

ВОЕННО-КОСМИЧЕСКАЯ АКАДЕМИЯ
имени А.Ф. Можайского

МОДЕЛИРОВАНИЕ СИСТЕМ УПРАВЛЕНИЯ

Курс лекций

Часть 2

**ПОСТРОЕНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ
ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ**



Санкт-Петербург
2014

Авторы:
Н.А. Осипов, А.М. Барановский, А.Ю. Цветков, А.Б. Кузнецов

Рецензенты:
кандидат технических наук, доцент
Белозеров В.А.
кандидат технических наук, доцент
Ананенко В.М.

Моделирование систем. В 2 ч. Ч. 2. Построение математических моделей по экспериментальным данным: курс лекций / Н.А. Осипов, А.М. Барановский, А.Ю. Цветков, А.Б. Кузнецов. – СПб.: ВКА имени А.Ф. Можайского, 2013. – 99 с.

Курс лекций соответствует учебной программе дисциплины "Моделирование систем управления" по специальности ФГОС "Специальные организационно-технические системы".

Во второй части курса лекций описываются методы построения математических моделей по экспериментальным данным, рассматриваются модели временных рядов, основные положения регрессионного анализа, а также научно-технические основы идентификации систем.

© ВКА имени А.Ф. Можайского, 2014

Подписано к печ. 20.11.2013
Гарнитура Times New Roman
Уч.-печ. л. 13

Формат печатного листа 448x300/8
Авт. л. 6
Заказ 2393

Бесплатно

Типография ВКА имени А.Ф. Можайского

ОГЛАВЛЕНИЕ

Введение	5
1. Оценивание параметров модели, заданной распределением случайной величины	7
1.1. Метод сравнения	9
1.2. Метод максимального правдоподобия.....	12
1.3. Байесовский метод оценивания	16
2. Обработка массива данных	23
2.1. Первичная обработка исходных данных и построение гистограммы	23
2.2. Расчет статистических параметров распределения	24
2.3. Подбор теоретического распределения и его параметров	25
3. Модели временных рядов.....	31
3.1. Структура временного ряда	31
3.2. Модели случайных процессов	33
3.3. Численные характеристики временных рядов	35
4. Практический анализ временных рядов	38
4.1. Выделение тренда	38
4.2. Выделение сезонных эффектов на фоне тренда	40
4.3. Метод скользящего среднего	40
5. Исследование структуры стационарного временного ряда	43
5.1. Цели и методы анализа	43
5.2. Интерпретация графика коррелограммы	44
6. Регрессионные модели.....	51
6.1. Определение коэффициентов модели по экспериментальным данным... 52	
6.2. Проверка адекватности модели	55
7. Планирование активных факторных экспериментов	59
7.1. Определение коэффициентов модели при ортогональных планах..... 59	
7.2. Дробные факторные эксперименты	64
8. Исследование влияния качественных факторов	66
8.1. Общие сведения о качественных факторах	66
8.2. Проверка гипотезы об отсутствии влияния фактора при неизвестном распределении результирующего фактора.....	68
8.3. Проверка гипотезы об отсутствии влияния фактора при нормальном распределении результирующего признака	71
9. Определение параметров систем	74
9.1. Классификация задач и методов идентификации.....	74

9.2. Определение параметров статической системы по полной выборке	81
9.3. Определение параметров динамической системы по полной выборке....	83
9.4. Последовательная идентификация	89
9.5. Оценивание разброса коэффициентов модели.....	92
9.6. Идентификация параметров динамических систем в режиме принудительных информационных колебаний	93
Список литературы.....	98

ВВЕДЕНИЕ

Экспериментальные исследования, выполняемые в науке и технике, включают в себя как измерительную часть, так и обработку полученных данных с их детальным анализом. Практические знания из области проведения и организации эксперимента, умения и навыки в работе с измерительными приборами, владение аппаратом статистического анализа результатов требуются и в деятельности инженера-практика, и в деятельности инженера-исследователя.

Эмпирические исследования являются основным источником объективной информации о характеристиках процессов, протекающих в реальных объектах, в том числе в автоматизированных системах, средствах и комплексах.

