,76; 152,69; 156,76; 155,03; 153,28; 150,59.

Решение. Пусть ξ – случайная величина, характеризующая результат измерения силы тока I . Поскольку среднеквадратическое отклонение $\sigma[\xi]$ измерения ξ известно, то будем использовать выражение

$$P\left(\bar{x} - \frac{\sigma[\xi]}{\sqrt{n}} U_{\alpha/2} < \Theta < \bar{x} + \frac{\sigma[\xi]}{\sqrt{n}} U_{\alpha/2}\right) = P_{\text{дов}},$$

где \bar{x} – точечная оценка математического ожидания $M[\xi]$ по выборке объема n ; $U_{\alpha/2}$ – квантиль стандартного нормального распределения уровня $\alpha/2$, $\alpha = 1 - P_{\text{дов}} = 0,05$.

$$\text{Математическое ожидание } \hat{M}[\xi] = \frac{1}{n} \sum_{i=1}^n x_i = \frac{(154,64 + 158,25 + \dots + 150,59)}{10} = 156,3.$$

Квантиль стандартного нормального распределения: $U_{0,05/2} = U_{0,025} = 1,96$. Таким образом, для доверительной вероятности 0,95 интервальная оценка математического ожидания $M[\xi]$ результата измерения ξ определяется неравенством

$$P(154,44 < M[\xi] < 158,16) = 0,95.$$

То есть с доверительной вероятностью $P_{\text{дов}} = 0,95$ действительное значение силы тока I будет в интервале $154,44 \div 158,16$ мА.

6 ОСНОВЫ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА

6.1. Классификация задач многомерного статистического анализа

На практике часто встречаются объекты с двумя и более случайными величинами, и при разработке моделей сложных систем необходимость многомерного статистического анализа становится очевидной. В таких моделях приходится изучать не только характерные особенности отдельных случайных *факторов*, но и их взаимодействие. При этом возможны различные подходы к выявлению и оценке такого взаимодействия.

Объект исследования можно представить в виде "черного ящика" с "входами" и "выходами" (рисунок 6.1), среди которых различают:

$X = (x_1, x_2, \dots, x_k)$ – вектор *входных контролируемых* переменных, которыми *можно управлять* в исследовании;

$Z = (z_1, z_2, \dots, z_p)$ – вектор *входных контролируемых* переменных, которыми *невозможно управлять* в исследовании;

$E = (e_1, e_2, \dots, e_f)$ – вектор *входных неконтролируемых и неуправляемых* переменных (*шум*);

$Y = (y_1, y_2, \dots, y_g)$ – вектор *выходных* переменных.

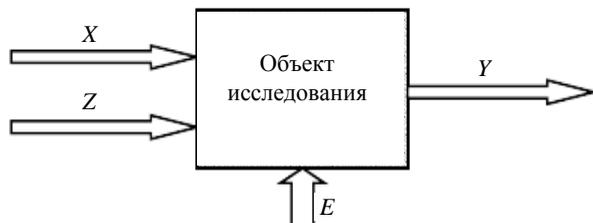


Рисунок 6.1 – Объект исследования в виде "черного ящика"

Переменные X , Z и Y называются факторами. Если фактор принимает фиксированные, детерминированные значения, то они называются *уровнями фактора*. Факторы Z могут быть случайными. Переменные E тоже случайные, хотя могут и не описываться законами распределения (могут быть не стохастического, не вероятностного вида), но происхождение их не является предметом исследования, они происходят из-за погрешностей эксперимента или моделирования. Если факторы Z случайны, то и выходные переменные Y необходимо рассматривать как случайные. В частности, такого рода случайности могут происходить из-за наложения шума E .

Рассмотрим некоторые виды задач статистического анализа, которые могут возникнуть при изучении сложных *многофакторных систем*:

- 1 Существует ли связь между отдельными факторами.
- 2 Если между факторами есть связь, то насколько она тесная.

3 Если между какими-то факторами есть связь, то какой функцией ее можно представить.

4 Какие входные факторы оказывают на определенные выходные наибольшее влияние.

5 Какие входные факторы можно отбросить из процесса изучения на основании их слабого, сравнимого с шумом, влияния.

6 Существуют ли неучтенные факторы, которые необходимо рассматривать ввиду их существенного влияния на выходные.

7 Существуют ли обобщенные факторы, которыми можно заменить несколько рассматриваемых.

8 Как связаны между собой зашумленные факторы.

9 Каковы характеристики шума.

10 Как выделить "полезную" информацию из зашумленной.

Все эти задачи можно решить с помощью методов многомерного статистического анализа, включающего в себя:

- корреляционный анализ (КА);
- регрессионный анализ (РА);
- конфлюэнтный анализ;
- дисперсионный анализ (ДА),
- факторный анализ (ФА);
- теорию фильтрации.

Каждый из этих разделов – совокупность методов и приемов математической статистики. Исторически так сложилось, что четкого разграничения указанных разделов и методов между разделами не существует. Выбор методов диктуется лишь конкретной практической задачей. Так, например, задачи 1, 2 вышеприведенного списка решаются методами КА, задача типа 3 – РА. Задачи 4–6 относятся к задачам ДА, 7 – ФА, а 8, 9 – конфлюэнтного. Теория фильтрации позволяет решить задачи типа 10 и частично 9.

Каждый из перечисленных видов статистического анализа, по сути, предназначен для обоснования той или иной *статистической математической модели* изучаемого оригинала. Корреляционный анализ позволяет получить корреляционную модель, регрессионный анализ – регрессионную модель, конфлюэнтный – конфлюэнтную, дисперсионный – дисперсионную. Возможность решения перечисленных задач вытекает из физического смысла основных числовых характеристик случайных величин: дисперсии или среднего квадратического отклонения и коэффициента корреляции.

Конечной целью и основной *задачей корреляционного анализа* и является построение совместного закона распределения системы случайных величин.

Задачу регрессионного анализа составляет исследование вида и формы корреляции: какой функцией можно представить связь случайных величин (*регрессию* – уравнение для условной средней).

Некоторые авторы считают регрессионный анализ частью корреляцион-

ного, некоторые считают его методы самостоятельными. Однако эти методы могут предложить лишь то или иное выражение связи, они не отвечают на вопрос о наличии собственно функциональной зависимости. Такая постановка вопроса возможна только в профессиональном плане исследования объекта, а не в формальном математическом.

Задачей конфлюэнтного анализа является изучение структуры случайных величин, находящихся в некотором взаимодействии [2]. Наибольшее развитие у классика многомерного статистического анализа Р. Фишера получил *дисперсионный анализ*, основной задачей которого является сравнение дисперсий разных случайных величин или различных способов вычисления дисперсий.

Факторный анализ решает задачу поиска минимального числа обобщенных факторов, заменяющих исходное множество.

Задачей теории фильтрации является выделение исходного сигнала из искаженной или неполной информации [2].

6.2 Регрессионный анализ

Часто в практике инженерных и научных исследований результат испытания характеризуется набором случайных величин, т. е. многомерной случайной величиной. Например, влияние погодных условий характеризуется температурой, давлением, влажностью, направлением и скоростью ветра.

Многомерные случайные величины характеризуются многомерными законами распределения. Наряду с числовыми характеристиками, применимыми для одномерных СВ (математическим ожиданием, дисперсией и др.), многомерные случайные величины определяются дополнительными числовыми характеристиками, описывающими зависимость между ее компонентами.

В дальнейшем ограничимся рассмотрением двумерной СВ (X, Y) , компоненты X и Y которой фактически являются одномерными СВ.

Зависимость между случайными величинами, при которой каждому значению x случайной величины X однозначно ставится в соответствие единственное значение y , называется *функциональной*. Например, зная напряжение U на участке электрической цепи сопротивлением R , можно однозначно определить величину тока ($I = U/R$); зная скорость движения поезда v , можно однозначно определить время T его движения по перегону.

Однако часто на практике одному значению СВ X может соответствовать не одно, а множество значений СВ Y , характеризующихся для каждого $X = x$ условным распределением с плотностью вероятностей $f(y|X=x)$. Такая зависимость называется *статистической*. Примером статистической зависимости является зависимость величины потребления электроэнергии на перегоне электровозом (Y) от массы поезда (X), т.к. одному значению массы состава могут соответствовать различные значения потребления электроэнергии (кроме массы состава на электропотребление влияют ско-

рость движения, профиль и состояние пути, состояние подвижного состава, направление и скорость ветра и прочие факторы).

Таким образом, при фиксированном значении x_i влияющей величины X значения зависимой величины Y по-прежнему являются случайными и определяются условным законом распределения (при условии, что $X = x_i$). Изменение условного закона распределения СВ Y при изменении значения X может проявляться как при изменении вида распределения (рисунок 6.2, а), так и при изменении его числовых характеристик (рисунок 6.2, б), например, математического ожидания.

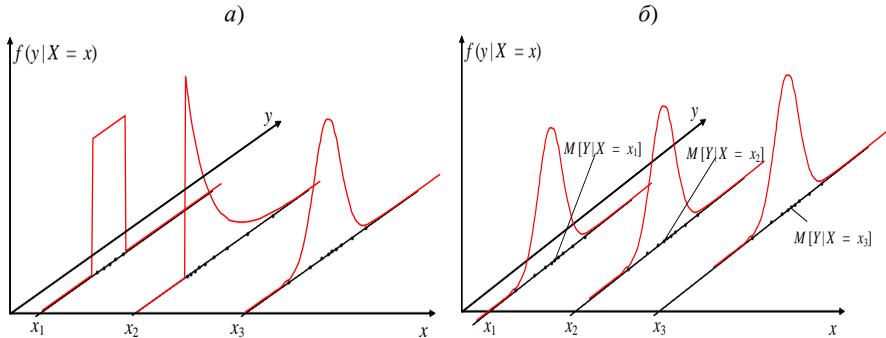


Рисунок 6.2 – Примеры статистической зависимости: с изменением вида закона распределения (а), с изменением только лишь математического ожидания (б)

Если условный закон распределения [например, $f(y|X=x)$] одной из величин не изменяется при изменении значения другой случайной величины X , то такие величины называются *независимыми*. Например, температура воздуха Y завтра в полдень не зависит от скорости движения автомобиля X . Независимыми являются случайные величины, не связанные причинно-следственными связями.

Исследование зависимостей между случайными величинами является наиболее часто используемым приложением математической статистики и применяется главным образом для предсказания значения одной СВ по значению другой. Чаще всего ограничиваются изучением *регрессионной* зависимости, которая каждому значению одной СВ ставит в соответствие условное математическое ожидание другой СВ $M[Y|X=x]$. Предметом *регрессионного* анализа является нахождение вида уравнения регрессионной зависимости между случайными величинами. Предметом же *корреляционного* анализа является оценка тесноты связи между СВ, а также оценка качества описания зависимости выбранным уравнением регрессии [14].

Регрессионный анализ предназначен для решения единственной задачи получения теоретического *уравнения регрессии*, вид которого задается, исходя из особенностей изучаемой системы случайных величин.

Регрессия – функциональная зависимость, *аппроксимирующая* (заменяющая) статистическую зависимость средних значений рассматриваемых факторов (переменных).

Рассмотрим последовательные шаги решения общей задачи регрессионного анализа.

1 *Задание вида* линии *регрессии* – важнейший этап. Основные соображения о виде функции должны диктоваться только реальными *физическими* свойствами изучаемого явления, например, возрастание/убывание, выпуклость/вогнутость, наличие замечательных точек (известные значения координат отдельных точек, экстремумов и перегибов), асимптот, периодов, свойств симметрии, гладкости и т.д. *И только дополнительными мотивами могут служить графические построения выборочной корреляционной зависимости.*

2 Вычисление *выборочных оценок* параметров предполагаемого теоретического закона распределения: \bar{x} , \bar{y} , σ_x , σ_y , r_{xy} .

3 *Проверка гипотезы* о равенстве нулю коэффициента корреляции. Этот шаг нужен для выяснения вопроса о необходимости дальнейшего анализа. Если исследуемые случайные величины по результатам выборки можно считать некоррелированными, то бесполезно пытаться строить регрессию и следует прекратить анализ или вернуться к пункту 1.

4 *Проверка гипотез* о законах распределения исследуемых случайных величин. Это необходимо для метода наибольшего правдоподобия и может быть проделано с помощью критерия согласия Пирсона.

5 Для отыскания *параметров линии регрессии* используется *метод наибольшего правдоподобия*, дающий несмещенные их оценки и не предъявляющий каких-либо требований к законам распределения случайных величин. Правда, применить этот метод можно только тогда, когда последние *известны*. Поэтому и необходим предыдущий шаг.

6 Уравнение регрессии можно считать окончательным только после подтверждения дисперсионным анализом, т. е. "погрешность" замены статистической регрессии аппроксимирующей зависимостью. Проверка производится по критерию Фишера с $(k - v, N - k)$ степенями свободы.

Пример. Пусть дана выборка значений двумерной СВ $(X, Y) = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$, где n – объем двумерной выборки. Первым шагом в построении эмпирического уравнения регрессии между СВ является графическое отображение значений двумерной СВ в виде точек $(x_1, y_1), \dots, (x_n, y_n)$ на плоскости X - Y , называемое *диаграммой рассеяния* (корреляционным полем) (рисунок 6.3).

Визуальный анализ диаграммы рассеяния и предметная постановка задачи (физический смысл рассматриваемых величин) позволяет сделать предположение о виде уравнения регрессии. Если предполагается, что зависимость между СВ X и Y линейна (рисунок 6.3, а), то теоретическая модель регрессионной зависимости между СВ задается уравнением

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = \overline{1, n},$$

где y_i – i -е выборочное значение СВ Y ;

β_0, β_1 – параметры линейной регрессии, требующий определения;

x_i – i -е выборочное значение СВ X ;

e_i – ошибка, вызванная отклонением i -го наблюдения СВ Y от условного среднего $M[Y|X=x_i]$. Ошибки e_i обуславливаются упрощением вида зависимости Y от X (без учёта всех влияющих факторов), а также возможной ошибкой в выборе формы регрессии, которая может описываться другим уравнением).

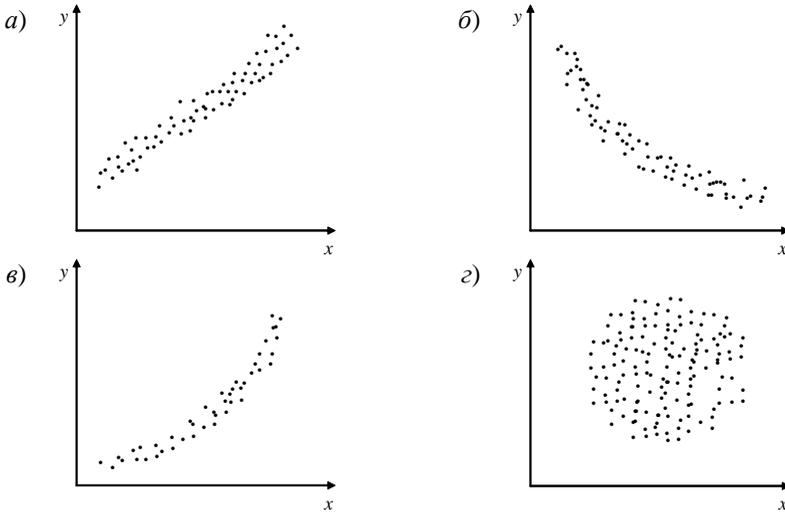


Рисунок 6.3 – Диаграмма рассеяния, соответствующая линейной (а), экспоненциальной (б), параболической (в) регрессионной зависимости и отсутствию регрессионной зависимости (г)

Метод наименьших квадратов. Для оценивания параметров уравнения регрессии β_0, β_1, \dots наиболее часто применяется метод наименьших квадратов (МНК). Идею МНК продемонстрируем на примере определения параметров линейной регрессионной зависимости.

Для случая линейной регрессии оценки параметров $\hat{\beta}_0$ и $\hat{\beta}_1$ получаются минимизацией по $\hat{\beta}_0$ и $\hat{\beta}_1$ суммы квадратов отклонений e_i выборочных значений СВ Y от значений, предсказываемых уравнением регрессии Y на X . Полученные значения $\hat{\beta}_0$ и $\hat{\beta}_1$ вычисляются по формулам

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}; \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

В результате оценка уравнения линейной регрессии (эмпирическое уравнение линейной регрессии, прямая, полученная МНК) будет иметь вид

$$\bar{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Метод наименьших квадратов применим для оценивания параметров уравнений регрессии произвольного вида (гиперболической, параболической, экспоненциальной, логарифмической и т. д.).

Пошаговый регрессионный анализ. В случае, если по виду диаграммы рассеяния сложно выдвинуть предположение о виде регрессионной зависимости, рекомендуется использовать пошаговый регрессионный анализ. Для этого уравнение регрессии выбирают как можно более сложным, содержащим большое количество слагаемых. Например,

$$\bar{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 / x + \hat{\beta}_2 x + \hat{\beta}_3 x^2 + \hat{\beta}_4 x^3 + \hat{\beta}_5 e^x.$$

Стандартная пошаговая процедура позволяет в записанном уравнении регрессии выбрать те слагаемые, совокупность которых достаточно качественно предсказывает среднее значение зависимой СВ, и отбросить те слагаемые, которые существенно не улучшают предсказание СВ Y .

Особенно полезен пошаговый регрессионный анализ в случае множественной регрессии, когда СВ Y зависит от нескольких независимых случайных величин X_1, X_2, X_3 и т. д. При этом пошаговая процедура позволяет из множества влияющих (независимых) СВ исключить несущественные и тем самым упростить уравнение регрессии (данную процедуру можно реализовать в программах статистической обработки данных, например Statistica или Statgraphics [15, 16]).

6.3 Корреляционный анализ

Помимо предположения о форме уравнения регрессионной зависимости между случайными величинами и нахождения параметров уравнения исследователю необходимо оценить, насколько удачно выбранная регрессионная модель объясняет существующую зависимость между исследуемыми СВ.

Корреляционный анализ – группа статистических методов установления формы и тесноты корреляционной связи между факторами.

Коэффициент корреляции. Основной числовой характеристикой, определяющей тесноту линейной связи между двумя случайными величинами, является коэффициент корреляции

$$r = \frac{M[XY] - M[X]M[Y]}{\sigma[X]\sigma[Y]},$$

где $M[XY]$ – математическое ожидание произведения СВ X и Y .

Коэффициент корреляции является безразмерной величиной и может принимать значения из интервала $-1 \leq r \leq 1$. Для линейно независимых слу-

чайных величин X и Y коэффициент корреляции равен нулю. Чем теснее линейная зависимость между двумя СВ, тем больше коэффициент корреляции отличается от нуля. Экстремальные значения -1 или 1 коэффициента корреляции соответствуют линейной функциональной зависимости между двумя СВ (отрицательной и положительной соответственно). Положительная зависимость указывает на увеличение значения одной величины при увеличении значения другой величины, например стаж и производительность труда рабочих, время подготовки к экзамену и полученная оценка. Для отрицательной зависимости характерно уменьшение значения одной из величин при увеличении значения другой величины.

Таким образом, можно говорить, что коэффициент корреляции характеризует тесноту линейной регрессионной зависимости между двумя СВ, т. е. близость ее к линейной функциональной зависимости. По заданной двумерной выборке оценку коэффициента корреляции \hat{r} можно найти по формуле

$$\hat{r} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Коэффициент детерминации. Для характеристики качества описания зависимости между двумя СВ произвольным уравнением регрессии используется коэффициент детерминации \hat{R}^2 . Чем лучше линия регрессии аппроксимирует точки на диаграмме рассеяния, тем больше значение \hat{R}^2 и тем надежнее уравнение регрессии может быть применено для практических расчетов.

Оценка коэффициента детерминации может принимать значения в интервале от нуля до единицы ($0 \leq \hat{R}^2 \leq 1$) и определяется выражением

$$\hat{R}^2 = \frac{\sum_{i=1}^n (\bar{y}(x_i) - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

где n – объем выборки; $\bar{y}(x_i)$ – значение СВ Y , предсказываемое уравнением регрессии; $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ – безусловное среднее СВ Y ; y_i – i -е выборочное значение СВ Y .

Отметим, что коэффициент детерминации не является мерой какой-либо зависимости априорно (как это справедливо для коэффициента корреляции, который всегда характеризует степень линейной зависимости между двумя СВ), он лишь оценивает степень приближения выбранного уравнения регрессионной зависимости к действительной зависимости между двумя

СВ. Близость коэффициента детерминации к нулю указывает на то, что выбранное уравнение регрессии (модель зависимости) незначимо объясняет существующую зависимость между СВ. Равенство же коэффициента детерминации единице указывает на то, что выбранное уравнение регрессии полностью (однозначно) определяет зависимость между СВ. Если значение коэффициента детерминации больше 0,7, то считают, что выбранное уравнение регрессии хорошо описывает зависимость, существующую между случайными величинами. Если же коэффициент детерминации меньше 0,3, то уравнение регрессии незначительно описывает зависимость между случайными величинами, если таковая существует.

Проверка значимости оценок коэффициентов корреляции и детерминации. Оценки коэффициентов корреляции и детерминации сами являются случайными величинами, так как для различных выборок из одной и той же генеральной совокупности могут принимать различные значения. При малых объемах выборок эти различия будут особенно существенными. Поэтому при нахождении оценок коэффициентов корреляции и детерминации используется проверка значимости этих оценок, которая позволяет сделать вывод о существенности описания действительной зависимости выбранным уравнением регрессии.

Проверка значимости \hat{r} . Чтобы сделать статистический вывод о значимости оценки коэффициента корреляции (при проверке линейной регрессионной зависимости) выдвигается нулевая гипотеза об отсутствии линейной зависимости между исследуемыми СВ (т. е. $H_0: r=0$; $H_A: r \neq 0$). Если гипотеза H_0 отклоняется, то считается, что уравнение регрессии Y на X действительно имеет линейный вид.

Для проверки гипотезы H_0 вычисляется t -статистика:

$$\hat{t} = \hat{r} \sqrt{\frac{n-2}{1-\hat{r}^2}}.$$

При условии справедливости гипотезы H_0 рассчитанная t -статистика имеет распределение Стьюдента с $n-2$ степенями свободы. Значение \hat{t} сравнивается с критическим значением $t_{\alpha, \nu}$ при $\nu = n-2$ степенях свободы. Если расчетное значение \hat{t} по абсолютной величине превосходит табличное для заданного уровня значимости α , то нулевая гипотеза H_0 о линейной независимости двух СВ отклоняется, и с вероятностью ошибки α можно утверждать, что существует линейная зависимость.

Проверка значимости \hat{R}^2 . При выполнении процедуры проверки значимости оценки коэффициента детерминации выдвигается нулевая гипотеза о том, что предложенное уравнение регрессии никак не отражает реальную зависимость между СВ, т. е. при H_0 $\hat{R}^2 = 0$. Альтернативная гипотеза

за заключается в том, что выбранная модель зависимости (уравнение регрессии) $\bar{y}(x_i)$ в достаточной степени объясняет действительную зависимость между случайными величинами, т. е. при $H_0 \hat{R}^2 > 0$.

Для проверки значимости оценки коэффициента детерминации используется статистика

$$\hat{F} = \hat{R}^2 \frac{n-m}{(m-1)(1-\hat{R}^2)},$$

имеющая \hat{F} -распределение Фишера с $v_1 = m - 1$ и $v_2 = n - m$ степенями свободы. Здесь m – число неизвестных параметров предполагаемого уравнения регрессии ($\beta_0, \beta_1, \dots, \beta_{m-1}$). Значение статистики, вычисленное по формуле, сравнивается с критическим значением $F_{v_1, v_2, \alpha}$, найденным по таблицам квантилей распределения Фишера при заданном уровне значимости и соответствующем числе степеней свободы. Если $\hat{F} > F_{v_1, v_2, \alpha}$, то нулевая гипотеза отклоняется, вычисленный коэффициент детерминации значимо отличается от нуля, и с вероятностью ошибки α можно утверждать, что между исследуемыми величинами существует зависимость предложенного вида, и полученное уравнение регрессии может использоваться в дальнейших исследованиях.

Пример. По результатам десяти совместных измерений скорости движения электровоза X , км/ч, и соответствующего потребления электроэнергии Y , МВт·ч/100 км, представленных в таблице 6.1, следует исследовать зависимость между данными величинами с целью прогнозирования величины потребления электроэнергии электровоза при заданной скорости движения.

Таблица 6.1– Исходные данные

| | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| X | 40,23 | 19,63 | 29,01 | 89,14 | 74,96 | 57,89 | 34,33 | 21,01 | 16,69 | 9,24 |
| Y | 50,66 | 67,82 | 60,95 | 65,53 | 46,84 | 52,91 | 43,71 | 70,52 | 67,02 | 89,96 |

Решение. На величину потребления электроэнергии электровозом Y , помимо скорости движения X , влияние оказывает масса состава, профиль и качество железнодорожного полотна, качество подвижного состава, направление и скорость ветра и другие факторы. Поэтому зависимость между величиной потребления электроэнергии электровозом Y и скоростью движения X является статистической: на одной скорости движения при различных дополнительных условиях потребление электроэнергии электровозом может принимать различные значения. Диаграмма рассеяния исследуемых величин представлена на рисунке 6.4.

Характер расположения точек на диаграмме рассеяния позволяет сделать предположение о параболической регрессионной зависимости

$$\bar{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2.$$

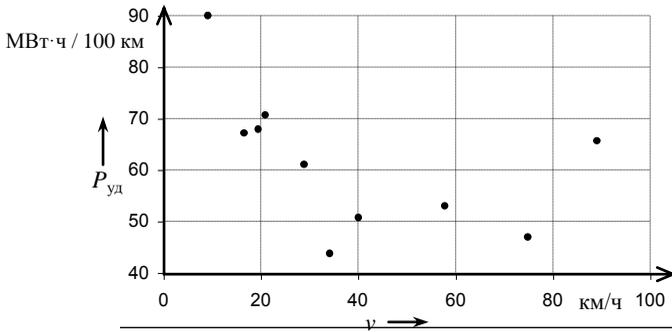


Рисунок 6.4 – Диаграмма рассеяния случайных величин X и Y

Оценки параметров β_0 , β_1 и β_2 найдем методом наименьших квадратов:

$$\begin{cases} \sum_{i=1}^n y_i - \hat{\beta}_0 n - \hat{\beta}_1 \sum_{i=1}^n x_i - \hat{\beta}_2 \sum_{i=1}^n x_i^2 = 0; \\ \sum_{i=1}^n y_i x_i - \hat{\beta}_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 - \hat{\beta}_2 \sum_{i=1}^n x_i^3 = 0; \\ \sum_{i=1}^n y_i x_i^2 - \hat{\beta}_0 \sum_{i=1}^n x_i^2 - \hat{\beta}_1 \sum_{i=1}^n x_i^3 - \hat{\beta}_2 \sum_{i=1}^n x_i^4 = 0. \end{cases}$$

Для вычисления значений сумм составим расчетную таблицу 6.2.

Таблица 6.2 – Результаты промежуточных вычислений

| N | x_i | y_i | x_i^2 | $x_i y_i$ | x_i^3 | x_i^4 | $x_i^2 y_i$ |
|----------|--------|--------|---------|-----------|-----------|-------------|-------------|
| 1 | 40,23 | 50,66 | 1618,45 | 2038,05 | 65110,36 | 2619389,79 | 81990,82 |
| 2 | 19,63 | 67,82 | 385,34 | 1331,31 | 7564,16 | 148484,53 | 26133,55 |
| 3 | 29,01 | 60,95 | 841,58 | 1768,16 | 24414,24 | 708257,06 | 51294,31 |
| 4 | 89,14 | 65,53 | 7945,94 | 5841,34 | 708301,06 | 63137956,13 | 520697,42 |
| 5 | 74,96 | 46,84 | 5619,00 | 3511,13 | 421200,36 | 31573178,98 | 263194,03 |
| 6 | 57,89 | 52,91 | 3351,25 | 3062,96 | 194003,98 | 11230890,64 | 177314,75 |
| 7 | 34,33 | 43,71 | 1178,55 | 1500,56 | 40459,58 | 1388977,51 | 51514,37 |
| 8 | 21,01 | 70,52 | 441,42 | 1481,63 | 9274,24 | 194851,70 | 31128,95 |
| 9 | 16,69 | 67,02 | 278,56 | 1118,56 | 4649,10 | 77593,50 | 18668,83 |
| 10 | 9,24 | 89,96 | 85,38 | 831,23 | 788,89 | 7289,33 | 7680,57 |
| Σ | 392,13 | 615,92 | 21745,5 | 22484,9 | 1475766 | 111086869,2 | 1229617,6 |

После подстановки значений система уравнений примет вид

$$\begin{cases} 615,92 - 10\hat{\beta}_0 - 392,13\hat{\beta}_1 - 21745,47\hat{\beta}_2 = 0; \\ 22484,93 - 392,13\hat{\beta}_0 - 21745,47\hat{\beta}_1 - 1475765,97\hat{\beta}_2 = 0; \\ 1229617,6 - 21745,5\hat{\beta}_0 - 1475765,97\hat{\beta}_1 - 111086869,2\hat{\beta}_2 = 0. \end{cases}$$

Решив систему уравнений известными методами или с помощью MATHCAD, получим следующее решение: $\hat{\beta}_0 = 103$; $\hat{\beta}_1 = -2,096$; $\hat{\beta}_2 = 0,01875$, а уравнение регрессии примет вид

$$\bar{y}(x) = 103 - 2,096x + 0,01875x^2.$$

На рисунке 6.5 представлена диаграмма рассеяния случайных величин X и Y с нанесённой линией регрессии.

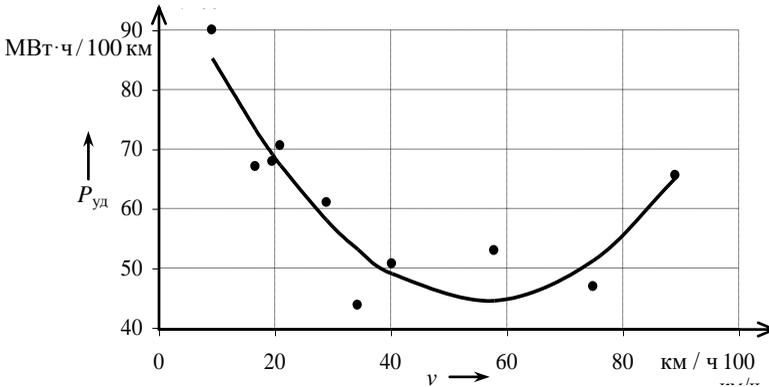


Рисунок 6.5 – Диаграмма рассеяния случайных величин X и Y с нанесённой линией регрессии

Оценим качество описания зависимости между потреблением электроэнергии электровозом (Y) и скоростью его движения (X), полученным уравнением регрессии с помощью коэффициента детерминации:

$$\hat{R}^2 = \frac{\sum_{i=1}^n (\bar{y}(x_i) - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 0,8447.$$

Расчётное значение коэффициента детерминации $\hat{R}^2 = 0,844708$ указывает на удовлетворительность описания зависимости между величиной скорости (X) и потреблением электроэнергии (Y), выбранным уравнением регрессии. Проверим, однако, значимость оценки коэффициента детерминации с помощью статистики Фишера:

$$\hat{F} = 0,8447 \cdot \frac{10 - 3}{(3 - 1)(1 - 0,8447)} = 19,04.$$

Вывод. Критическое значение статистики Фишера для степеней свободы $v_1 = 3 - 1 = 2$ и $v_2 = 10 - 3 = 7$ и уровня значимости $\alpha = 0,05$ составляет $F_{0,05;2;7} = 4,737$. Поскольку расчётное значение статистики Фишера больше критического ($\hat{F} = 19,03818 > 4,737 = F_{0,05;2;7}$), то вычисленный коэффициент детерминации значимо отличается от нуля, и выбранное уравнение регрессионной зависимости может быть использовано в практических целях.

6.4 Дисперсионный анализ

Дисперсионный анализ разработан Р. Фишером в 1920 году и представляет собой группу методов математической статистики для анализа результатов наблюдений, зависящих от нескольких одновременно действующих

факторов, как случайных, так и детерминированных, как наблюдаемых, так и ненаблюдаемых.

На современном этапе развития теории вероятностей основными причинами случайности (непредсказуемости) явлений считаются невозможность учета всех факторов, влияющих на рассматриваемое явление, и ограниченная точность физических измерений.

При прогнозировании случайных явлений исследователи стараются учитывать как можно большее количество влияющих факторов. Например, при определении температуры окружающей среды на следующий день учитывается температура за прошедшую неделю, влажность, давление, расположение атмосферных фронтов, скорость и направление ветра, температура в прилегающих областях, фаза Луны, солнечная активность и другие факторы. Однако учесть все факторы (например, солнечную активность) бывает сложно или невозможно. Поэтому ограничиваются лишь значимо влияющими факторами. Вопрос о значимости влияния (в частности, случайную величину) тех или иных факторов может быть решен **методами дисперсионного анализа**.

Основной целью дисперсионного анализа является проверка свойств распределения некоторой случайной величины, основанное на выборочных данных. Происхождение названия “дисперсионный анализ” связано с тем, что при исследовании статистической значимости различия между средними двух (или нескольких) групп на самом деле сравниваются их дисперсии.

Сущностью дисперсионного анализа является расчленение общей дисперсии на части, одна из которых обусловлена случайной ошибкой (то есть внутригрупповой изменчивостью), а вторая связана с различием средних значений. Последняя компонента дисперсии затем используется для анализа статистической значимости различия между средними значениями. Если это различие значимо, нулевая гипотеза отвергается и принимается альтернативная гипотеза о существовании различия.

Переменные, значения которых определяются с помощью измерений в ходе эксперимента (например, количество отказавших транзисторов), называются *зависимыми* переменными. Переменные, которыми можно управлять при проведении эксперимента (например, значение напряжения, тока, частоты, температуры), называются *факторами* или *независимыми* переменными (дисперсионный анализ удобно выполнять в программах статистической обработки данных, например Statistica или Statgraphics [15, 16]).

При разработке математического аппарата Фишеру потребовалось сделать следующие предположения дисперсионного анализа, необходимые для обоснованности выкладок и выводов:

1) исследуемые факторы *независимы*. С точки зрения способов отбора информации это означает независимость выборочных результатов наблюдения (отдельных выборок или слоев – они не преобразуются друг в друга с помощью какого-либо алгоритма);

2) исследуемые факторы, каждый по отдельности, подчиняются *нормальным* законам распределения;

3) *дисперсии* исследуемых факторов *однородны* (априори приблизительно одного порядка).

В зависимости от количества учитываемых факторов дисперсионный анализ может быть одно- или многофакторным. Предметом *однофакторного дисперсионного анализа* является проверка того, зависит ли математическое ожидание случайной величины ξ от некоторого фактора. Например, влияет ли форма импульсов электромагнитных помех на помехоустойчивость компьютера системы управления.

Многофакторный дисперсионный анализ позволяет проверить, оказывают ли влияние на математическое ожидание случайной величины ξ определенные факторы (которые можно варьировать в ходе эксперимента), а также выбрать наиболее важные факторы и оценить степень их влияния. Например, какие из факторов (частота импульсов помех, их длительность, амплитуда, форма, каналы проникновения и пр.) значительно ухудшают качество работы электронного оборудования, и, следовательно, какие меры должны быть предприняты для увеличения помехозащищенности и помехоустойчивости оборудования в первую очередь.

Однофакторный дисперсионный анализ. Идея однофакторного дисперсионного анализа заключается в разложении общей дисперсии (разброса) случайной величины ξ на два независимых слагаемых – факторную дисперсию (разброс, порождаемый воздействием исследуемого фактора A) и остаточную дисперсию, обусловленную другими неучтенными воздействиями, т.е. $D[\xi] = D_{\text{факт}} + D_{\text{ост}}$. Чем сильнее фактор A влияет на значение величины ξ в массе всех влияющих факторов, тем больше факторная дисперсия $D_{\text{факт}}$ по отношению к остаточной дисперсии $D_{\text{ост}}$. В результате сравнения факторной и остаточной дисперсий по критерию Фишера $F = D_{\text{факт}} / D_{\text{ост}}$ приходят к выводу о значимости влияния фактора A .

С целью изучения влияния фактора A на случайную величину ξ результаты эксперимента разбиваются в зависимости от вариации фактора A на k групп по n_i измерений в каждой. Результаты измерений обозначаются через x_{ij} , где i – номер уровня фактора A ($i = \overline{1, k}$), j – номер результата измерения ($j = \overline{1, n_i}$).

На основе имеющейся статистической информации требуется проверить нулевую гипотезу $H_0: M[\xi|A_1] = M[\xi|A_2] = \dots = M[\xi|A_k]$ о равенстве условных математических ожиданий величины ξ при различных значениях уровня фактора A против альтернативной гипотезы H_a о том, что не все условные математические ожидания совпадают (а следовательно, их различия вызваны различием уровней влияющего фактора A).