Целью обработки экспериментальных данных (ЭД) является выявление закономерностей в характеристиках исследуемых объектов и процессов. Результаты обработки ЭД позволяют оценить качество объекта, они необходимы для оперативного управления процессами, решения задач адаптации объекта к изменившимся условиям или формирования требований ко вновь создаваемым системам.

Получение экспериментальной информации связано с решением ряда проблем по организации регистрации первичных параметров, их сбора и обработки. Те данные, которые можно непосредственно зарегистрировать, обычно лишь косвенно отражают существенные свойства изучаемого процесса или объекта. Многие показатели качества автоматизированных систем носят случайный характер и по этой причине не могут быть непосредственно измерены. Ряд событий в системах происходит крайне редко, и получить для них достаточный объем эмпирических данных (в частности, получить данные по отказам систем с высокой надежностью) невозможно.

Методы обработки ЭД начали разрабатываться более двух веков тому назад в связи с необходимостью решения практических задач по агробиологии, медицине, экономике, социологии. Полученные при этом результаты составили фундамент такой научной дисциплины, как математическая статистика. В последние 20–30 лет математический аппарат обработки ЭД получил значительное развитие в связи с необходимостью решения принципиально новых задач. К настоящему времени он включает множество различных направлений, которые выходят за пределы классической математической статистики. Многие методы нашли применение при исследовании технических и человеко-машинных систем, а также при обработке результатов имитационного (статистического) моделирования.

Для освоения материала данного курса лекций необходимы начальные сведения по математическому анализу и теории вероятностей. Выполнение практических задач предполагает применение типовых пакетов прикладных программ для проведения расчетов (например, табличного процессора MS Excel, пакета символьной математики MathCAD) или специализированных пакетов обработки статистических данных типа STATISTICA, SPSS.

В написании курса лекций принимали участие: кандидат технических наук, доцент Н. А. Осипов (введение, разделы 1–8), кандидат технических наук, доцент А. Б. Кузнецов (раздел 9), кандидат технических наук, доцент А. М. Барановский (раздел 9), кандидат технических наук А. Ю. Цветков (разделы 3–5).

1. ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛИ, ЗАДАННОЙ РАСПРЕДЕЛЕНИЕМ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Как правило, математические модели сложных систем получают итерационным путем. Например, точность модели системы автоматического управления (САУ), полученной расчетным путем, составляет примерно 50%, поэтому модели уточняются постепенно. Основной информацией для уточнения или построения различного рода моделей являются опытные данные, получаемые в ходе проведения экспериментальных исследований.

Экспериментальные исследования событий и процессов основаны на наблюдениях, в ходе которых регистрируются различные факты искусственного и естественного происхождения.

Источниками экспериментальных данных являются:

- результаты наблюдения за реальными объектами и протекающими в них процессами. Наблюдения могут проводиться в ходе испытаний или в ходе обычной эксплуатации;
- результаты моделирования объектов. В первую очередь к ним следует отнести результаты имитационного моделирования;
- технические, экономические, научные отчеты и обзоры, публикуемые в различных изданиях, например сведения о результатах испытаний или о характеристиках однотипных устройств различных производителей;
- результаты опросов специалистов и другие источники.

Обработка ЭД, получаемых от различных источников, имеет много общего. Однако организация сбора и интерпретации ЭД специфична для конкретной предметной области. В дальнейшем обработка ЭД будет рассматриваться применительно к результатам наблюдения за функционированием автоматизированных систем управления (АСУ), их элементов или моделей.

Использованием экспериментальных данных для уточнения или построения моделей занимаются два направления: математическая статистика и теория автоматического управления. В математической статистике этому посвящены следующие разделы:

- оценивание параметров распределений;
- корреляционный анализ;
- регрессионный анализ;
- дисперсионный анализ;
- анализ временных рядов;

– планирование эксперимента.

В теорию автоматического управления входит теория идентификации, которая предназначена для построения или уточнения моделей САУ по результатам наблюдения за поведением систем.

Распределения случайных величин используются как модели сложных процессов. Например, распределение времени отказа системы рассматривают как модель надежности системы. Очень часто возникают задачи уточнения предполагаемых распределений. Уточнить можно только опытным путем, при котором получают оценки требуемых показателей.