Проверка нулевой гипотезы основывается на вычислении статистического критерия значимости

$$F = \frac{\frac{1}{k-1} \sum_{i=1}^k ((x_i - \bar{x})^2 n_i)}{\frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - x_i)^2},$$

который, при условии справедливости гипотезы H_0 , имеет распределение Фишера с $v_1 = k - 1$ и $v_2 = n - k$ степенями свободы.

Если значение критерия значимости F окажется больше критического F_{α, v_1, v_2} , определяемого квантилем распределения Фишера с v_1 и v_2 степенями свободы уровня α [17], то с вероятностью ошибки, меньшей α , можно утверждать, что гипотеза $H_0: M[\xi|A_1] = M[\xi|A_2] = \dots = M[\xi|A_k]$, неверна (не все условные математические ожидания равны между собой). Следовательно, с вероятностью ошибки, меньшей α , можно утверждать, что значения исследуемого фактора A влияют на значения случайной величины ξ .

Замечание. Дисперсионный анализ основан на нескольких допущениях [18]:

1) результаты измерений x_{ij} внутри i -й группы имеют нормальный закон распределения;

2) внутригрупповые дисперсии результатов измерений x_{ij} при различных уровнях фактора A одинаковы, т.е. $D_1 = D_2 = \dots = D_k$.

Примечание – Приведенные условия выполняются в случае, когда количество неучтенных факторов велико, а их совместное действие на величину ξ компенсируется (в соответствии с центральной предельной теоремой) и не зависит от значений уровней фактора A . Выполнение второго условия можно проверить с помощью статистического критерия значимости Бартлетта [5].

Пример 1. Методом однофакторного дисперсионного анализа проверить влияние помехозащитных мероприятий на качество работы микропроцессора (МК). Испытания проходили 24 микроконтроллера, из них 10 – незащищенных, 7 – с экранированным корпусом и 7 – с заземленным корпусом. Количество сбоев каждого из МК (в процессе испытания) представлено в таблице 6.3 и на рисунке 6.6.

Таблица 6.3 – Результаты испытаний микроконтроллеров

| Помехозащитные меры | Номер образца | | | | | | | | | | Число образцов в группе |
|---------------------|---|----|----|----|----|----|----|----|----|----|-------------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| | Количество сбоев образца в процессе испытания | | | | | | | | | | |
| 1 Без защиты | 11 | 9 | 18 | 11 | 17 | 12 | 14 | 16 | 15 | 9 | 10 |
| 2 Экранирование | 9 | 11 | 15 | 8 | 9 | 6 | 10 | | | | 7 |
| 3 Заземление | 11 | 5 | 8 | 9 | 7 | 4 | 10 | | | | 7 |

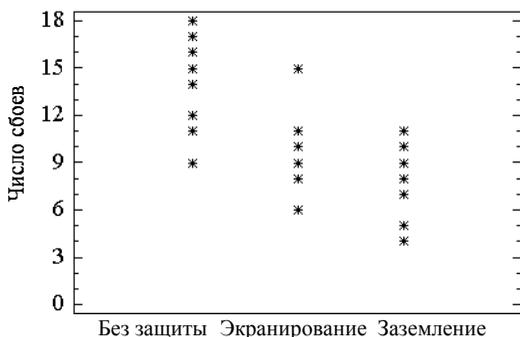


Рисунок 6.6 – Результаты испытаний микроконтроллеров

Рассмотрим случайную величину ξ – количество сбоев МК (микропроцессора) в процессе испытания. Для проверки влияния помехозащитных мероприятий на качество работы МК выдвинем нулевую гипотезу о том, что математические ожидания числа сбоев МК при проведении различных помехозащитных мероприятий не отличаются, т. е. при H_0

$$\begin{aligned} M[\xi | \text{ без мероприятий}] &= \\ &= M[\xi | \text{ экранирование}] = \\ &= M[\xi | \text{ заземление}]. \end{aligned}$$

Альтернативная гипотеза состоит в том, что помехозащитные мероприятия изменяют математическое ожидание числа сбоев, т.е. влияют на качество работы МК.

Для проверки нулевой гипотезы вычислим значение статистического критерия значимости \hat{F} . Здесь n – число испытываемых МК ($n = 24$), k – число групп МК ($k = 3$), n_1 – число МК в 1-й группе (помехонезащищенных, $n_1 = 10$), n_2 – число МК во 2-й группе (использующих экранирование, $n_2 = 7$), n_3 – число МК в 3-й группе (использующих заземление, $n_3 = 7$); x_{ij} – число сбоев j -го образца из i -й группы МК, \bar{x}_i – выборочное среднее число сбоев МК i -й группы.

$$\begin{aligned} x_1 &= (11+9+18+\dots+15+9)/10=13,2; & x_2 &= (9+11+15+\dots+6+10)/7=9,714; \\ x_3 &= (11+5+8+\dots+4+10)/7=7,714; \end{aligned}$$

$$\begin{aligned} \hat{F} &= \frac{\frac{1}{k-1} \sum_{i=1}^k ((x_i - \bar{x})^2 n_i)}{\frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - x_i)^2} = \\ &= \frac{\frac{1}{3-1} \left((13,2-10,583)^2 \cdot 10 + (9,714-10,583)^2 \cdot 7 + (7,714-10,583)^2 \cdot 7 \right)}{\frac{1}{24-3} \left((11-13,2)^2 + \dots + (9-13,2)^2 + (9-9,714)^2 + \dots + (10-7,714)^2 \right)} = \frac{65,688}{8,6884} = 7,56. \end{aligned}$$

Найдем критическое значение критерия значимости, как квантиль распределения Фишера уровня $\alpha = 0,05$ с $v_1 = k - 1 = 3 - 1 = 2$ и $v_2 = n - k = 24 - 3 = 21$ степенями свободы. Поскольку расчетное значение критерия значимости \hat{F} больше критического ($\hat{F} = 7,56 > F_{0,05;2;21} = 3,467$), то с вероятностью, меньшей $\alpha = 0,05$, можно утверждать, что проводимые помехозащитные мероприятия существенно влияют на качество работы МК в условиях внешних электромагнитных помех.

В то же время перечисленные помехозащитные меры могут оказаться несущественными для защиты МК от воздействия помех, поступающих по цепям питания, для чего необходимы дополнительные исследования.

Пример 2. Имеются данные о заработной плате 20 работников предприятия (таблица 6.4).

Таблица 6.4 – Данные о заработной плате работников предприятия

| | | | | | | | | | | |
|----------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| № п/п | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Зарплата, тыс. у. е. | 1,3 | 1,7 | 2,3 | 2,7 | 3,0 | 3,2 | 3,4 | 3,5 | 3,6 | 3,7 |
| № п/п | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Зарплата, тыс. у. е. | 3,9 | 4,2 | 4,4 | 4,5 | 4,7 | 4,8 | 5,0 | 5,2 | 5,7 | 5,8 |

Используя правило сложения дисперсий, определить степень влияния уровня образования работников на заработную плату, если работники со средним специальным образованием имеют заработную плату до 3,5 тыс. у. е., с высшим – более 3,5 тыс. у. е.

В таблице 6.5 приведены основные расчетные результаты.

Таблица 6.5 – Распределение заработной платы работников в зависимости от уровня образования

| Среднее специальное | | | | Высшее | | | |
|---------------------|--------------------------------|-------------------|-----------------------|--------------|--------------------------------|-------------------|-----------------------|
| № п/п | Зарплата (x_i), тыс. у. е. | $x_i - \bar{x}_1$ | $(x_i - \bar{x}_1)^2$ | № п/п | Зарплата (x_i), тыс. у. е. | $x_i - \bar{x}_2$ | $(x_i - \bar{x}_2)^2$ |
| 1 | 1,3 | -1,3 | 1,69 | 1 | 3,6 | -1,0 | 1,00 |
| 2 | 1,7 | -0,9 | 0,81 | 2 | 3,7 | -0,9 | 0,81 |
| 3 | 2,3 | -0,3 | 0,09 | 3 | 3,9 | -0,7 | 0,49 |
| 4 | 2,7 | 0,1 | 0,01 | 4 | 4,2 | -0,4 | 0,16 |
| 5 | 3,0 | 0,4 | 0,16 | 5 | 4,4 | -0,2 | 0,04 |
| 6 | 3,2 | 0,6 | 0,36 | 6 | 4,5 | -0,1 | 0,01 |
| 7 | 3,4 | 0,8 | 0,64 | 7 | 4,7 | 0,1 | 0,01 |
| 8 | 3,5 | 0,9 | 0,81 | 8 | 4,8 | 0,2 | 0,04 |
| | | | | 9 | 5,0 | 0,4 | 0,16 |
| | | | | 10 | 5,2 | 0,6 | 0,36 |
| | | | | 11 | 5,7 | 1,1 | 1,21 |
| | | | | 12 | 5,8 | 1,2 | 1,44 |
| <i>Итого</i> | 21,1 | | 4,57 | <i>Итого</i> | 55,5 | | 5,73 |

Определим внутригрупповые средние и дисперсии:

$$\bar{x}_1 = \frac{\sum_{i=1}^8 x}{n} = \frac{21,1}{8} = 2,6 \text{ тыс. у. е.}, \bar{x}_2 = \frac{\sum_{i=1}^{12} x}{n} = \frac{55,5}{12} = 4,6 \text{ тыс. у. е.},$$

$$\sigma_1^2 = \frac{\sum_{i=1}^8 (x_i - \bar{x}_1)^2}{n} = \frac{4,57}{8} = 0,57 \text{ (тыс. у. е.)}^2, \sigma_2^2 = \frac{\sum_{i=1}^{12} (x_i - \bar{x}_2)^2}{n} = \frac{5,73}{12} = 0,48 \text{ (тыс. у. е.)}^2.$$

Внутригрупповые дисперсии показывают вариацию заработной платы по группам, вызванные всеми возможными факторами (стаж работы, квалификация, должность, производительность труда и др.), кроме образования работников.

Средняя внутригрупповых дисперсий

$$\bar{\sigma}_i^2 = \frac{\sum \sigma_i^2 \cdot N_i}{\sum N} = \frac{0,57 \cdot 8 + 0,48 \cdot 12}{20} = \frac{10,32}{20} = 0,52 \text{ (тыс. у. е.)}^2.$$

Межгрупповая дисперсия исчисляется по формуле $\delta^2 = \sum (\bar{x}_i - \bar{x})^2 N_i / \sum N$.

Определим общую среднюю и межгрупповую дисперсию:

$$\bar{x} = \frac{76,6}{20} = 3,83 \text{ тыс. у. е.}, \quad \delta^2 = \frac{(2,6 - 3,83)^2 \cdot 8 + (4,6 - 3,83)^2 \cdot 12}{20} = 0,96 \text{ (тыс. у. е.)}^2.$$

Межгрупповая дисперсия показывает вариацию заработной платы, обусловленную влиянием образования работников.

Общую дисперсию определяем на основании правила сложения дисперсий:

$$\sigma^2 = \bar{\sigma}_i^2 + \delta^2, \quad \sigma^2 = 0,52 + 0,96 = 1,48 \text{ (тыс. у. е.)}^2.$$

Определим долю межгрупповой дисперсии в общей с помощью эмпирического коэффициента детерминации:

$$\hat{R}^2 = \frac{\delta^2}{\sigma^2}, \quad \hat{R}^2 = \frac{0,96}{1,48} = 0,649.$$

На 64,9 % вариация заработной платы обусловлена влиянием уровня образования работников и на 35,1 % ($1 - 0,649 = 0,351$) – влиянием прочих факторов, не учтенных в группировке.

Эмпирическое корреляционное отношение

$$\hat{R} = \sqrt{0,649} = 0,81.$$

Эмпирическое корреляционное отношение показывает тесноту связи между результативным и факторным признаками. Оно равно 0,81, следовательно, связь между заработной платой и уровнем образования работников – тесная.

Подобным образом может быть проверена значимость влияния технологий производства и обслуживания на надежность электронных устройств уровня зарплаты и стажа на производительность труда и пр.

Многофакторный дисперсионный анализ. Многофакторный дисперсионный комплекс – это совокупность исходных наблюдений, позволяющих статистически оценить действие и взаимодействие нескольких изучаемых факторов на изменчивость результативного признака. Эффект взаимодействия составляет ту часть общего варьирования, которая вызвана различным действием одного фактора при разных градациях другого. Специфическое действие сочетаний в эксперименте выявляется тогда, когда при одной градации первого фактора второй действует слабо или угнетающе, а при другой градации он проявляется сильно и стимулирует развитие результатив-

ного признака. Например, если исследуется вероятность ДТП на определенном участке дороги, то необходимо учитывать состояние дорожного покрытия, время года (гололед), время суток (день или ночь), видимость в направлении движения (плохая обзорность из-за кустов и деревьев), состояние водителя, техническое состояние всех систем транспортного средства, водительские навыки (стаж, реакции), влияние окружения (пассажиров, пешеходов), действия других водителей в динамике ДТП и многих других влияющих факторов.

Дисперсионный анализ данных многофакторного комплекса проводится в два этапа. *Первый* этап – разложение общей вариации результативного признака на варьирование вариантов и остаточное: $C_Y = C_V + C_Z$. На *втором* этапе сумма квадратов отклонения для вариантов разлагается на компоненты, соответствующие источникам варьирования: главные эффекты изучаемых факторов и их взаимодействия. В двухфакторном опыте $C_V = C_A + C_B + C_{AB}$, в трехфакторном – $C_V = C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}$.

Дисперсионный анализ двухфакторного анализа по изучению градаций факторов A (число вариантов l_A) и B (число вариантов l_B), проведенного в n опытах, осуществляется в следующие этапы:

1 Определяются суммы и средние по вариантам и по опыту.

2 Вычисляются общая сумма квадратов отклонений, сумма квадратов для вариантов и остатка:

$$N = l_A l_B n; C = (\sum X)^2 / N; C_Y = \sum X^2 - C; C_V = \sum (V^2 / n) - C; C_Z = C_Y - C_V.$$

3 Для вычисления сумм квадратов по факторам A , B и взаимодействию AB составляется вспомогательная таблица, в которую записываются суммы по вариантам. Суммируя цифры, находят суммы A , суммы B и вычисляют суммы квадратов отклонений для главных эффектов и взаимодействия:

$$C_A = \sum \frac{A^2}{l_B n} - C; C_B = \sum \frac{B^2}{l_A n} - C; C_{AB} = C_V - C_A - C_B.$$

Если взять отношения частных дисперсий к общей, получим равенства

$$\frac{C_V}{C_Y} + \frac{C_Z}{C_Y} = \frac{C_Y}{C_Y} = 1; \quad \frac{C_A}{C_Y} + \frac{C_B}{C_Y} + \frac{C_{AB}}{C_Y} + \frac{C_Z}{C_Y} = \frac{C_Y}{C_Y} = 1.$$

Каждое из этих отношений будет показывать долю участия отдельной частной дисперсии в образовании общей дисперсии. А так как каждая частная дисперсия соответствует одному из частных влияний, то отношение частной дисперсии к общей измеряет долю данного влияния в общем суммарном статистическом влиянии всех факторов, определяющих развитие данного результативного признака. Поэтому доля (выраженная в относительных единицах или в процентах) каждой частной дисперсии в общей их сумме может быть принята за показатель силы влияния, того влияния, которое характеризуется данной частной дисперсией – или одной из факториальных, или случайной.

Таким образом, сила влияния фактора (факторов) в дисперсионном анализе измеряется отношением дисперсий частных к общей: $\eta_i^2 = C_i / C_Y$.

Отношение одной из факториальных дисперсий (C_V, C_A, C_B, C_{AB}) или случайной дисперсии (C_Z) к общей (C_Y) можно назвать основным показателем силы влияния факторов – организованных и неорганизованных.

Рассмотрим подготовку к проведению дисперсионного анализа с точки зрения обеспечения необходимых требований его математического аппарата.

Независимость исследуемых факторов является важнейшим условием дисперсионного анализа, пренебрегать которым нельзя из-за опасности получить бессмысленные или неверные выводы. В случае априорной неопределенности в этом плане необходимо провести *корреляционный анализ* на базе отдельного *специально поставленного эксперимента*. Тогда можно будет опираться хотя бы на некоррелированность факторов (для коррелированных факторов проводить дисперсионный анализ бессмысленно).

Если исследуемые факторы оказались зависимыми или коррелированными, то необходимо попытаться подобрать другие факторы или выбрать только независимые. Большую помощь в этом могут оказать *метод главных компонент* и *факторный анализ*.

Однородность дисперсий в слоях при различных значениях исследуемых факторов можно проверить по критерию Фишера.

Таким образом, рекомендуется следующая *последовательность действий при дисперсионном анализе по выборочным результатам*:

1 Проверка независимости (или некоррелированности) исследуемых факторов методами корреляционного анализа. При необходимости – обеспечение некоррелированности с помощью замены факторов.

2 Проверка нормального распределения исследуемых факторов по критерию согласия Пирсона. При необходимости – замена факторов.

3 Проверка однородности дисперсий по критерию Фишера. При необходимости – замена факторов.

4 Разбиение общей дисперсии на составляющие в соответствии с задачей исследований.

5 Вычисление необходимых межгрупповых и остаточных дисперсий и проверка гипотез о значимости их различия с помощью критерия Фишера.

Достаточно полное описание разнообразных классических примеров применения многомерного статистического анализа можно найти в литературе [18].

6.5 Статистическая проверка адекватности математических моделей

В подразд. 2.2 изложен общий подход к оценке адекватности математических моделей механических систем и процессов. Для адекватности математической модели реальному поведению оригинала рассогласование соот-

ветствующих параметров должно удовлетворять двум критериям: *точности* и *непротиворечивости*. Таким образом, необходим алгоритм проверки этих двух критериев для величины *рассогласования* результатов *контрольного вычислительного эксперимента* с результатами *натурного эксперимента* в тех же условиях: $\Delta u = u_{\text{модели}} - u_{\text{оригинала}}$. Эта величина на практике принимает дискретные значения, так как данные об оригинале регистрируются аппаратурой в конечном числе точек замера. Задачей контрольного вычислительного эксперимента в этом случае является получение данных о тех же параметрах в тех же узловых точках. В результате для статистического анализа предлагается множество значений *рассогласования* $\{\Delta u_i; i = 1, 2, \dots, N\}$ – выборка из генеральной совокупности истинной величины *рассогласования*.

Как известно даже на бытовом уровне, для повышения *точности измерений* проводят не одно измерение, а несколько. Это делается не из-за того, что какое-то из них может оказаться ошибочным, а из-за замечательного свойства *дисперсии* средней арифметической величины измерений – уменьшаться с ростом числа повторений опытов:

$$D_N = \frac{D}{N} \quad \text{или} \quad \sigma_N = \frac{\sigma}{\sqrt{N}},$$

где D и σ – дисперсия и среднее квадратическое отклонение в одном опыте (измерении); D_N и σ_N – дисперсия и среднее квадратическое отклонение результата осреднения замеров по N опытам. Поэтому с помощью большего числа опытов достигают меньшего рассеивания (среднего квадратического отклонения) данных, т.е. большей точности.

Однако одной величины *среднего квадратического отклонения* для оценки точности результатов недостаточно. Такая оценка страдает неполнотой, так как не учитывает, насколько часто встречаются большие и малые, положительные и отрицательные *рассогласования*.

Итак, точность следует определять единой оценкой всего множества наблюдаемых значений случайной величины *рассогласования* результатов вычислительного эксперимента и "истинного" значения наблюдаемой величины. По своему смыслу в качестве такой "истинной" единой оценки должно выступать *математическое ожидание* *рассогласования*, которое обозначим a . Но об этом "истинном" значении *рассогласования* мы ничего не можем знать достоверно, остается о нем судить лишь с определенной вероятностью по ограниченному числу опытов. Поэтому наиболее полную оценку точности (вернее, погрешности) вычислительного эксперимента дает *доверительный интервал* для математического ожидания *рассогласования*. Так, например, может звучать вывод о точности в этом случае: с доверительной вероятностью 0,98 гарантируется погрешность коэффициента сцепления не более 0,3°. Критерием оценки точности тогда является соблюдение этой

парой значений условий, приемлемых с точки зрения целей исследования.

Единственным практическим *недостатком* такой оценки может быть лишь необходимость знать закон распределения исследуемого рассогласования. Однако, во-первых, для оценки погрешности по подавляющему большинству параметров технических систем можно считать такое распределение нормальным, хотя бы приблизительно в некоторой области, а во-вторых, можно практиковать построение несимметричных доверительных интервалов, отражающих разную степень строгости требований по точности. Доверительный интервал для оценки истинного значения рассогласования a по найденному значению выборочной средней величины рассогласования Δu при неизвестном значении σ , но известной несмещенной выборочной оценке среднего квадратического отклонения s :

$$\overline{\Delta u} - t_{\gamma, N-1} \frac{s}{\sqrt{N}} < a < \overline{\Delta u} + t_{\gamma, N-1} \frac{s}{\sqrt{N}},$$

где $\overline{\Delta u} = \frac{1}{N} \sum_{j=1}^r N_j \Delta u_j$; $s^2 = \frac{1}{N-1} \sum_{j=1}^r N_j (\Delta u_j - \overline{\Delta u})^2$; $t_{\gamma, N-1}$ определяется по распределению Стьюдента; N_j – число попаданий в j -й интервал (из r) наблюдаемых рассогласований Δu_j ; Δu_j – середина j -го интервала; N – общее число наблюдаемых значений Δu .

Размер доверительного интервала тем меньше, чем меньше доверительная вероятность γ и чем больше число опытов N . Естественно, при планировании вычислительного эксперимента следует стремиться к тому, чтобы такая оценка погрешности (т.е. доверительный интервал) не выходила за границы требуемой с точки зрения целей исследования погрешности $\pm \delta$, чего можно добиться разумным увеличением числа опытов N и уменьшением доверительной вероятности γ . Иными словами, следует стремиться к тому, чтобы доверительный интервал целиком укладывался внутри допустимой погрешности (например, от $-\delta$ до $+\delta$).

Если такого условия не удастся выполнить на данной серии опытов, то следует или увеличить число опытов N , или уменьшить доверительную вероятность γ . Однако последнее значительно слабее влияет на результат, тем более, что значения доверительной вероятности $\gamma < 0,7$ применять нежелательно, так как это означает, что почти треть значений рассогласований может выходить за границы доверительного интервала (треть рассогласований принимает неконтролируемые значения).

Однако оценки точности с помощью доверительного интервала для математического ожидания рассогласования тоже недостаточно. Даже в том случае, когда рассеивание результатов мало, и гарантируется с определен-

ной доверительной вероятностью, может существовать *систематическая погрешность*. Ее присутствие свидетельствует о *закономерности* рассогласования между оригиналом и моделью и не позволяет пользоваться ею.

Оценка систематической ошибки делается по величине выборочного среднего рассогласований, так как эта точечная оценка по своему смыслу характеризует среднее значение рассогласований, "*присутствующее постоянно*". Указанную оценку можно получить с помощью критерия Стьюдента, построенного на t -распределении, в предположении идеального случая: истинное значение математического ожидания погрешности равно нулю, т.е. $a = 0$. По указанному критерию сравниваются две величины:

$$t = \frac{\overline{\Delta u}}{s} \sqrt{N} \text{ и } t_{\text{крит}}(1-\alpha, N-1),$$

где $t_{\text{крит}}(1-\alpha, N-1)$ – предельное значение критерия, определяемое по таблице распределения Стьюдента при уровне значимости α (вероятности совершить ошибку первого рода: отвергнуть верную гипотезу об $a = 0$) с $N - 1$ степенями свободы.

Если $|t| < t_{\text{крит}}$, то систематическая ошибка *незначима*, т.е. не существенна и можно считать $a = 0$. В случае противоположного неравенства: $|t| > t_{\text{крит}}$, систематическая ошибка *значима*, т.е. не может считаться нулевой. В последнем случае математическая модель может считаться достаточно точной только при выполнении двух условий: доверительный интервал не выходит за границы требуемой с точки зрения целей исследования погрешности $\pm \delta$, а выборочное среднее погрешностей Δu пренебрежимо мало с точки зрения целей исследования, чтобы учитывать такую систематическую ошибку.

Непротиворечивость со статистической точки зрения может означать незначимость рассогласования, иными словами, неподверженность каким-либо закономерностям, непринципиальность – хаотичность. Последний термин и служит основой для построения критерия оценки непротиворечивости с помощью критерия согласия Пирсона χ^2 . Для этого достаточно, чтобы рассогласование между моделью и оригиналом имело характер *простой ошибки измерений*, т.е. подчинялось нормальному закону распределения с нулевым математическим ожиданием $a = 0$.

По критерию Пирсона χ^2 для этого сравниваются $\hat{\chi}^2$ и $\chi_{\alpha, \nu}^2$. Если $\hat{\chi}^2 < \chi_{\alpha, \nu}^2$, то различие статистического и гипотетического (нормального) законов распределения *незначимо*. То есть при заданном уровне значимости α гипотезу о поведении рассогласования эксперимента и "истины", как случайной ошибки измерений, можно принять и можно считать результаты вычислительного эксперимента не противоречащими реальному поведению оригинала. В случае противоположного неравенства: $\hat{\chi}^2 > \chi_{\alpha, \nu}^2$ расхождение

значимо (закономерно, не может считаться случайным), и гипотезу следует отвергнуть, т.е. результаты вычислительного эксперимента противоречат реальному поведению оригинала.

Только в том случае, когда выполнены условия и требуемой точности, и непротиворечивости, можно считать результаты вычислительного эксперимента адекватными реальному поведению оригинала с доверительной вероятностью γ и уровнем значимости α в эксперименте из N опытов.

Таким образом, можно составить **алгоритм проверки адекватности математической модели реальному поведению оригинала** с помощью статистических критериев, предварительно задав допустимую погрешность $\pm\delta$, уровень значимости α и доверительную вероятность γ , исходя из целей исследования. В этом алгоритме строго соблюдается последовательность проверки статистических критериев, каждый следующий из которых опирается на вывод предыдущего.

1 Выбирается один из параметров объекта, для которого есть результаты наблюдения $\{u_{\text{оригинала } i}\}$ в N точках, и соответствующий параметр $\{u_{\text{модели } i}\}$, полученный в контрольном вычислительном эксперименте в тех же условиях в тех же точках.

2 Вычисляются разности $\Delta u_i = u_{\text{модели } i} - u_{\text{оригинала } i}$.

3 Вся область значений Δu разбивается на r интервалов таким образом, чтобы в каждой из них попало не менее пяти значений Δu_i .

4 Производится расчет количества попадания Δu_i в каждый j -й ($1 \leq j \leq r$) интервал – частот N_j .

5 Определяются статистические оценки параметров распределения случайной величины Δu : выборочное среднее $\overline{\Delta u} = \frac{1}{N} \sum_{j=1}^r N_j \Delta u_j$, где Δu_j – середина j -го интервала; и несмещенная оценка дисперсии

$$s^2 = \frac{1}{N-1} \sum_{j=1}^r N_j (\Delta u_j - \overline{\Delta u})^2.$$

6 Для проверки *непротиворечивости*, т.е. подчиненности рассогласования нормальному закону распределения, применяется критерий согласия Пирсона χ^2 . Вычисленное значение $\hat{\chi}^2$ сравнивается с табличным значением распределения Пирсона $\chi_{\alpha, \nu}^2$, где наиболее употребительные значения $\alpha = 0,05$ или $\alpha = 0,01$. Если $\hat{\chi}^2 < \chi_{\alpha, \nu}^2$, то распределение Δu незначимо отличается от нормального, т.е. результаты вычислительного эксперимента можно считать не противоречащими реальному поведению оригинала. Если $\hat{\chi}^2 > \chi_{\alpha, \nu}^2$, то значимое отличие распределения Δu от нормального свиде-

тельствует о противоречии результатов вычислительного эксперимента реальному поведению оригинала и исследования адекватности следует прекратить.

Замечание. Критерий согласия Пирсона χ^2 не требует сведений о законах распределения, а все последующие пункты требуют знания закона распределения и основаны на том, что Δu распределено по нормальному закону распределения с нулевым математическим ожиданием $a = 0$ и средним квадратическим отклонением s .

7 Для оценки систематической ошибки проверяется гипотеза о равенстве нулю математического ожидания ($a = 0$) рассогласования Δu с помощью критерия Фишера, для чего вычисляется статистика $t = \frac{\overline{\Delta u}}{s} \sqrt{N}$, которая сравнивается с $t_{\text{крит}}(1-\alpha, N-1)$, определяемым по таблице распределения

Стьюдента при уровне значимости 0,05 или 0,01 и числе степеней свободы $N - 1$. Если $|t| > t_{\text{крит}}$, то дальнейшие исследования адекватности нужно прекратить, так как это означает существование недопустимой систематической погрешности между результатами вычислительного эксперимента и реальным поведением оригинала. Если $|t| < t_{\text{крит}}$, то систематическая погрешность отсутствует и можно продолжать исследования.

8 Для оценки *точности* математической модели строится доверительный интервал для математического ожидания рассогласования при заданной доверительной вероятности γ (обычно 0,8; 0,9; 0,99; или 0,999):

$$\overline{\Delta u} - t_{\gamma, N-1} \frac{s}{\sqrt{N}} < a < \overline{\Delta u} + t_{\gamma, N-1} \frac{s}{\sqrt{N}},$$

где $t_{\gamma, N-1}$ определяется по таблице распределения Стьюдента. Если радиус доверительного интервала не превосходит допустимый погрешности, т.е.

$$t_{\gamma, N-1} \frac{s}{\sqrt{N}} \leq \delta,$$

то математическую модель можно считать достаточно точной по отношению к оригиналу.

В дополнение к этому можно построить доверительный интервал для среднего квадратического отклонения рассогласования

$$s \sqrt{\frac{N-1}{\chi_{1-\frac{\gamma}{2}, N-2}^2}} < \sigma < s \sqrt{\frac{N-1}{\chi_{\frac{\gamma}{2}, N-2}^2}}.$$

9 Если по п. 7 можно считать математическую модель не противоречащей оригиналу, а по п. 8 – достаточно точной, то результаты расчетов адекватны реальному поведению оригинала.

Рассмотренные методы позволяют поставить любой технологический процесс под *математически строгий контроль*. Это означает, что, организуя специальным образом сбор статистического материала о параметрах производства, можно делать научно обоснованные (с оценкой ошибок) выводы о качестве процесса, тенденциях его изменения и о качестве продукции.

Регулирование качества технологических процессов на основе текущего контроля получило широкое развитие в 60-х годах XX века в основном в виде техники *контрольных карт*. Сбор и обработка статистической информации для них предельно формализованы, а преднамеренная ее фальсификация требует серьезной математической подготовки и недоступна на производстве. В последние годы стали появляться новые статистические методы контроля, учитывающие экономику и безопасность.

Если в качестве математического ожидания a рассматривать *нормативное значение* контролируемого параметра технологического процесса (на которое он *должен* быть настроен), то за соблюдением этой настройки можно будет следить на основании замеров выборочной средней \bar{x} . Тогда, исходя из заданной доверительной вероятности γ и величин допуска в меньшую и в большую сторону, можно будет отмечать случаи нарушения заданных требований к технологическому процессу.

Если закон распределения не удастся подобрать даже приближенно, то используется неравенство Чебышева:

$$P(|\bar{x} - a| < \delta) \geq 1 - \frac{\sigma^2}{\delta^2},$$

правда, в этом случае необходимо знание дисперсии σ^2 , которую можно оценить дополнительными *предварительными* исследованиями процесса.

Контрольная карта представляет собой график изменения контролируемого параметра по времени, на котором заранее нанесены *контрольные границы* (границы допуска) в соответствии с границами доверительного интервала. На рисунке 6.7 показана для примера контрольная карта средних \bar{x} . Контроль заключается в нанесении статистических данных на карту и принятии решения о вмешательстве в технологический процесс. Процесс считается *статистически подконтрольным*, пока значения \bar{x} лежат внутри контрольных границ. При выходе \bar{x} за эти пределы производство следует приостановить для перенастройки или переналадки оборудования.

Существуют разработанные ряд *методов контроля и управления случайными процессами* [17], основной целью которых было обеспечить экономичное обслуживание по состоянию. Следует особо отметить, что все эти математически строгие результаты доведены до алгоритмов, пригодных к практическому использованию.

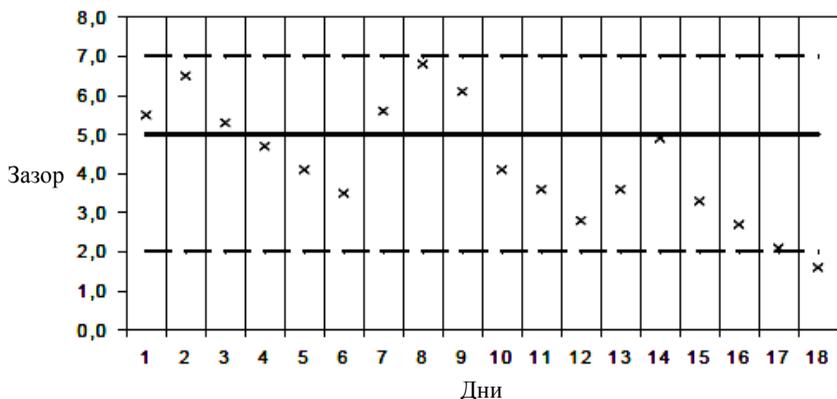


Рисунок 6.7 – Контрольная карта средних значений контролируемого параметра

Например, пусть $\xi(t)$ – случайный монотонно не убывающий во времени t процесс, о развитии которого можно судить по некоторому контролируемому признаку, наблюдаемое значение которого в момент времени t_k обозначено как x_k . Это соответствует обычной технологической практике, когда контрольные мероприятия проводятся через определенные промежутки времени, не обязательно равные. Суммарные затраты на наблюдение и регулировку системы в период ее работоспособности, т.е. пока $x_k < L$ (L – пороговое значение, соответствующее отказу) составляют сумму C , а штраф (потери при отказе, т.е. выходе системы за порог отказа, когда $x_k > L$) – величину A .

Если для такого процесса изменение признака $\Delta x_k = x_k - x_{k-1}$ на очередном интервале времени контроля не зависит от текущего времени, т.е. не зависит от предыстории и распределено по известному (хотя бы эмпирически) закону $P(\Delta x_k < z) = F(z)$, то минимум удельных затрат на контроль и обслуживание достигается при прекращении эксплуатации (для ремонта, замены и т.п.) в момент времени t_{k-1}^* , когда впервые будет нарушено неравенство или, что то же самое, когда впервые выполнится неравенство

$$1 - P(\Delta x < L - x_{k-1}) \leq \frac{C}{A(k-1)}.$$

Алгоритм действий при организации управления качеством по такому методу рассмотрим на численном примере, проиллюстрированном на рисунке 6.8. На левом графике приведен эмпирический закон распределения $F(z)$, полученный предварительным сбором статистической информации,

по которому с помощью *единичного жребия* можно симитировать разовое изменение контролируемого признака Δx_k . На правом графике представлено монотонное развитие процесса от состояния полностью исправной системы (признак $x = 0$) к отказу (признак $x = L$). Видно, как с течением времени контролируемый признак ("дефектность", количество отказавших подсистем и т. п.) стремится к порогу L , за которым система считается отказавшей.

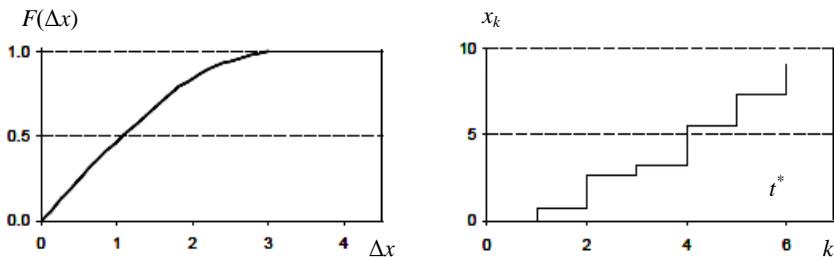


Рисунок 6.8 – Пример метода контроля состояния

Второй рассматриваемый здесь метод контроля и управления технологическими процессами весьма перспективен для учета изменения интенсивности отказов, когда она растет у *стареющей* системы. Пусть интенсивность отказов непостоянна ($\lambda \neq \text{const}$, $\lambda'(t) \geq 0$), т. е. увеличивается со временем по известному (хотя бы эмпирическому) закону: $\lambda = \lambda(t)$.