Значение параметра, вычисленное по ограниченному объему ЭД, является случайной величиной, т. е. значение такой величины от выборки к выборке может меняться. Следовательно, в результате обработки ЭД определяется не значение параметра, а только лишь его приближенное значение – статистическая оценка параметра. Получить статистическую оценку параметра теоретического распределения означает найти функцию от имеющихся результатов наблюдения, которая и даст приближенное значение искомого параметра.

Известны и используются два подхода к оцениванию: точечное и интервальное оценивание.

При точечном оценивании находят наилучшую оценку показателя и указывают погрешность оценки (дисперсию).

При интервальном оценивании определяют интервал, который накроет истинное значение параметра с заданной (доверительной) вероятностью.

При точечном оценивании стремятся получить такие оценки, которые будут удовлетворять трём свойствам:

- несмещенности;
- состоятельности;
- эффективности.

Оценка является несмещенной, если ее математическое ожидание равно истинному значению оцениваемого параметра.

Оценка называется состоятельной, если она сходится по вероятности к истинному значению, т.е. стремится к нему, когда объем выборки стремится к бесконечности.

Оценка эффективна, если она имеет минимальную дисперсию из всех оценок.

В данном курсе лекций будут рассмотрены следующие основные методы оценивания:

1. Метод сравнения.
2. Метод максимального правдоподобия.

3. Байесовский метод.

1.1. Метод сравнения

Метод сравнения заключается в приравнивании опытных (выборочных) и теоретических величин, имеющих одинаковый физический смысл (например, число отказов в интервале, вероятность событий, суммарная наработка).

Приравнивание дает уравнение (или систему уравнений), а параметры определяются как решение этих уравнений.

Известно другое название указанного метода – метод моментов, в котором в качестве сравниваемых величин выступают начальные или центральные моменты.

Для случайной величины X теоретический начальный момент k -го порядка:

$$\nu_k = M[x^k] - \text{матожидание от } k\text{-й степени случайной величины.}$$

Если $x_1, x_2 \dots x_N$ – реализация выборки, т.е. опытные (выборочные) значения случайной величины, то выборочный начальный момент k -го порядка

$$\bar{\nu} = \frac{1}{N} \sum_{i=1}^N x_i^k .$$

Теоретический центральный момент K -го порядка

$$\mu_K = M[(x - M[x])^K],$$

Выборочный центральный момент K -го порядка:

$$\bar{\mu}_K = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^K ,$$

где $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$; $N-1$ – число степеней свободы.

Величину $(N-1)$ можно объяснить следующим образом. Было N независимых результатов в выборке, после определения \bar{x} один результат может быть вычислен по другим результатам и по \bar{x} . Один элемент выборки использован на определение \bar{x} и “пропал”. Оценки при делении на $N-1$ должны получиться более точными. Покажем это на примере.

Пример 1.1. Пусть случайная величина X может принимать значения 0, 1, 2 и 3, причем вероятности событий равенства X значениям известны: 0.1, 0.4, 0.4 и 0.1 соответственно.

Решение. Определим математическое ожидание и дисперсию случайной величины:

$$M[\tilde{x}] = 0 \cdot 0,1 + 1 \cdot 0,4 + 2 \cdot 0,4 + 3 \cdot 0,1 = 1,5;$$

$$D[\tilde{x}] = 0,1(0 - 1,5^2) + 0,4(1 - 1,5^2) + 0,4(2 - 1,5^2) + 0,1(3 - 1,5^2) = 0,65.$$

Допустим, что распределение случайной величины неизвестно. В этом случае можно оценить дисперсию (определить приближенное значение – оценку), выбрав X определенное число раз (объем выборки N).

Пусть $N = 2$; $x_1 = 1$; $x_2 = 2$. Тогда оценка математического ожидания определяется как среднее: $\bar{x} = 1,5$.

Оценка дисперсии

$$D[\hat{x}] = \frac{1}{2-1}(2-1,5)^2 + (1-1,5)^2 = 0,5.$$

При делении не на $\frac{1}{N-1}$, а на $\frac{1}{N}$ получили бы 0,25, что дальше от истинного значения дисперсии, которое равно 0,65.

Оценим методом моментов параметры гамма-распределения. Выражение плотности имеет следующий вид:

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

где α и k – параметры; $\Gamma(k)$ – гамма-функция.