Обозначим:

T_1 – среднее время замены исправной системы;

T_2 – среднее время замены неисправной системы (обычно на практике $T_2 > T_1$ из-за затрат на ликвидацию последствий отказа);

T_3 – момент времени *предупредительной замены* системы;

$F(\omega) = 1 - e^{-\int_0^\omega \lambda(t) dt}$ – вероятность безотказной работы системы;

$p(t, \Delta t)$ – вероятность того, что в момент времени t система исправна и проработает еще в течение времени Δt ;

$p(t, \Delta t) \Rightarrow \max$ – критерий оптимальности на всем сроке жизни системы с учетом многократного ее восстановления. Тогда оптимальное T_3 можно определить из равенства

$$\frac{T_1}{T_2 + \Delta t} = 1 - \frac{1}{1 - F(T_3) + \lambda(T_3) \int_0^{T_3} [1 - F(\omega)] d\omega}.$$

7 ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

7.1 Статистические методы планирования эксперимента

7.1.1 Проблемы построения эксперимента

Планирование эксперимента появилось в агробиологии и связано с именем английского статистика и биолога сэра Рональда Эйлмера Фишера, упоминавшегося в разделе дисперсионного анализа. В начале XX века на агробиологической станции в Ротамстеде (Великобритания) начались исследования влияния удобрений на урожайность различных сортов зерновых. Ученые вынуждены были считаться как с большой изменчивостью объектов исследования, так и с большой продолжительностью опытов (около года). В этих условиях не было иного пути, кроме разработки продуманного плана эксперимента для уменьшения негативного влияния указанных факторов на точность выводов. Применив статистические знания к биологическим проблемам, Фишер пришел к разработке собственных принципов теории статистического вывода и положил начало новой науке о планировании и анализе экспериментов.

Исследование сложных технических систем (например, аэродинамики и систем безопасности автомобилей) требует постановки и проведения сложных и дорогостоящих экспериментов (например, летных испытаний и продувок в аэродинамических трубах, краш-тестов и т.д.) с исследованием влияния множества параметров. Для этого необходимо **планирование эксперимента** – совокупность действий, объединенных целью исследования и направленных на разработку стратегии экспериментирования от начальных до заключительных этапов изучения объекта исследований (от получения априорной информации до создания работоспособной математической модели или определения оптимальных условий) [2, 22].

С другой стороны, между объектом исследования и результатами эксперимента, полученными и обработанными в целях этого исследования, существует огромная дистанция. Дело не только в том, что эксперимент может быть поставлен разным образом, а его результаты обработаны различными методами. Дело еще в том, что объект исследования – оригинал – многообразен и бесконечно сложен в своих проявлениях, только часть из которых мы можем наблюдать в эксперименте – модели. То есть необходимо четко себе представлять, что результаты эксперимента – это лишь более или менее представительная выборка при наблюдении за генеральной совокупностью. А это требует использования методов математической статистики для проведения успешного анализа результатов эксперимента. Особую роль здесь играют *модели дисперсионного и регрессионного анализа*.

Ситуация еще более усложняется, если исследуется не единичный объект, а сложная система. В этом случае наибольшим затруднением становится

ся множественность целей подсистем и элементов, а также множественность критериев оценки их работы (многокритериальность). Различные критерии могут вступать в противоречие друг с другом, однако даже их частичное игнорирование может полностью исказить систему как целое. Например, если для транспортного маршрутного предприятия приоритет – наполняемость салона, то для пассажиров – удобство и безопасность поездки.

Эффективность авторемонтного завода не может быть оценена каким-либо единым показателем, например, доходом. Завод работает не в изолированном пространстве, а в среде предприятий, которые противятся росту цены ремонта. По-видимому, кроме дохода необходимо учитывать и себестоимость, и производительность труда, и общий объем работ, и что-то еще. Иногда требуется учитывать дополнительные факторы. Так, например, резкое увеличение производительности труда приведет к простоям из-за отсутствия фронта работ. Таким образом, при неправильной организации труда (игнорировании множественности задач и критериев деятельности подсистем) безусловно прогрессивный рост производительности труда может обернуться общим проигрышем.

Возникает необходимость разработки такой постановки эксперимента, которая обеспечила бы выполнение сразу нескольких условий: практическую осуществимость (быстрота, экономичность), достаточную полноту (для обоснования выводов), наглядность (удобство интерпретации).

В теории *планирования эксперимента* сформировалась система основных терминов и понятий. Прежде всего имеется в виду, что изучаемая система может быть представлена "черным ящиком". Кроме того, вводятся следующие определения.

Эксперимент – это система операций, воздействий и (или) наблюдений, направленных на получение информации об объекте при исследовательских испытаниях [2, 22].

Опыт – это отдельная часть эксперимента, воспроизводящая исследуемое явление в определенных задаваемых условиях при возможности регистрации его результатов.

Пассивный эксперимент – эксперимент в отсутствии управляемых факторов: выходные факторы зависят только от неуправляемых входных и неконтролируемых факторов.

Активный эксперимент – эксперимент в отсутствии неуправляемых входных факторов: выходные факторы зависят только от управляемых входных и неконтролируемых факторов.

Теория планирования эксперимента занимается именно активным экспериментом и предполагает планирование значений управляемых входных факторов X в области их возможных значений. Таким образом, *планом эксперимента* является некоторая совокупность уровней факторов X , построенная для определенных целей исследования. Обычно он изображается в

виде матрицы плана, в которой строки отвечают опытам, а столбцы – факторам, т.е. элемент x_{ij} матрицы плана обозначает определенный уровень j -го фактора в i -м опыте:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{21} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nk} \end{bmatrix}.$$

Для наиболее экономичного получения достоверного результата, удовлетворяющего заданным требованиям, необходимо решить ряд проблем постановки эксперимента:

1 Собственно эксперимент:

- а) формулировка целей и задач эксперимента;
- б) выбор наблюдаемого выходного фактора;
- в) выбор управляемых факторов;
- г) выбор уровней этих факторов (количественных или качественных, фиксированных или случайных);
- д) подбор сочетаний уровней факторов.

2 Планирование эксперимента:

- а) определение необходимого числа опытов;
- б) определение порядка проведения отдельных опытов;
- в) выбор метода рандомизации;
- г) составление математической модели для описания результатов.

3 Анализ результатов эксперимента:

- а) сбор и обработка данных;
- б) вычисление статистик (выборочных функций) для проверки гипотез;
- в) интерпретация результатов эксперимента.

Для успешного решения первой из этих проблем применяются *методы теории моделирования*, третьей – *методы обработки информации*. Методам решения второй проблемы и частично первой посвящена *теория планирования эксперимента*.

Основные **принципы планирования эксперимента**, изложенные ниже, позволяют сделать эксперимент практически реализуемым и информативным.

1 *Отказ от полного перебора возможных входных состояний*. Для получения исчерпывающей информации о поведении выходных факторов Y , вообще говоря, необходимо проведение такого числа опытов в эксперименте, чтобы все уровни каждого из входных факторов встречались друг с другом во всех возможных комбинациях. Понятно, что с ростом числа факторов и уровней объем такого эксперимента быстро увеличивается, а для непрерывно изменяющегося хотя бы одного фактора становится бесконечным. Поэтому от полного перебора отказываются сознательно, строго отбирая

лишь *существенные факторы* и вводя дискретную сетку уровней непрерывных факторов, исходя из необходимого вида математической модели.

2 *Последовательное усложнение математической модели* (принцип последовательного планирования). При постановке сложного эксперимента бывает весьма полезно пользоваться приемами последовательного анализа, следуя Е.С. Вентцель [23]: "Основной принцип теории планирования эксперимента состоит в том, что любое принятое заранее решение должно пересматриваться с учетом полученной новой информации".

3 *Сопоставление с шумом*. Точность регрессионной модели должна сопоставляться с интенсивностью случайного неконтролируемого воздействия. Действительно, бессмысленно ставить дорогостоящий эксперимент для получения точной модели, если результаты эксперимента сами по себе обладают большой итоговой погрешностью. Конечно, повторные опыты могут существенно увеличить точность эксперимента, но здесь имеется в виду именно итоговая погрешность всего эксперимента с учетом повторения опытов.

4 *Рандомизация* (принцип приведения к случайности влияния факторов) – это обеспечение случайности влияния действующих на систему факторов, не поддающихся или поддающихся с трудом учету и контролю. В рандомизированной системе влияние действующих факторов можно считать вполне случайным, что позволяет учитывать их статистически. Простейший прием рандомизации плана эксперимента состоит в случайном переборе уровней факторов. Понятие рандомизации ввел Фишер, заметивший, что разделение общей дисперсии на частные, соответствующие исследуемым факторам, тем эффективнее, чем случайнее вклад каждого фактора. Этот же вывод можно отнести и к неучтенным на первых порах факторам, что подталкивает к учету как можно большего числа факторов. Иными словами, рандомизация – это искусственный перевод неслучайных факторов, которые невозможно учесть в эксперименте, в разряд случайных, что позволяет избавиться от систематических погрешностей, связанных с этими факторами.

5 *Оптимальность плана* (наличие критерия оптимальности). Цели исследований можно достичь только в таком эксперименте, который практически осуществим, достаточно информативен и нагляден. Все это требует применения оптимизации, т.е. формулировки критерия оптимальности. При выработке критерия оптимальности плана эксперимента безусловно следует учесть и минимизацию объема и стоимости эксперимента, и максимизацию объема и эффективности (в статистическом смысле) получаемой информации. К последним в современном планировании эксперимента принято относить критерии оптимальности планов для регрессионного анализа. Вообще говоря, даже для одной и той же задачи могут быть сформулированы несколько критериев оптимальности. Из нескольких десятков таких *критериев* отметим наиболее распространенные [2], связанные с точностью оце-

нивания коэффициентов регрессии: D -оптимальности, который обеспечивает минимизацию обобщенной дисперсии оценок коэффициентов регрессионной модели (минимизацию объема эллипсоида рассеяния); A -оптимальности, обеспечивающий минимизацию суммы дисперсий оценок коэффициентов регрессионной модели, и E -оптимальности, осуществляющий минимизацию наибольшей оси эллипсоида рассеяния.

7.1.2 Назначение плана эксперимента

Целью планирования эксперимента было сформулировано как получение максимума достоверной информации при минимуме затрат. Под затратами здесь могут пониматься не только материальные затраты непосредственно на эксперимент, но и затраты времени и ресурсов.

При планировании эксперимента следует иметь в виду, что информация должна быть получена *достоверной* и сформулированной наиболее компактно. Эти условия можно выполнить только с помощью дисперсионного и регрессионного анализа, т. е. с помощью подбора дисперсионных и регрессионных математических моделей, наилучшим образом соответствующих результатам эксперимента. То есть необходимо получение таких оценок параметров моделей, статистически определенных по результатам эксперимента, которые обладают свойствами *эффективности, состоятельности и несмещенности*.

Эффективность статистической оценки, как известно, тем выше, чем меньше ее дисперсия. А поскольку дисперсия оценки любого параметра зависит от множества факторов: условий постановки эксперимента, его погрешности и, как видно из нижеприведенных примеров, от его плана, то очевидна возможность управления общей дисперсией через дисперсии от влияния этих отдельных факторов.

Перечисленные требования к эксперименту, конкретизированные в терминах математической статистики, позволяют осознанно выбирать наиболее удачный план с точки зрения применения тех или иных методов статистического анализа.

Пример. При наличии погрешности весов, для уменьшения ошибки измерения может быть использован способ взвешивания трех предметов A, B, C заключается в предварительном взвешивании пустой чашки весов, а затем отдельных предметов.

План такого эксперимента показан в таблице 7.1. Здесь матрицей плана является матрица из $+1$ и -1 , представляющих два уровня каждого фактора: $x_{ij} = +1$ и $x_{ij} = -1$ (для каждого предмета это означает присутствие или отсутствие его на чашке весов).

Таблица 7.1 – Пример плана эксперимента

| № опыта | A | B | C | Результат |
|---------|-----|-----|-----|-----------|
| 0 | -1 | -1 | -1 | y_0 |
| 1 | +1 | -1 | -1 | y_1 |
| 2 | -1 | +1 | -1 | y_2 |
| 3 | -1 | -1 | +1 | y_3 |

В итоге вес A определится как $y_1 - y_0$, а дисперсия этого результата – формулой

$$\sigma^2 \{A\} = \sigma^2 \{y_1 - y_0\} = 2\sigma^2 \{y\},$$

где $\sigma \{y\}$ – погрешность однократного взвешивания.

Этот же эксперимент можно провести по плану, в котором взвешивание пустой чашки весов заменено взвешиванием всех трех предметов вместе (таблица 7.2). В этом случае вес предмета A определится выражением $0,5(y_1 - y_2 - y_3 + y_4)$, а дисперсия $\sigma^2 \{A\} = \sigma^2 \left\{ \frac{1}{2}(y_1 - y_2 - y_3 + y_4) \right\} = \sigma^2 \{y\}$ окажется вдвое меньше, т.е. результат вдвое точнее!

Таблица 7.2 – Пример плана эксперимента

| № опыта | A | B | C | Результат |
|---------|-----|-----|-----|-----------|
| 0 | +1 | +1 | +1 | y_4 |
| 1 | +1 | -1 | -1 | y_1 |
| 2 | -1 | +1 | -1 | y_2 |
| 3 | -1 | -1 | +1 | y_3 |

Такой эффект достигнут за счет того, что каждый вес теперь получается в результате четырех взвешиваний, а не двух. Иными словами, для того, чтобы определить вес каждого предмета по первому плану, требуется произвести только два опыта: замер с пустой чашей весов и с эти предметом, зато по второму плану все четыре опыта участвуют в определении веса каждого предмета. Таким образом, стоящее в знаменателе формулы дисперсии число степеней свободы выросло с 2 до 4. В этом смысле первый план фактически распадается на три отдельных плана однофакторных двухуровневых экспериментов, а второй оказывается планом трехфакторного двухуровневого эксперимента.

При изучении систем с большим количеством факторов приходится иметь дело с понятием полноты плана, отражающим перебор всех уровней исследуемых факторов. Поэтому введем это понятие. *Полным факторным экспериментом* называется такой эксперимент, план которого обеспечивает полный перебор всевозможных сочетаний всех уровней и всех факторов. План полного факторного эксперимента называют *полным планом*. Объем полного плана (т.е. количество необходимых опытов) легко подсчитать:

$$h_1 h_2 \dots h_k = \prod_{j=1}^k h_j,$$

где k – число факторов, а h_j – число уровней каждого j -го фактора. Так, например, планы в таблицах 7.1 и 7.2 неполные.

Если уровни исследуемых факторов распределены в планах симметрично, т.е. все встречаются одинаковое число раз, то такие планы называются сбалансированными. Симметрия эта не геометрическая, а "смысловая": если в одном опыте встречается один уровень данного фактора, то в каком-то другом опыте должен встретиться и другой. Например, если есть опыт ABC , то должны быть опыты и BCA , и CAB , что обеспечит симметрию опытов по использованию уровней A , B , C факторов.

7.1.3 Планирование объема эксперимента

Допустим, что эксперимент имеет целью определение с контролируемой погрешностью значения единственного параметра x наблюдаемого явления. Изложим по порядку условия сложения различные методы определения объема плана такого эксперимента.

1 Простейший подход к планированию объема эксперимента выражается известной поговоркой: "Семь раз отмерь, один – отрежь!". Математическое обоснование этой житейской мудрости ясно из следующего.

Поскольку погрешность связана с рассеиванием параметра, т.е. с отклонением этой случайной величины от своего математического ожидания, то простейший подход к оценке погрешности можно свести к точечной оценке рассеивания. На языке математической статистики это выражается известным соотношением

$$\sigma\{\bar{x}\} = \frac{\sigma}{\sqrt{N}},$$

где σ – среднее квадратическое отклонение контролируемого параметра x в одиночном опыте (измерении); \bar{x} – его среднее выборочное значение по N одинаковым опытам; $\sigma\{\bar{x}\}$ – среднее квадратическое отклонение среднего выборочного.

Итак, если априори известно (например, в виде паспортных данных измерительной техники) значение σ контролируемого параметра при одиночном опыте, то вышеприведенная формула позволяет определить необходимое число опытов для обеспечения требуемой погрешности δ :

$$N = \frac{\sigma^2}{\delta^2},$$

при этом несущественен закон распределения x .

2 Точечные оценки контролируемого параметра обладают всеми необходимыми свойствами несмещенности, состоятельности, эффективности и достаточности только при больших объемах эксперимента. Кроме того, определение объема эксперимента по дисперсии не дает вероятностной оценки обеспечения заданной погрешности. А такая оценка необходима, поскольку результаты любого эксперимента следует рассматривать как выборку из генеральной совокупности проявления исследуемого процесса. Поэтому более качественную оценку требуемого объема эксперимента следует проводить с помощью доверительных интервалов. Однако для их построения необходимо знать закон распределения этого параметра; в отличие от предыдущих методов – это и есть "плата" за качество оценки.

Поскольку контролируемый параметр чаще всего оценивают с помощью выборочного среднего, постольку для оценки его точности можно исполь-

зовать доверительный интервал для математического ожидания этого контролируемого параметра. Этот интервал обычно строится симметрично относительно выборочного среднего, и его радиус непосредственно определяет точность оценки (вернее, погрешность) при заданной "уверенности" – доверительной вероятности γ . Так, например, для нормально распределенного параметра погрешность его средней выборочной оценки (радиус доверительного интервала, половина его) определяется одной из формул

$$\delta = u_{0,5\gamma} \frac{\sigma}{\sqrt{N}} \text{ или } \delta = t_{\gamma} \frac{s}{\sqrt{N}},$$

(первой – при априори известном значении σ , второй – при неизвестном σ , но найденной выборочной несмещенной его оценке – s), где $u_{0,5\gamma}$ – аргумент стандартизованной функции нормального закона распределения; γ – назначенная доверительная вероятность; t_{γ} – из таблицы распределения Стьюдента с $N - 1$ степенями свободы при вероятности $p = 1 - \gamma$. Отсюда вытекают формулы для определения объема эксперимента, необходимого для обеспечения погрешности δ с доверительной вероятностью γ :

$$N = u_{0,5\gamma}^2 \frac{\sigma^2}{\delta^2} \text{ или } N = \frac{t_{\gamma}^2 s^2}{\delta^2}.$$

В общем случае контролируемой величины, распределенной по другому закону (не нормальному), следует исходить из общего определения доверительного интервала и строить его на основании именно этого закона распределения. Можно также пользоваться приближенными законами распределения.

В таблице 7.3 приведены формулы для определения объема эксперимента, необходимого для проверки соответствующих конкурирующих гипотез [α – вероятность ошибки I рода (отвергнуть верную гипотезу); β – вероятность ошибки II рода (принять неверную гипотезу)].

Таблица 7.3 – Формулы для определения объема эксперимента

| Конкурирующая гипотеза | Требуемый объем эксперимента | |
|------------------------|--|---|
| | при известном σ | при неизвестном σ |
| $a \neq a_0$ | $u_{1-\frac{\alpha}{2}}^2 \frac{\sigma^2}{(\bar{x} - a_0)^2}$ | $t_{1-\frac{\alpha}{2}}^2 \frac{s^2}{(\bar{x} - a_0)^2}$ |
| $a > a_0$ | $u_{1-\alpha}^2 \frac{\sigma^2}{(\bar{x} - a_0)^2}$ | $t_{1-\alpha}^2 \frac{s^2}{(\bar{x} - a_0)^2}$ |
| $a < a_0$ | $u_{\frac{\alpha}{2}}^2 \frac{\sigma^2}{(\bar{x} - a_0)^2}$ | $t_{\alpha}^2 \frac{s^2}{(\bar{x} - a_0)^2}$ |
| $a = a_1$ | $\left(u_{\frac{1-\beta}{2}}^2 + u_{\frac{1-\alpha}{2}}^2 \right) \frac{\sigma^2}{(a_1 - a_0)^2}$ | $\left(t_{1-\alpha}^2 + t_{1-\beta}^2 \right) \frac{s^2}{(a_1 - a_0)^2}$ |

7.2 Особые методы планирования эксперимента

7.2.1 Специальные приемы планирования эксперимента

Классическая теория планирования эксперимента представляет собой набор приемов составления матрицы дискретного эксперимента с конечным числом уровней факторов. В этом смысле планирование эксперимента с непрерывными значениями факторов принципиально невозможно и приходится искусственно вводить *дискретные уровни*. Если рассматриваемые факторы имеют вид простых измеримых параметров, то проблема состоит лишь в том, чтобы выбрать наименьшее число наиболее представительных значений в качестве уровней. Куда более сложным представляется случай неизмеримых, качественных факторов. Для их дискретизации необходимы неформальные процедуры, основанные на глубоком знании особенностей объекта. Приходится применять методы **ранжирования** и **экспертных оценок**.

1 Для многофакторных экспериментов большое число уровней может стать непреодолимым препятствием не только при планировании, но и при постановке отбора и обработке данных. Само по себе большое число факторов увеличивает лишь размерность задачи и усложняет ее трактовку. Хуже обстоит дело при многоуровневых факторах – план и эксперимент в этом случае теряют наглядность. Поэтому на практике стараются, прежде всего, уменьшить именно число уровней. Конечно, это не является самоцелью.

Объем эксперимента определяется k -произведением $\prod_{j=1}^k h_j$, где k – число факторов; h_j – число уровней каждого фактора.

2 Еще одной серьезной проблемой при планировании эксперимента является выполнение требований дисперсионного анализа, особенно в части *независимости* исследуемых факторов. На практике изучение сложных объектов обычно происходит в режиме пассивного наблюдения. При этом выбор изучаемых факторов часто бывает ограничен только теми, которые легко регистрировать. В этих условиях априори гарантировать независимость наблюдаемых параметров *невозможно*. Решению этой проблемы помогает *метод главных компонент*, который позволяет выбрать исследуемые факторы на основе их *некоррелированности*. Поскольку для дисперсионного анализа непосредственное значение имеет именно отсутствие корреляции между исследуемыми параметрами, то последний факт можно использовать вместо громоздких и неформальных исследований их независимости. В конечном итоге можно даже удовлетвориться ненулевой, но относительно *слабой корреляцией*, оценку которой следует делать для каждой конкретной задачи отдельно.

Основная идея метода главных компонент состоит в замене переменных,

характеризующих исследуемые факторы. Пусть исходная система из k наблюдаемых входных факторов x_j ($j = 1, 2, \dots, k$) построена по итогам N наблюдений и представляется матрицей результатов эксперимента:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{21} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NK} \end{bmatrix},$$

в которой столбцами являются наблюдаемые значения входных факторов. По этим данным можно рассчитать корреляционную матрицу $R = (r_{ij})$, составленную из выборочных оценок коэффициентов корреляции:

$$r_{ij} = \frac{1}{s_i s_j} \frac{1}{N-1} \sum_{n=1}^N (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j),$$

где s_i – выборочная оценка среднего квадратического отклонения i -го фактора. Заметим, что диагональные элементы матрицы R равны единице, а сама матрица – симметричная:

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{1k} \\ \dots & \dots & \dots & \dots \\ r_{N1} & r_{N2} & \dots & 1 \end{bmatrix},$$

Замена переменных $z_i = \sum_{j=1}^k a_{ij} x_j$ подбирается таким образом, чтобы но-

вые переменные z_i были некоррелированными. Для этого достаточно, чтобы матрица ковариаций новых переменных стала диагональной. Для того, чтобы выбор шкал измерения факторов не привносил в расчеты своих масштабов, следует максимально унифицировать системы отсчета и нормировать переменные.

Математический метод решения такой задачи известен и реализуется с помощью стандартного программного обеспечения. Из уравнения $\det(R - \nu E) = 0$, где E – единичная матрица, находят собственные значения матрицы R : $\nu = s^2(z_i)$, которые позволяют вычислить собственные векторы a_i той же матрицы. В свою очередь собственные векторы определяют матрицу искомого преобразования, как матрицу, составленную из векторов a_i .

3 При планировании эксперимента немаловажное значение имеет удачный выбор факторов, который позволил бы уменьшить его объем. Задачу выбора наименьшего числа наиболее представительных (т.е., как мы теперь знаем, независимых – некоррелированных) факторов помогает

решить факторный анализ. количества независимых (некоррелированных) факторов.

На практике, как правило, просто подбирают наиболее приемлемый набор "высших" факторов. Конечно, выбор здесь неоднозначен, поэтому пытаются отыскать наиболее легко интерпретируемые, комплексные факторы. Собственно *факторным анализом* и является такой поиск.

4 Существует еще один способ выявления наиболее значимых факторов изучаемого процесса – *планирование отсеивающих экспериментов*, состоящее в постепенном отсеивании несущественных факторов из первоначально выбранного большого их числа. Наиболее распространенным методом для этого является *метод случайного баланса*. Задача отсеивающего эксперимента состоит в том, чтобы провести ранжирование всех факторов.

7.2.2 Методы экспертных оценок

Методы экспертных оценок занимают в планировании эксперимента особое место благодаря тому, что их можно применять на любой стадии исследовательской работы и в самых разнообразных целях. Здесь следует упомянуть лишь самые распространенные вопросы, которые решаются с их помощью:

- формулировка постановки задачи исследований;
- выбор основных факторов;
- выбор наблюдаемых и регистрируемых параметров;
- структурирование сложных систем;
- ранжирование критериев и целей деятельности отдельных элементов сложных систем и т.д.

Группы экспертов (от 1 до 30 специалистов) создаются с учетом их квалификации по специально разработанным методикам.

Основным требованием всех методов экспертных оценок является строгая формулировка вопросов экспертизы, содержащих единообразно понимаемые и единичные (не распространенные или комплексные) темы.

Ниже, без дополнительных рассуждений, приводится краткий обзор основных аспектов теории экспертных оценок.

Этапы экспертизы:

- формирование множества оцениваемых объектов (вопросов);
- формирование экспертной комиссии;
- выбор процедуры экспертного оценивания (выбор метода);
- обработка результатов экспертного оценивания;
- анализ результатов экспертизы и принятие решения.

Классификация методов экспертизы:

- по количеству экспертов: индивидуальные, коллективные (следует учитывать, что при количестве экспертов менее 7 информация характеризуется неполнотой, поэтому более целесообразно 10 и больше экспертов);

– по методу опроса (возможны сочетания методов): индивидуальный, очный (интервью) с обратной связью или без нее, открытый (ответ в произвольной форме), закрытый (ответ типа да/нет или выбор ответа из предложенного набора), прямой (до экспертов доводится цель экспертизы), косвенный (до экспертов не доводится цель экспертизы);

– по информированности экспертов: с заданным множеством объектов оценивания, с неопределенным множеством объектов оценивания;

– по организации процедуры: простые, сложные (с декомпозицией проблемы на простые вопросы).

Методы проведения экспертизы (формы организации экспертизы).

Метод мозговой атаки (метод мозгового штурма) – коллективный открытый метод, при котором на поставленную цель экспертизы эксперты отвечают генерацией произвольных идей: без критики, без аргументации, может быть, с развитием или с фантазией. Далее идеи лаконизируются; не соответствующие цели, отсеиваются, систематизируются, рассматриваются прямо противоположные; оценивается возможность реализации. В заключение, после отбора руководителями экспертизы нескольких идей, они могут подвергаться всей группой экспертов конструктивной критике по *конкретным* позициям: логичность, этичность, эстетичность, экологичность, социальность и т.п.

Метод дискуссии – коллективный открытый метод, при котором во время демократического обсуждения заданной тематики происходит генерация идей, их формулировка и обобщение.

Метод комиссий – коллективный открытый метод, при котором предусматривается предварительный отбор состава экспертов и разбиение их на комиссии по направлениям, предоставление всей требуемой информации и в заключение – коллегиальная оценка по принятой шкале.

Метод суда – метод, близкий к методу комиссий, но состав отбирается не административным образом, а с обеспечением противоположных мнений и интересов, и информация предполагается субъективной.

Метод деловых совещаний – коллективный открытый метод, при котором ведущий формулирует проблему, участники высказывают дополнительную информацию, альтернативы, идеи и оценки, а окончательный выбор решения или оценки делает и оглашает ведущий.

Анкетирование и интервьюирование – метод, который может быть любым по количеству экспертов и по методу опроса, предъявляет особые требования к лаконичности и однозначности вопросов и ответов, проводится в письменной или устной форме чаще всего независимым опросом экспертов.

Метод Дельфи – анонимное анкетирование в несколько этапов (2–3) с обратной связью, когда всем экспертам сообщаются результаты обработки (но не частные мнения) ответов предыдущего этапа с аргументацией крайних суждений.

Метод решающих матриц – предполагает декомпозицию сложной проблемы на несколько уровней, в которых возможны альтернативные решения, связанные с решениями на соседних уровнях; эксперты строят в конечном итоге матрицу оценок, сумма которых равна 1.

Перечисленные методы могут комбинироваться друг с другом.

Обработка результатов экспертизы может быть проведена по следующей схеме, отдельные элементы которой могут не использоваться.

1 Показатели обобщенного мнения:

1.1 Средняя балльная оценка $M_j = \frac{1}{m_j} \sum_{i=1}^{m_j} C_{ij}$, где m_j – число экспертов, оценивших j -й вопрос; C_{ij} – оценка по балльной шкале $0 \leq C \leq B$ (где B – наивысшая балльная оценка в принятой шкале), данная i -м экспертом j -му вопросу ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$).

1.2 Частота k -й оценки j -го опроса $K_j^k = \frac{m_j^k}{m_j}$.

1.3 Сумма рангов j -го вопроса $S_j = \sum_{i=1}^{m_j} R_{ij}$, где R_{ij} – ранг.

2 Показатели степени согласованности экспертов:

2.1 Коэффициент вариации мнений определенному j -му вопросу $V_j = \frac{\sigma_j}{M_j}$, где $\sigma_j = \sqrt{\frac{1}{m_j - 1} \sum_{i=1}^{m_j} (C_{ij} - M_j)^2}$; согласованность тем выше, чем меньше V_j .

2.2 Коэффициент конкордации $W = \frac{12}{m^2(n^3 - n) - m \sum_{i=1}^n T_i} \sum_{j=1}^n d_j^2$, где $d_j = s_j - \frac{1}{n} \sum_{j=1}^n s_j$; $T_i = \sum_{l=1}^L d_l^2 (t_l^2 - t_l)$, t_l – кратность l -го ранга; полная согласованность экспертов – при $W = 1$, согласованности нет при $W = 0$.

3 Статистическая значимость показателя согласованности W по критерию Пирсона: $\chi_n^2 = \frac{1}{mn(n+1) - \frac{1}{n} \sum_{i=1}^m T_i} \sum_{j=1}^n d_j^2$ с $n - 1$ степенями свободы.

4 Коэффициент активности при ответе на j -й вопрос: $K_{Эj}^k = \frac{m_j}{m_j}$.

5 Учет компетентности экспертов может быть проведен с помощью следующих оценок.

5.1 Самооценка по балльной или шкальной системе таких объективных

показателей, как уровень, широта и стаж специализации по вопросам экспертизы и смежным. Самооценка производится по специальным таблицам, не допускающим произвола и задающим каждому уровню фактора свою балльную оценку, например, по балльной шкале из таблицы 7.4 [2] с вычислением коэффициента осведомленности $K_{oc} = B / 10$.

Таблица 7.4 – Градация самооценки по балльной системе

| Градация | Балл |
|--|---------|
| Не знаком с вопросом | 0 |
| Плохо знаком, но вопрос входит в сферу знаний и деятельности | 1, 2, 3 |
| Удовлетворительно знаком, тесно связан со смежной узкой специализацией | 4, 5, 6 |
| Хорошо знаком, принимает практическое участие | 7, 8, 9 |
| Вопрос входит в узкую специализацию | 10 |

5.2 Самообъективная оценка вычисляется, как сумма оценок степени влияния всех видов источников информации по шкале, приведенной в таблице 7.5 (максимальное значение равно 1).

Таблица 7.5 – Самообъективная оценка по степени влияния всех видов источников информации

| Источники информации | Степень влияния источника | | |
|---|---------------------------|---------|--------|
| | высокая | средняя | низкая |
| Производственный опыт | 0,50 | 0,40 | 0,20 |
| Проведенный теоретический анализ | 0,30 | 0,20 | 0,10 |
| Обобщение отечественной литературы | 0,05 | 0,04 | 0,03 |
| Обобщение зарубежной литературы | 0,05 | 0,04 | 0,03 |
| Личное знакомство с состоянием вопроса за рубежом | 0,05 | 0,04 | 0,03 |
| Интуиция | 0,05 | 0,04 | 0,03 |

5.3 Коэффициент компетентности i -го эксперта ($i = 1, 2, \dots, m$) по j -му вопросу ($j = 1, 2, \dots, n$) может быть вычислен по формуле $k_{ij} = \sum_{v=1}^s \theta_{vi} R_{vj} / \sum_{v=1}^s R_{vj}$,

составляющие которой определяются следующей процедурой. Единым образом (руководителем или каждым экспертом для себя, или взаимно) для каждого i -го эксперта определяются уровни осведомленности $0 < \theta_{vi} < 10$ по каждой v -й из s специализаций (дисциплин, сфер деятельности), к которым можно отнести все вопросы экспертизы. Каждый j -й вопрос экспертизы получает оценку $0 \leq R_{vj} \leq 2$ по уровню важности для него каждой v -й специализации с помощью двухбалльной непрерывной шкалы: несущественно – 0; полезно – 2; весьма существенно – 2.

8 ОСНОВЫ ОПТИМИЗАЦИИ ВАРИАНТОВ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

8.1 Методы векторной оптимизации сложных систем

При решении большинства задач проектного моделирования сложных систем (СЛС) возникает необходимость их оптимизации по совокупности откликов вероятностных и имитационных моделей, оценивая при этом по различным критериям эффективность вариантов их функционирования [19]. Такая оптимизация получила название векторной или многокритериальной. Ее особенностью является наличие не одного решения, а целого множества решений, каждое из которых может быть выбрано в качестве оптимального. В общем виде задачи векторной оптимизации СЛС могут быть записаны следующим образом:

$$Y_i = f_i(x) \rightarrow \max; \quad i = \overline{1, k}; \quad Y_j = f_j(x) \rightarrow \min; \quad j = \overline{k+1, m}, \quad x \in G.$$

Выделение главного критерия. Выделяется главный критерий. Например, пусть $Y_1 = f_1(x) \rightarrow \max$, тогда

$$Y_i = f_i(x) \geq \hat{Y}_i^*; \quad i = \overline{2, k}; \quad Y_j = f_j(x) \leq \hat{Y}_j^*; \quad j = \overline{k+1, m}, \quad x \in G.$$

Как видим, для остальных критериев качества вариантов СЛС стремятся, чтобы они были либо больше, либо меньше некоторого граничного значения (соответственно \hat{Y}_i^* и \hat{Y}_j^*). Часть контроля качества вариантов СЛС осуществляется по главному критерию, но при условии, что по всем остальным критериям качества отклика СЛС находится в допустимых пределах.

Метод последовательных уступок. Все отклики вариантов СЛС можно расположить в порядке уменьшения их важности. Например, пусть важность откликов расположена следующим образом: Y_1, Y_2, \dots, Y_n и пусть для всех Y_i необходимо добиться их максимизации из возможных вариантов. Алгоритм метода сводится к следующему. Вначале находится решение, обращающее в максимум главный критерий $\hat{Y}_1^* = f_1^*(x) \rightarrow \max$. Затем из априорных соображений назначается некоторая «уступка» Δf_1 , требующая выполнения неравенства $f_1(x_1) \geq f_1^*(x) - \Delta f_1$, где $f_1^*(x) = \max f_1(x)$ и находим такое решение x_1 , при котором $f_2(x) \rightarrow \max$.

Далее снова назначается «уступка» по критерию $Y_2(x) = f_2(x)$. Например, пусть это будет Δf_2 , с помощью которой можно максимизировать $f_2^*(x) = \max f_2(x)$. Этот процесс продолжается до тех пор, пока по всем откликам не будет достигнута некоторая близость к максимальным значениям с учетом сделанных уступок: $\Delta f_i; \quad i = \overline{1, n}$.

Методы приведения откликов к одному типу и масштабу. В ряде случаев исследователю удобно, чтобы все отклики были однотипны (требовали либо максимизации, либо минимизации). На практике очень часты случаи, когда какой-либо отклик, например $Y_2(x)$, требует минимизации ($Y_2(x) \rightarrow \min$), а все остальные отклики требуют максимизации ($Y_i(x) \rightarrow \max$; $i \neq 2$; $i = 1, 3, \dots, n$). В таких случаях для откликов, требующих минимизации, находят обратную величину: $Y_2^*(x) = 1/Y_2(x)$.

Однако этой операции может оказаться недостаточно, поскольку каждый отклик имеет различный смысл и его значение определяется с различной размерностью (килограммы, метры, секунды, миллиметры, метры). Поэтому отклики трудно сравнивать друг с другом даже в случае, когда они одного типа. В таких случаях используется операция «нормировка». Находят для каждого отклика максимальное значение:

$$Y_{mi} = \max_i Y_i \quad \text{или} \quad Y_{mj}^* = \max_j Y_j^*; \quad i, j = \overline{1, n}.$$

Проводится нормировка каждого отклика:

$$Y_i^{**} = Y_i / Y_{mi} \quad \text{— для откликов, требующих максимизации;}$$

$Y_j^{**} = Y_j^* / Y_{mj}^*$ — для откликов, полученных с помощью операции приведения к одному типу.