Известно, что теоретические начальные моменты первого и второго порядков определяются выражениями $\nu_1 = \frac{k}{\alpha}$, $\nu_2 = \frac{k(k+1)}{\alpha^2}$.

Из опытных значений (выборки) можно получить выборочные начальные моменты $\bar{\nu}_1, \bar{\nu}_2$.

Приравнивая теоретические начальные моменты и выборочные, получим систему из двух уравнений. В результате решения этой системы относительно искомых параметров можно получить выражения для расчета их оценок:

$$k = \frac{\bar{\nu}_1^2}{\bar{\nu}_2 - \bar{\nu}_1^2}, \alpha = \frac{\bar{\nu}_1}{\bar{\nu}_2 - \bar{\nu}_1^2}.$$

Наиболее просто метод моментов выглядит при оценивании самих моментов. Выборочное оценивание моментов и является их оценкой.

Одними из основных характеристик распределения случайной величины являются математическое ожидание и дисперсия. Согласно методу моментов их оценки равны выборочным начальному моменту 1-го порядка и центральному моменту 2-го порядка.

Если $a = M[x]$, $\sigma^2 = M[(x - \bar{x})^2]$, то их оценки соответственно будут определяться следующим образом:

$$\hat{a} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2.$$

Можно показать, что эти оценки не смещены и состоятельны.

Действительно,

$$M[\hat{a}] = \frac{1}{N} \sum_{i=1}^N \underbrace{M[x_i]}_a = \frac{1}{N} N \cdot a = a,$$

следовательно, оценка не смещена;

$$D[\hat{a}] = \frac{1}{N^2} \sum_{i=1}^N D[x_i] = \frac{N}{N^2} \sigma^2 = \frac{\sigma^2}{N},$$

т.е. оценка \hat{a} стремится к a при $N \rightarrow \infty$, значит, оценка состоятельна.

Покажем несмещенность оценки для дисперсии:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N-1} \sum_{i=1}^N ((x_i - a) - (\bar{x} - a))^2 = \\ &= \frac{1}{N-1} \left[\sum_{i=1}^N (x_i - a)^2 - 2(\bar{x} - a) \sum_{i=1}^N (x_i - a) + N(\bar{x} - a)^2 \right] = \\ &= \frac{1}{N-1} \left[\sum_{i=1}^N (x_i - a)^2 - N(\bar{x} - a)^2 \right]. \end{aligned}$$

Математическое ожидание оценки

$$M[\hat{\sigma}^2] = \frac{1}{N-1} \left[\sum_{i=1}^N \underbrace{M[(x_i - a)^2]}_{\sigma^2} \right] - N \underbrace{M[(\bar{x} - a)^2]}_{\frac{\sigma^2}{N}} = \frac{1}{N-1} (N\sigma^2 - \sigma^2) = \sigma^2,$$

т.е. оно совпало с истинным значением, следовательно, оценка не смещена.

В математической статистике доказано, что оценка дисперсии состоятельна. Также доказано, что при нормальном распределении случайной величины X оценки математического ожидания и дисперсии эффективны.

Пример 1.2. Рассмотрим использование метода сравнения. Пусть требуется оценить \hat{T}_c – среднее время безотказной работы при плане [NUT]. Результатом испытаний является m – число отказов, и закон распределения времени отказов экспоненциальный.

Решение. Известна теоретическая вероятность безотказной работы

$$P(T) = e^{-\frac{T}{T_c}}.$$

Выборочная вероятность безотказной работы

$$1 - \frac{m}{N}.$$

Приравняем теоретическую и выборочную вероятности и получим искомую оценку среднего времени безотказной работы:

$$e^{-\frac{T}{\hat{T}_c}} = 1 - \frac{m}{N}.$$

После преобразований можно получить:

$$-\frac{T}{\hat{T}_c} = \text{Ln} \left(1 - \frac{m}{N} \right).$$

$$\hat{T}_c = -\frac{T}{\text{Ln} \left(1 - \frac{m}{N} \right)}.$$

Таким образом, при вычислении оценки средней наработки до отказа исходной информацией являются общее число испытываемых объектов и число отказавших из них.

1.2. Метод максимального правдоподобия

Метод максимального правдоподобия, предложенный Р. Фишером в 1912 г., основан на исследовании вероятности получения выборки наблюдений (x_1, x_2, \dots, x_n) .