В итоге все компоненты вектора параметров $\{Y_i^{**}\}$; $i = \overline{1, n}$ будут иметь одинаковый тип (в данном случае требовать максимизации) и изменяться в одном и том же диапазоне их значений $[0, 1]$ (нормированных максимальным значением). Поэтому их можно сравнивать.

Метод составного критерия. Когда все отклики имеют один и тот же масштаб изменения, и они все одного типа, исследователь может формировать составной отклик $Y_0(x)$ согласно следующей процедуре. Исследователь сам определяет важность каждого критерия с помощью весовых коэффициентов важности (δ_i). Предполагается, что $\delta_i \geq 0$ $i = \overline{1, n}$, а сумма коэффициентов важности равна единице: $\delta_i \geq 0$; $\sum_{i=1}^n \delta_i = 1$. Определяется составной

(обобщенный) отклик варианта СС: $Y_0(x) = \sum_{i=1}^n \delta_i Y_i(x)$. Если $Y_i(x) \rightarrow \max$, то

и $Y_0(x) \rightarrow \max$.

Несмотря на удобную форму записи, обобщенные критерии качества вариантов СЛС имеют существенные недостатки, связанные с субъективностью выбора весовых коэффициентов важности откликов (δ_i), а также с тем фактом, что недостатки эффективности вариантов СЛС по одним критериям могут компенсироваться за счет преимуществ по другим критериям.

Нормативные методы. Они являются обобщением рассмотренных выше методов и состоят в предварительном получении нормативов $nr f_i$, $i = \overline{1, n}$ на основе приближенного решения многоцелевой задачи и приближения к этим нормативам по некоторой заранее заданной метрике $\rho(f(x), nr f) \rightarrow \min$, где весовые коэффициенты определяются различными способностями. Например, в качестве метрики выбираются:

$$\rho_1^2(f(x), nr f) = \sum_{i=1}^n [f_i(x) - nr f_i]^2; \quad \rho_2(f(x), nr f) = \sum_{i=1}^n |f_i(x) - nr f_i|;$$

$$\rho_3(f(x), nr f) = \max_i |f_i(x) - nr f_i|.$$

Выбор метрики исследователь осуществляет из априорной информации об исследуемой сложной системе.

Методы логического объединения критериев. Используются только в тех случаях, когда отклики $Y_i = f_i(x)$, $i = \overline{1, n}$ могут принимать только два значения – 0 или 1:

$$Y_i(x) = f_i(x) = \begin{cases} 1, & \text{если } i\text{-я цель исследования достигнута,} \\ 0, & \text{если } i\text{-я цель исследования не достигнута.} \end{cases}$$

В таких случаях используются два вида обобщенных откликов:

– в виде конъюнкции откликов $Y_i(x)$, если общая цель состоит в достижении всех целей одновременно:

$$Y_{01}(x) = \prod_{i=1}^n f_i(x) = \prod_{i=1}^n Y_i(x);$$

– в виде дизъюнкции откликов $Y_i(x)$, если общая цель состоит в достижении хотя бы одной из целей:

$$Y_{02}(x) = 1 - \prod_{i=1}^n f_i(x) = 1 - \prod_{i=1}^n Y_i(x).$$

Во всех методах параметр X выступает в качестве обобщенной характеристики. На практике X может, в свою очередь, быть многомерный величиной. К ней также могут быть применены способы отображения ее в виде скалярной величины, путем приведения к одному типу и нормировки значений ее компонентов, приводящей к тому, что у X все компоненты изменяется на одном и том же интервале значений $[0, 1]$, и используют в этом случае составной параметр X_0 . Способы такой модификации параметра X к обобщенной величине можно найти в работах [20, 21]. Оставив в стороне методу приведения многомерного вектора параметров $\{X_j\}, \dots, j = \overline{1, m}$ к обобщенному скаляру, вначале рассмотрим методики оптимизации целевой функции моделирования от одной переменной X .

8.2 Оптимизация нелинейной функции отклика от одной переменной

Рассмотрим задачу максимизации различных видов целевых функций. В случае максимизации линейной целевой функции (ЦФ) одной переменной имеем $C(x) = C_0 + C_1 x$, график который приведен на рисунке 8.1, а. Если на x не налагается никаких ограничений, то при любом $C_1 \neq 0$ значение ЦФ можно сделать сколь угодно большим. Если же $a_1 \leq x \leq a_2$, то $x = a_1$ оптимально при $c_1 \leq 0$, а при $c_1 \geq 0$ оптимальным будет $x = a_2$. Следовательно, оптимальным является решение, соответствующее экстремальной точке. Однако, если функция $C(x)$ нелинейна, то возможны другие ситуации, изображенные на рисунке 8.1, б. Предположим, что:

- 1) значения x лежат в пределах конечного сегмента $I \equiv [a_1, a_2]$ т. е. $a_1 \leq x \leq a_2$, где $-\infty < a_1 < a_2 < +\infty$;
- 2) функция $C(x)$ непрерывна и ограничена сверху для всех x на сегменте.

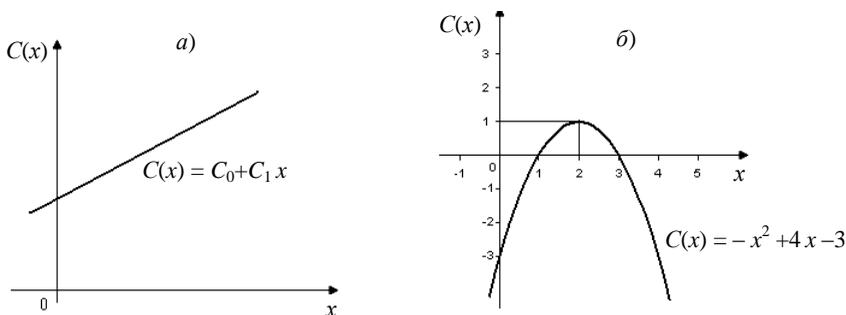


Рисунок 8.1 – Примеры нелинейных функций отклика

Из этих допущений следует, что $C(x)$ достигает максимума при значении x , находящемся на сегменте I . Рассмотрим случай, когда функция $C(x)$ является монотонной на сегменте I , и всегда достигается максимум x на I в одной из экстремальных точек сегмента. Если $C(x)$ является выпуклой функцией x на I , т.е. для любых x_1 и x_2 на I , где $x_1 < x_2$, и для любого $p (0 \leq p \leq 1)$ функция $C(x)$ удовлетворяет условию

$$p C(x_1) + (1 - p) C(x_2) \geq C [p x_1 + (1 - p)x_2] - \text{выпуклая.}$$

Рассмотрим функцию $C(x)$, график который изображен на рисунке 8.2. Максимальная значение $C(x)$ достигается при $x = f$. Предположим, что проверка значения $C(x)$ начинается со значения $x = a$. Если первое проверяемое значение x меньше a , тогда каждое проверяемое значение x все больше и больше будет приближаться к a , но не будет превышать a , поскольку $C(a)$ есть *локальный максимум*.

Далее предположим, что проверяемое значение x находится на сегменте $[d, e]$. Тогда, по-видимому, последовательный поиск закончится сразу же, поскольку как при небольшом увеличении, так и при небольшом уменьшении значения x функция $C(x)$ не меняет своей величины и ее улучшение не представляется возможным. Как видно из рисунка 8.2, рассмотрение функции $C(x)$ на сегменте $[0, g]$ является более общей ситуацией, чем это действительно необходимо для достижения задачи максимизации. График $C(x)$ можно разделить на 5 различных участков, где функция является одномеральной. При этом на каждом j участке ($j = \overline{1,5}$) модами являются соответственно: $\text{Mod}_1=C(a)$; $\text{mod}_2=C(c)$; $\text{mod}_3=C(d)$; $\text{mod}_4=C(f)$; $\text{mod}_5=C(d)$.

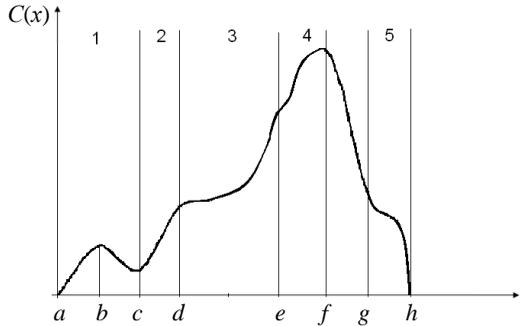


Рисунок 8.2 – Функция отклика, обладающая многими локальными оптимумами

Экстремумом функции $C(x)$ будет мода mod_4 . На практике исследователи стараются поделить область изменения функции на ряд унимодальных участков, а затем из значений мод выбирают ту, которая является максимальной на всем участке определения x . В нашем примере экстремум $C(x)$ достигается при $x = f$ и он равен $\text{ekst}C = C[f]$. Обычно исследователи для поиска экстремума унимодальной функции от одной переменной пользуются методами *дихотомии* и *золотого сечения*. Рассмотрим алгоритмы реализации этих методов.

Экстремумом функции $C(x)$ будет мода mod_4 . На практике исследователи стараются поделить область изменения функции на ряд унимодальных участков, а затем из значений мод выбирают ту, которая является максимальной на всем участке определения x . В нашем примере экстремум $C(x)$ достигается при $x = f$ и он равен $\text{ekst}C = C[f]$. Обычно исследователи для поиска экстремума унимодальной функции от одной переменной пользуются методами *дихотомии* и *золотого сечения*. Рассмотрим алгоритмы реализации этих методов.

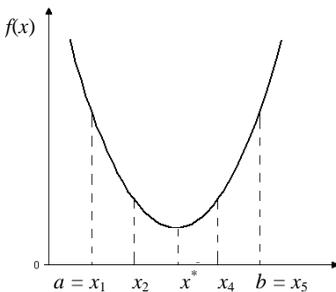


Рисунок 8.3 – Пример функции отклика с одним минимумом

Метод дихотомии. Пусть целевая функция $f(x)$ имеет на отрезке $[a, b]$ минимум в точке x^* (рисунок 8.3). Вычисление абсциссы x^* точки минимума $f(x)$ фактически сводится к вычислению отрезка $[a_n, b_n]$, длина которого $|b_n - a_n| < \varepsilon$, ε – заданная точность вычисления и $x \in [a_n, b_n]$.

Сущность метода дихотомии заключается в следующем:

- 1) полагают $x_1 = a$ 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, $f(x_i) = \min\{f(x_i), i = \overline{1,5}\}$. В силу унимодальности $f(x)$ на отрезке $[a, b]$ точка x^* находится на отрезке $[x_2, x_4]$, длина которого по построению вдвое меньше длины отрезка $[a, b]$.

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

В качестве примера найдем методом дихотомии минимум функции $f(x) = \sin x + \sqrt{x}$ на отрезке $[3, 6]$. Точность ε принять равной 1.

График функции $f(x)$ имеет вид, представленный на рисунке 8.4, а.

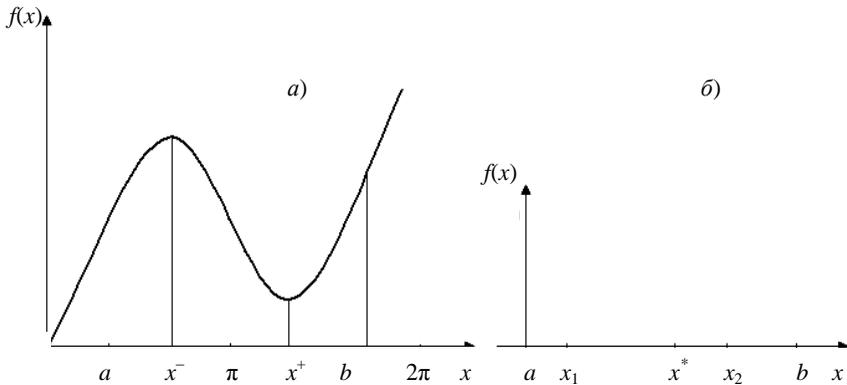


Рисунок 8.4 – Примеры:

а – функции отклика с двумя экстремумами;

б – выбора точек деления отрезка по методу золотого сечения

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 $f(x_i)$, $i = \overline{1,5}$:

$$f(x_1) = 1,87; f(x_2) = 1,36; f(x_3) = 1,14; f(x_4) = 1,43; f(x_5) = 2,17;$$

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 ε , и мы вынуждены повторить все действия,

полагая теперь, что новый отрезок $[a, b]$, на котором находится x^* , равен $[3,75; 5,25]$.

Вторая итерация:

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

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

$$3) \min\{f(x_i), i = \overline{1,5}\} = f(x_3) = 1,14.$$

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 $x^* = (4,88+4,12)/2 = 4,5$,

и $f(x^*) \approx 1,1438$. Безусловно, $\varepsilon = 1$ обеспечивает нам очень грубое решение задачи. На практике, как правило, задают ε 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 находится из квадратного уравнения $x^2 + dx - d^2 = 0$, и следовательно, $x_1 \approx 0,38d$ и $x_2 \approx 0,62d$ (см. рисунок 8.4, б).

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:

$$1) \text{ вычисляем } x_1 = a + 0,38(b-a) \text{ и } x_2 = a - 0,62(b-a);$$

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

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

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

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

3) выбираем отрезок, которому принадлежит точка x^* . Назовем 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 ε $x^* = (b_1 + a_1)/2$.

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

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

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

Решение. 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 ε , поэтому следует выполнить еще 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 ε $x^* = (4,15 + 4,86) / 2 = 4,505$, а значит $f(x^*) = 1,1439$.

8.3 Максимизация функции отклика от многих переменных

Вычислительный алгоритм максимизации функции отклика. В тех случаях, когда вычислительный алгоритм максимизации функции отклика $C(x_i)$ зависит от значений компонентов вектора параметров (x_1, x_2, \dots, x_n) , исследователю приходится постулировать, что функция $C(x_i)$ $i = \overline{1, m}$ достигает конечного максимума при конечных значениях $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$. Сокращенно обозначим: $\{x_i\}$ – через X , $\{\hat{x}_i\}$ – через \hat{X} , $C(x_i)$ – через $C(X)$. Кроме того, исследователь должен предположить следующие допущения:

1) при всех значениях $\{x_i\}$ функция $C(X)$ однозначна и конечна;

2) при всех значениях X каждая частная производная $\frac{\partial C(X)}{\partial X_j}$ однознач-

на, конечна и непрерывна, так что функция $C(X)$ также непрерывна;

3) функция $C(X)$ обладает конечным максимумом \hat{C} ;

4) при возможных значениях $C(X)$, например C , существует такое конечное число M_C , что $\{x_j\} \leq M_C$, если $C(X) \geq C$.

Из теории дифференциального исчисления известно, что необходимым условием существования максимума является выполнение первых трех допущений. Это означает, что функция $C(X)$ имеет максимум в точке \hat{x} только в том случае, если $\frac{\partial C(X)}{\partial X_j} = 0$ для всех $j = 1, 2, \dots, n$. Многие вычислитель-

ные методы максимизации $C(X)$ имеют следующую стандартную форму:

Шаг 1. Выбирают качестве исходной произвольную пробную точку X^0 .

Шаг 2. Расчеты прекращают, если в пробной точке X^k частные производные равны нулю, иначе определяют значение Y_j^k для $j = 1, \dots, n$ и переходят к шагу 3 (где k – номер пробной точки).

Шаг 3. Определяют новую пробную точку $X_j^{k+1} = X_j^k + Y_j^k$ и переходят к шагу 1.

Возвращаются к выполнению шага 2, заменив X^k на X^{k+1} . При этом определяют направление d_j^n для $j = 1, \dots, n$ на основе информации о поведении $C(X)$ вблизи проверяемой точки X^n , т. е. на локальных свойствах изменения $C(X)$. Выбор длины шага t^h основан на информации о поведении $C(X)$ за пределами окрестности точки X^n при движении в выбранном направлении $d_1^k, d_2^k, \dots, d_n^k$. Наконец, направление и длина шага объединяются в величины: $Y_j^k = t^k d_j^k; \quad j = 1, 2, \dots, n$.

Рассмотрим следующий **пример**.

Нужно максимизировать функцию $C(X_1, X_2) = -(X_1 - 3)^2 - 4(X_2 - 2)^2$.

Исследуя $C(X)$, можно установить, что единственным оптимальным решением является $\hat{X}_1 = 3; \hat{X}_2 = 2$, при которых $C(\hat{X}) = 0$. Поскольку $C(X)$ является квадратичной функцией, оптимальное значение X можно вычислить непосредственно из необходимых условий $\frac{\partial C(X)}{\partial X_j} = 0; \quad i = \overline{1, 2}$. На рисунке 8.5 приведен график линий

уровня $C(X_1, X_2)$. Начало алгоритма осуществляется при $X_1^0 = X_2^0 = 0$ и при $C(X^0) = -25$. Линии уровня имеют вид эллипсов, уравнения для которых

$$\frac{\partial C}{\partial X_1} = -2(X_1 - 3); \quad \frac{\partial C}{\partial X_2} = -8(X_2 - 2).$$

Обе эти производные в точке X^0 строго положительны. Поэтому на шаге 2 направление дальнейшего изменения значений X_1 и X_2 приведено на рисунке 8.6. В процессе подготовки к выбору направления d_1 и d_2 проводится касательная к линиям уровня $C(X)$ в точке X^0 . В итоге имеет место система уравнений

$$\frac{\partial C}{\partial X_1}(X_1 - X_1^0) + \frac{\partial C}{\partial X_2}(X_2 - X_2^0) = 0,$$

При $X_1^0 = X_2^0 = 0$ система уравнений имеет решение

$$6X_1 + 16X_2 = 0, \quad \text{или} \quad X_2 = -\frac{3}{8}X_1.$$

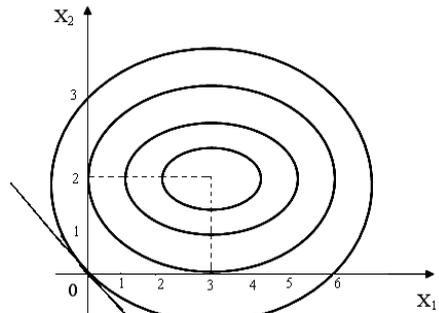


Рисунок 8.5 – Касательная к линиям уровня

Сама же касательная показана рисунке 8.6, а. Направление $d_1 d_2$ изображено в виде вектора, выходящего из исходной точки. Из рисунка 8.6, б видно, что значение $C(X)$ на каждой итерации улучшается. При возрастании номера шага k пробные точки лежат на спирали, все больше и больше приближаясь к окружности с центром в начале координат и с единичным радиусом.