По методу максимального правдоподобия оценки получаются из условия максимизации функции правдоподобия (вероятности или плотности результата эксперимента).

Функция правдоподобия обозначается следующим образом:

– для дискретных случайных величин:

$$P(\text{результат}, \theta) = P(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^N P(x_i, \theta) = L(\text{результат}, \theta),$$

– для непрерывных случайных величин:

$$P(x_1, \dots, x_n, \theta) = \Delta x^N \prod_{i=1}^N f(x_i, \theta) = L(\text{результат}, \theta),$$

где f – плотность распределения.

Таким образом, в общем виде можно записать:

$$L(\text{результат}, \theta) = \begin{cases} \prod_{i=1}^N P(x_i, \theta), \\ \prod_{i=1}^N f(x_i, \theta). \end{cases}$$

Максимум функции правдоподобия можно найти, приравняв ее производную к нулю.

В целях упрощения вычислений можно перейти от функции правдоподобия к ее логарифму $\text{Ln } L$. Такое преобразование допустимо, так как функция правдоподобия – положительная функция, и она достигает максимума в той же точке, что и ее логарифм.

Если параметр распределения есть векторная величина $\theta = (\theta_1, \theta_2, \dots, \theta_n)$, то оценки максимального правдоподобия находят из системы уравнений

$$\begin{cases} \partial \text{Ln} L(\theta_1, \theta_2, \dots, \theta_n) / \partial \theta_1 = 0; \\ \partial \text{Ln} L(\theta_1, \theta_2, \dots, \theta_n) / \partial \theta_2 = 0; \\ \dots \\ \partial \text{Ln} L(\theta_1, \theta_2, \dots, \theta_n) / \partial \theta_n = 0. \end{cases}$$

Для проверки того, что точка оптимума соответствует максимуму функции правдоподобия, необходимо найти вторую производную от этой функции. И если вторая производная в точке оптимума отрицательна, то найденные значения параметров максимизируют функцию.

Таким образом, нахождение оценок максимального правдоподобия включает следующие этапы:

- построение функции правдоподобия (ее натурального логарифма);
- дифференцирование функции по искомым параметрам и составление системы уравнений;
- решение системы уравнений для нахождения оценок;
- определение второй производной функции, проверку ее знака в точке оптимума первой производной и формирование выводов.

Пример 1.3. Требуется оценить \hat{T}_c – среднее время безотказной работы при плане испытаний [NUN]. Результатом испытаний являются t_1, t_2, \dots, t_N – моменты отказов (наработка) каждого объекта. Закон распределения времени отказов экспоненциальный.

Решение. Запишем произведение плотностей для непрерывной случайной величины:

$$L(t_1, \dots, t_N, T_c) = \frac{1}{T_c} e^{-\frac{t_1}{T_c}} \cdot \frac{1}{T_c} e^{-\frac{t_2}{T_c}} \dots \frac{1}{T_c} e^{-\frac{t_N}{T_c}} = \frac{1}{T_c^N} e^{-\frac{\sum t_i}{T_c}},$$

Далее запишем логарифмическую функцию правдоподобия:

$$\ln L = -N \ln T_c - \frac{\sum t_i}{T_c}.$$

Найдем ее производную:

$$\frac{\partial \ln L}{\partial T_c} = -\frac{N}{T_c} + \frac{\sum t_i}{T_c^2}.$$

Определим искомую оценку:

$$\frac{N}{\hat{T}_c} = \frac{\sum t_i}{\hat{T}_c^2}.$$

В итоге оценка средней наработки будет определяться выражением

$$\hat{T}_c = \frac{1}{N} \sum_{i=1}^N t_i.$$

В качестве другого примера оценим вероятность наступления события m раз в N случаях.

Функция правдоподобия в соответствии с биномиальным законом распределения имеет следующий вид:

$$L(m, P) = C_N^m P^m (1 - P)^{N-m}.$$

Согласно методу максимального правдоподобия получим искомую оценку:

$$\ln L(m, P) = \ln C_N^m + m \ln P + (N - m) \ln (1 - P);$$

$$\frac{\partial L(m, P)}{\partial P} = \frac{m}{P} - \frac{N - m}{1 - P};$$

$$\frac{m}{\hat{P}} = \frac{N - m}{1 - \hat{P}};$$

$$\hat{P} = \frac{m}{N}.$$

Оценим точность полученной оценки, т.е. определим ее дисперсию.