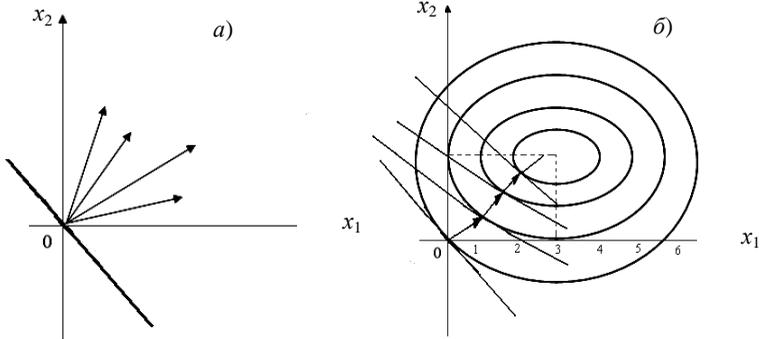


Рисунок 8.6 – Направления движения (а) и оптимальная длина шага (б)

Метод скорейшего подъема. Позволяет учитывать информацию о виде целевой функции при движении от произвольной точки \hat{X} к значению \hat{X}_h , обеспечивая его движение за минимальное число итераций. Приняв в качестве $d_j^k = \frac{\partial C}{\partial X_j}$, в пробной точке X^k величину приращения, равную $t^k \geq 0$, можно максимизировать целевую функцию путем добавления приращений на каждом k -м шаге оптимизации: $C(X_1^k + t^k d_1^k, \dots, X_n^k + t^k d_n^k)$.

В методе применена идея использования в качестве направления набор частных производных, обычно именуемый градиентом функции $C(X)$ в точке X^k , обозначаемый символом $\nabla C(X)$:

$$\nabla C(x) = \text{grad } C(X) = \left(\frac{\partial C(X)}{\partial X_1}, \frac{\partial C(X)}{\partial X_2}, \dots, \frac{\partial C(X)}{\partial X_n} \right)$$

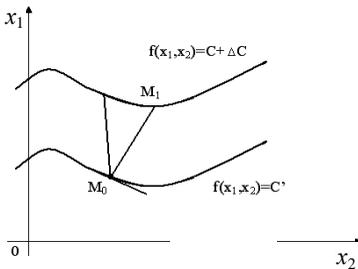


Рисунок 8.7 – Линии уровня функции двух переменных

Вектор $[\text{grad } C(X)]$ называется анти-градиентом функции $C(X)$. Он указывает направление наискорейшего убывания функции $C(X)$.

Для простоты рассмотрим функцию двух переменных $C(X_1, X_2)$. На рисунке 8.7 изображены две линии уровня:

$$C(X_1, X_2) = S; \quad C(X_1, X_2) = S + \Delta S.$$

Как известно, линии уровня функции $C(X)$ – это множество точек, в которых данная функция $C(X_1, X_2)$ принимает одно и то же фиксированное значение. Рассмотрим какую-либо точку $M_0(X_1^0, X_2^0)$ на линии $C(X_1, X_2) = S$ и точку $M_1(X_1^1, X_2^1)$ на линии $C(X_1, X_2) = S + \Delta S$. Приращение функции ΔC при переходе из точки M_0 в точку M_1 равно $\Delta C = \Delta S$, и это приращение не зависит от положения точки M_1 на линии $C(X_1, X_2) = S + \Delta S$. Отношение приращения функции к длине отрезка M_0M_1 назовем средней скоростью изменения функции $C(X_1, X_2)$ на отрезке M_0M_1 . Очевидно, что средняя скорость ее изменения $\Delta C / |M_0M_1|$ зависит от длины отрезка M_0M_1 или от положения точки M_1 на линии $C(X_1, X_2) = S + \Delta S$. При этом скорость $\Delta C / |M_0M_1|$ достигает наибольшего значения там, где длина отрезка M_0M_1 , будет наименьшей. Известно, что длина отрезка наименьшая в том случае, когда направление отрезка M_0M_1 перпендикулярно касательной $C(X_1, X_2) = S$ в точке M_0 при $\Delta C \rightarrow 0$. Теперь, если рассматривать предел отношения $\lim \frac{\Delta C}{|M_0M_1|}$, то это означает опре-

деление производной. Итак, частные производные $\frac{\partial C(M_0)}{\partial X_1}$ и $\frac{\partial C(M_0)}{\partial Y_1}$ определяют максимальные скорости изменения $C(X_1, X_2) = C$ в точке M_0 по координатным осям.

Как видим, идея градиентного метода реализуется следующей последовательностью шагов при поиске экстремума функции $C(X)$.

Шаг 1. Выбирают начальную точку M_0 , принадлежащую области определения функции $C(X)$.

Шаг 2. Находят направление градиента функции $C(X)$ в точке M_0 :

$$C(M_0) = \left(\frac{\partial C(M_0)}{\partial X_1}, \frac{\partial C(M_0)}{\partial X_2}, \dots, \frac{\partial C(M_0)}{\partial X_n} \right).$$

Шаг 3. Задаем приращение из точки M_0 , на величину равную α , в точку M_1 . В результате получаем расчетные формулы:

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

Когда ищем минимум функции, необходимо перемещаться в направлении антиградиента, и в этом случае $\alpha < 0$. Если же ищем максимум функции $C(X)$, то $\alpha > 0$. Признаком достижения минимума (максимума) функции служит обращение в нуль градиента $grad C = 0$. Но на практике, как правило, ограничиваются приближенным решением. Поэтому критерием дости-

жения требуемого результата является выполнение условия

$$\sum_{j=1}^n \left(\frac{\partial C(X)}{\partial X_j} \right)^2 \leq \varepsilon, \text{ где } \varepsilon > 0 \text{ – наперед заданное число.}$$

Существуют и другие критерии. Например, после каждой итерации можно сравнивать достигнутое значение целевой функции с ее значением в предыдущей точке. Если выполняется неравенство $|C(M_n) - C(M_{n-1})| \leq \varepsilon$, то поиск экстремума функции $C(X)$ прекращается. На рисунках 8.6 и 8.7 приведен пример поиска экстремума методом скорейшего подъема.

В качестве **примера** рассмотрим поиск экстремума целевой функции закупки сырья двух видов фабрикой. Пусть фабрика закупает сырье в количестве соответственно X_1 и X_2 в тоннах. Расходы на хранение сырья на складе определяются по функциям: $C(X_1, X_2) = X_1^2 - 7X_1 + X_2^2 - 4X_2 - X_1X_2 + 35$ (у. е.). Необходимо определить объемы сырья (X_1^* и X_2^*), расходы на хранения которых будут минимальны. Примем точность вычислений, равную $\varepsilon = 1$. В качестве начальной M_0 выберем точку с координатами $X_1 = 1$; $X_2 = 1$, причем шаг установим равным $\alpha = 0,5$. Расходы на хранение сырья в количестве $X_1=1$; $X_2=1$ равны $C(M_0) = 25$. Вычислим значения компонент градиента.

$$\frac{\partial C(M_0)}{\partial X_1} = -2X_1 - X_2 - 2; \frac{\partial C(M_0)}{\partial X_2} = 2X_2 - X_1 - 4; \frac{\partial C'(M_0)}{\partial X_1} = -6; \frac{\partial C'(M_0)}{\partial X_2} = -3.$$

Переходим к точке $M_1 (X_1^1, X_2^1)$. Уравнение параметрических прямых имеет вид $X_1^1 = 1 - 0,5(-0,6) = 4$; $X_2^1 = 1 - 0,5 \cdot (-0,3) = 2,5$, откуда

$$\frac{\partial C(M_1)}{\partial X_1} = -1,5; \frac{\partial f(M_1)}{\partial X_2} = -3; C(M_1) = 9,25.$$

Разность $|C(M_1) - C(M_0)| = 15,75 > \varepsilon$, поэтому выполняем вторую итерацию – вычисляем значения целевой функции в точке $M_1 (X_1^2, X_2^2)$:

$$X_1^2 = 4 - 0,5 \cdot (-1,5) = 4,75; \quad X_2^2 = 2,5 - 0,5 \cdot (-0,3) = 4;$$

$$\frac{\partial C(M_2)}{\partial X_1} = -1,5; \quad \frac{\partial C(M_2)}{\partial X_2} = -0,75; \quad C(M_2) = 5,31.$$

Поскольку $|C(M_2) - C(M_1)| = 3,94 > \varepsilon$, необходимо выполнить третью итерацию:

$$X_1^3 = 4,75 - 0,5 \cdot (-1,5) = 5,5; \quad X_2^3 = 4 - 0,5(-0,75) = 4,4; \quad \frac{\partial C(M_3)}{\partial X_1} = -0,38;$$

$$\frac{\partial C(M_2)}{\partial X_2} = -4,74; \quad C(M_3) = 4,31; \quad |C(M_3) - C(M_2)| = 1 = \varepsilon.$$

Учитывая, что требуемая точность результата достигнута, процесс вычисления издержек прекращается. Таким образом, минимальные издержки на хранение при объемах сырья $X_1 = 5,5$ и $X_2 = 4,4$ составляют $C(M_3) = 4631$ у. е.

8.4 Нахождение экстремума функции отклика от многих переменных при наличии ограничений

Задача поиска экстремума функции отклика от многих переменных при наличии ограничений усложняется. Например, нужно найти минимум функции $C(X)$, при ограничениях $h_j(X) = \overline{b_j}$; $j = \overline{1, m}$. При этом предполагается выполнение условий применения методов оптимизации: функции $C(X)$ и $h_j(X)$ обладают непрерывными частными производными по аргументам X_1, \dots, X_n . В таких случаях используется метод *неопределенных множителей Лагранжа*. Суть метода состоит в следующем. Составляется вспомогательная функция:

$$L(X_1, X_2, \dots, X_n, \lambda_1, \lambda_2, \dots, \lambda_m) = C(X_1, X_2, \dots, X_n) + \sum_{j=1}^m \lambda_j (b_j - h_j(X_1, X_2, \dots, X_n)),$$

которая называется функцией Лагранжа, а множители λ_j , $j = \overline{1, m}$ называются неопределенными множителями Лагранжа. Обозначим $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ и запишем функцию Лагранжа в более компактном виде:

$$L(X, \Lambda) = C(X) + \sum_{j=1}^m \lambda_j \cdot (b_j - h_j(X))$$

Эту функцию Лагранжа нужно минимизировать. Составим систему уравнений:

$$\frac{\partial L(X, \Lambda)}{\partial X_i} = \frac{\partial C(X)}{\partial X_i} = \sum_{j=1}^m \lambda_j + \frac{\partial h_j(X)}{\partial X_i} = 0; \quad i = \overline{1, n};$$

$$\frac{dL(X, \lambda)}{d\lambda_j} = b_j - h_j(X) = 0; \quad j = \overline{1, m}.$$

Как видим, идея метода неопределенных множителей Лагранжа заключается в том, что исходная задача отыскания условного экстремума целевой функции $C(X)$ сводится к задаче безусловного экстремума функции Лагранжа $L(X, \Lambda)$, которая решается классическим методом – путем приравнивая к нулю частных производных по X_j ; $j = \overline{1, m}$ и λ_j ; $j = \overline{1, m}$.

Отметим, что при применении метода неопределенных множителей Лагранжа также существуют следующие проблемы. Во-первых, для отыскания точек экстремума функции $L(X, \Lambda)$ требуется проводить анализ производных

$$\frac{\partial^2 L(X, \Lambda)}{\partial X_i^2} \text{ и } \frac{\partial^2 L(X, \Lambda)}{\partial \lambda_j^2}.$$

Во-вторых, условия не являются достаточными для существования решений X^* .

Простота составления функции Лагранжа и возможность обобщения задачи оптимизации на случай ограничений-неравенств обусловила широкое применение метода неопределенных множителей Лагранжа в современных исследованиях СС.

Итак, если нужно найти решение задачи методом неопределенных множителей Лагранжа, мы должны выполнить следующие действия:

- 1) составить функцию Лагранжа;
- 2) вычислить частные производные и приравнять их к нулю, получив систему;
- 3) решить систему уравнений любым известным способом, получив тем самым стационарные точки;
- 4) исследовать функцию на наличие экстремума в полученных стационарных точках.

Рассмотрим следующий *пример*. 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:

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

Решение этой системы: $x_1^* = \sqrt{s}$; $x_2^* = \sqrt{s}$; $\lambda = \frac{2}{\sqrt{s}}$.

Более подробно с методикой оптимизации при нелинейной целевой функции откликов математических моделей можно ознакомиться в работе [8].

СПИСОК ЛИТЕРАТУРЫ

- 1 **Кубланов, М. С.** Математическое моделирование. Методология и методы разработки математических моделей механических систем и процессов : учеб. пособие. В 2 ч. Ч. I. Моделирование систем и процессов / М. С. Кубланов. – М. : МГТУ ГА, 2004. – 108 с.
- 2 **Кубланов, М. С.** Математическое моделирование. Методология и методы разработки математических моделей механических систем и процессов : учеб. пособие. В 2 ч. Ч. II. Планирование экспериментов и обработка результатов измерений / М. С. Кубланов. – М. : МГТУ ГА, 2004. – 125 с.
- 3 **Советов Б. Я.** Моделирование систем : учеб. для вузов / Б. Я. Советов, С. А. Яковлев. – 3-е изд., перераб. и доп. – М. : Высш. шк., 2001. – 343 с.
- 4 **Моисеева, Л. Т.** Методы математического моделирования процессов в машиностроении : курс лекций / Л. Т. Моисеева. – Курск, 2008. – 46 с.
- 5 **Понтрягин, Л. С.** Дифференциальные уравнения и их приложения / Л. С. Понтрягин. – М. : Наука, 1988. – 208 с.
- 6 **Тихонов, Э. Е.** Методы прогнозирования в условиях рынка : учеб. пособие / Э. Е. Тихонов. – Невинномысск, 2006. – 221 с.
- 7 **Галушко, В. Н.** Статистический анализ показателей удельного электропотребления троллейбусов от различных факторов / В. Н. Галушко, В. А. Пацкевич, Т. С. Королёнок // Проблемы и перспективы развития транспортных систем и строительного комплекса : III Междунар. науч.-техн. конф.– Гомель : БелГУТ, 2013. – С. 180–181.
- 8 **Папков, Б. В.** Надежность и эффективность электроснабжения : учеб. пособие / Б. В. Папков, Д. Ю. Пашали. – Уфа : Уфимс. гос. авиац. техн. ун-т, 2005. – 380 с.
- 9 **Цисарь, И. Ф.** MATLAB Simulink. Компьютерное моделирование экономики / И. Ф. Цисарь. – М. : Солон-Пресс, 2008. – 256 с.
- 10 **Гандер, В.** Решение задач в научных вычислениях с применением Maple и MATLAB / В. Гандер, И. Гржебичек. – М. : Вассамедина, 2005 г. – 520 с.
- 11 **Шевченко, Д. Н.** Теория вероятностей и математическая статистика : учеб.-метод. пособие для студентов электротехнических специальностей / Д. Н. Шевченко. – Гомель : БелГУТ, 2006. – 318 с.
- 12 **Папков, Б. В.** Надежность и эффективность электроснабжения : учеб. пособие / Б. В. Папков, Д. Ю. Пашали. – Уфа : Уфимс. гос. авиац. техн. ун-т, 2005. – 380 с.
- 13 **Основы теории надежности систем электроснабжения / В. В. Карпов [и др.]. – Омск : Изд-во ОмГТУ, 2003. – 72 с.**
- 14 **Анализ статистических данных на персональном компьютере : лаб. практик. / Т. В. Прищепова [и др.]; под ред. В. С. Серёгиной. – Гомель : БелГУТ, 2006. – 95 с.**
- 15 **Боровиков, В. П.** Популярное введение в программу STATISTICA / В. П. Боровиков. – М. : КомпьютерПресс, 1998. – 69 с.
- 16 **Жученко, Ю. М.** Статистическая обработка информации с применением персональных компьютеров : практик. руководство для студентов 5-го курса / Ю. М. Жученко. – Гомель : ГГУ им. Ф. Скорины, 2007. – 101 с.
- 17 **Барзилович, Е. Ю.** Оптимально управляемые случайные процессы и их приложения (теоретические основы эксплуатации авиационных систем по состоянию) / Е. Ю. Барзилович. – Егорьевск : ЕАТК ГА, 1996. – 299 с.

- 18 **Хикс, Ч. Р.** Основные принципы планирования эксперимента / Ч. Р. Хикс. – М. : Мир, 1967. – 406 с.
- 19 Задачи и модели исследования операций. Ч. 3. Технология имитации на ЭВМ и принятие решений : учеб. пособие / И. В. Максимей [и др.]. – Гомель : БелГУТ, 1999. – 150 с.
- 20 Имитационная модель технологических процессов на городском транспорте / И. В. Максимей [и др.] // Информатика. – 2006. – № 2. – 124–134 с.
- 21 **Федоров, В. В.** Теория оптимального эксперимента / В. В. Федоров.– М. : Наука, 1971. – 270 с.
- 22 **Рыжкин, А. А.** Основы теории надежности : учеб. пособие / А. А. Рыжкин, Б. Н. Слюсарь, К. Г. Шучев. – Ростов н/Д : Издательский центр ДГТУ, 2002. – 182 с.
- 23 **Вентцель, Е. С.** Исследование операций: задачи, принципы, методология / Е. С. Вентцель. – М. : Наука, 1980. – 208 с.
- 24 НИОКР лаборатории «Вычислительная механика» [Электронный ресурс] / Конечно-элементное моделирование лобового столкновения автомобилей ВАЗ-21093 и ЗИЛ-433362. – Гомель, 2015. – Режим доступа: <http://fea.ru/project/115>. – Дата доступа: 4.01.2015.
- 25 EDC [Электронный ресурс] / HVE-CSI Crash Reconstruction Software. – Гомель, 2015. – Режим доступа: <http://www.edccorp.com/promotions/HVE-CSI-April.html>. – Дата доступа: 23.01.2015.

Учебное издание

Галушко Виктор Николаевич

**МАТЕМАТИЧЕСКИЕ МОДЕЛИ
В ТРАНСПОРТНЫХ СИСТЕМАХ**

Учебно-методическое пособие

Редактор И. И. Э в е н т о в

Технический редактор В. Н. К у ч е р о в а

Подписано в печать 20.05.2015 г. Формат 60x84 ¹/₁₆.
Бумага офсетная. Гарнитура Times. Печать на ризографе.
Усл. печ. л. 9,07. Уч.-изд. л. 9,38. Тираж 50 экз.
Зак. № . Изд. № 93.

Издатель и полиграфическое исполнение:
Белорусский государственный университет транспорта.
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий

№ 1/361 от 13.06.2014.

№ 2/104 от 01.04.2014.

Ул. Кирова, 34, 246653 г. Гомель