Число m можно представить суммой случайных величин, принимающих значение 1 (при успехе) или 0 (в противном случае): $m = z_1 + z_2 + \dots + z_N$.

Так как случайные величины независимы, то

$$M[Z] = 0(1 - P) + 1 \cdot P = P.$$

Оценка математического ожидания

$$M[\hat{P}] = \frac{N \cdot P}{N} = P.$$

Следовательно, оценка не смещена.

Определим, состоятельна ли полученная оценка.

Дисперсия случайной величины Z

$$D[Z] = (0 - P)^2(1 - P) + (1 - P)^2 \cdot P = P(1 - P).$$

В свою очередь, дисперсия полученной оценки будет определяться следующим образом:

$$D[\hat{P}] = \frac{1}{N^2} N \cdot D[Z] = \frac{P(1 - P)}{N}.$$

Так как очевидно, что при $N \rightarrow \infty$ значение дисперсии стремится к нулю, искомая оценка является состоятельной.

Исследуем, является ли полученная оценка эффективной. Для этого используем неравенство Рао–Крамера, определяющее нижнюю границу выборочной дисперсии несмещенной оценки:

$$D[\hat{P}] \geq \frac{1}{I_N(P)},$$

где $I_N(P)$ – количество информации об оцениваемом параметре, которая содержится в выборке, $I_N(P) = -M \left[\frac{\partial^2 \text{Ln} L(m, P)}{\partial P^2} \right]$

Очевидно, что

$$\frac{\partial^2 \text{Ln}(m, N)}{\partial P^2} = -\frac{m}{P^2} - \frac{N - m}{(1 - P)^2}.$$

Тогда, учитывая, что $M[m] = NP$, получим

$$M \left[\frac{\partial^2 \text{Ln}(m, N)}{\partial P^2} \right] = -\frac{NP}{P^2} - \frac{N - NP}{(1 - P)^2} = \frac{-N(1 - P) - NP}{P(1 - P)} = -\frac{N}{P(1 - P)}.$$

В итоге получили количество информации

$$I_N(P) = \frac{N}{P(1 - P)}.$$

Следовательно, неравенство Рао–Крамера будет иметь следующий вид:

$$D[\hat{P}] \geq \frac{P(1 - P)}{N}.$$

Таким образом, нижняя граница выборочной дисперсии, определяемая неравенством, совпадает с ранее полученным выражением для дисперсии искомой оценки, что дает возможность сделать вывод об эффективности оценки m/N .

1.3. Байесовский метод оценивания

Теория субъективной вероятности, на которой построен байесовский подход к оцениванию параметров, широко применяется в системах искусственного интеллекта в целях учета различного рода неопределенностей. Современные сторонники байесовской теории считают, что байесовский подход не противоречит классической теории статистического анализа. Классическая теория оставляет задачу выбора на усмотрение исследователя, не используя его знаний на заключительной стадии оценивания. Байесовская теория формализует этот этап и может поэтому рассматриваться как формальное завершение классической теории.

Исходной информацией для получения байесовской оценки являются:

- 1) априорное распределение исследуемого параметра;
- 2) результаты испытаний;
- 3) функция потерь.

Методической основой процесса перехода от дополнительной (априорной) информации (АИ), представленной в виде априорного распределения, к апостериорной информации путем добавления экспериментальных данных является теорема Байеса.

На основе априорного распределения $f_A(p)$ и экспериментальных данных по теореме Байеса определяется апостериорное распределение:

$$f(p) = \frac{f_A(p) \cdot \Pr(\text{rez} / p)}{\int_0^1 f_A(p) \cdot \Pr(\text{rez} / p) dp},$$

где $\Pr(\text{rez}/p)$ – вероятность результата испытания при условии, что параметр равен p .

Из апостериорного распределения в зависимости от вида функции потерь определяется байесовская оценка. Этот процесс можно представить в виде уточнения АИ результатами эксперимента (рис. 1.1).

1.3.1. Выбор априорного распределения

Априорное распределение представляет собой дополнительную, субъективную информацию, которой необходимо задаться для получения байесовской оценки.

Существует несколько способов, позволяющих определить вид априорного распределения.

Рассмотрим способ, основная идея которого заключается в том, что для обеспечения гарантии достоверности АИ выбирают такое априорное распределение,

которое будет иметь наибольшее рассеивание при имеющихся предварительных данных. В качестве меры рассеивания удобно выбрать энтропию распределения:

$$H = - \int_{-\infty}^{\infty} f(x) \text{Ln}(f(x)) dx.$$

Решение задачи сводится к определению функции $f(x)$, подчиненной некоторым уравнениям связи, энтропия которой максимальна. АИ используется при составлении уравнений связи.

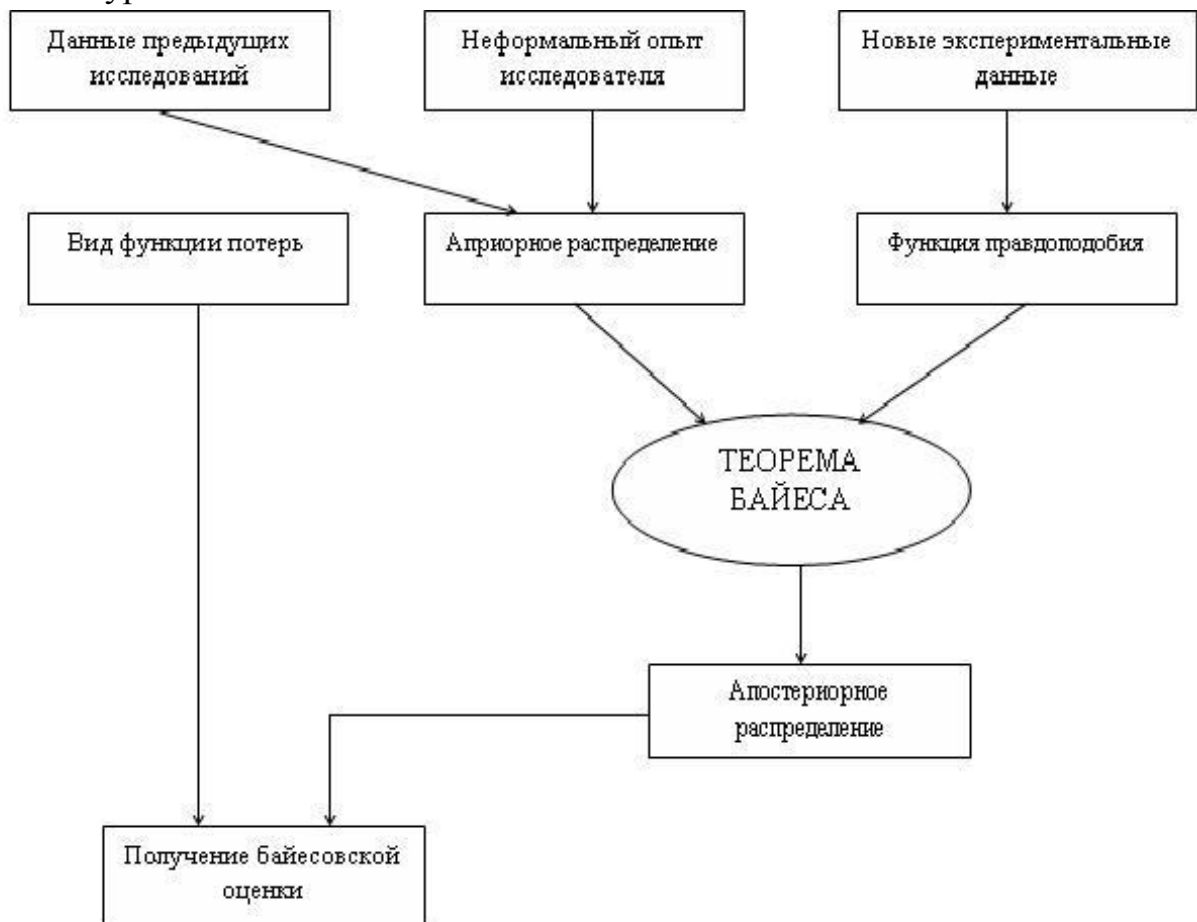


Рис. 1.1. Определение байесовской оценки

Предположим, что АИ задана в виде интервала $[x_H, x_B]$. Определим вид функции $f(x)$, согласующийся с опытными данными и обладающий наибольшим рассеиванием.

При наличии подобной АИ можно составить одно уравнение связи:

$$\int_{x_H}^{x_B} f(x) dx = 1.$$

Тогда функция правдоподобия, выраженная через множители Лагранжа, будет иметь следующий вид:

$$L = -f(x)\text{Ln}f(x) + \theta_1 f(x),$$

где θ_1 – неопределенный множитель Лагранжа. Дифференцируя данную функцию правдоподобия по $f(x)$, получим $-\text{Ln}f(x) - 1 + \theta_1 = 0$, откуда $f(x) = e^{-1+\theta_1}$.

Подставив последний результат в уравнение связи, получим

$$\int_{X_H}^{X_B} e^{-1+\theta_1} dx = e^{-1+\theta_1} x \Big|_{X_H}^{X_B} = e^{-1+\theta_1} (X_B - X_H) = 1,$$

тогда

$$e^{-1+\theta_1} = f(x) = \frac{1}{X_B - X_H}.$$

Таким образом, при представлении АИ в виде интервала, в качестве априорного распределения следует выбирать равномерное распределение.

При представлении АИ в виде оценки неизвестного параметра и ее дисперсии определить вид априорного распределения можно на основе принципа сопряженности. Его суть в нижеследующем.

В общем случае по теореме Байеса любое априорное распределение может быть использовано с любой функцией правдоподобия. В практических случаях обычно используют априорное распределение специальных форм, приводящих к простым оценкам. Для распределения результатов эксперимента можно подобрать такое семейство априорных распределений, при котором и апостериорные распределения будут принадлежать тому же семейству. Такое семейство называется сопряженным относительно распределения результатов эксперимента.

При оценивании вероятности наступления события m раз в N случаях распределение результатов является биномиальным. Известно, что для биномиального распределения сопряженным распределением является бета-распределение, т.е. в таком случае и апостериорное распределение будет также бета-распределением. Таким образом, для рассматриваемого варианта представления АИ в качестве априорного распределения целесообразно принять бета-распределение:

$$f_A(p) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)},$$

где $B(a,b)$ – бета-функция; a, b – параметры бета-распределения.

При определении параметров a и b обычно полагают априорную оценку математическим ожиданием априорного распределения, а ее дисперсию – дисперсией априорного распределения:

$$\begin{cases} \hat{p}_A = \frac{a}{a+b}; \\ \sigma_A^2 = \frac{a \cdot b}{(a+b)^2(a+b+1)}, \end{cases}$$

где \hat{p}_A, σ_A^2 – соответственно априорная оценка вероятности и ее дисперсия.

В результате решения системы можно получить:

$$a = \hat{p}_A \left(\frac{\hat{p}_A(1-\hat{p}_A)}{\sigma_A^2} - 1 \right); \quad b = a \frac{(1-\hat{p}_A)}{\hat{p}_A}.$$

Логично предположить, что априорную оценку можно считать и наиболее вероятным значением неизвестного параметра, т.е. модой априорного распределения: $\hat{p}_A = \frac{a-1}{a+b-2}$. В этом случае решение системы уравнений возможно лишь численными методами.

В случае, когда дисперсия априорной оценки неизвестна, параметры априорного распределения можно определить либо с помощью рациональных степеней уверенности исследователя в качестве АИ, либо с помощью субъективных уровней доверия к ней. Сущность способа использования рациональных степеней уверенности заключается в применении такого априорного распределения, которое гарантирует, что потери будут не больше ожидаемых.

Суть способа субъективных уровней доверия заключается в том, что для определения параметров априорного распределения вводится байесовская информация, которая содержится в априорном распределении об оцениваемом параметре. Определяется она также как и информация Фишера. Исследователь, используя свои знания и опыт, а также мнения экспертов, задает весовую значимость априорной информации по отношению к опытной:

$$S = \frac{I_{a,b}^B(p)}{I_N(p)},$$

где $I_{a,b}^B(p)$ – байесовская информация; $I_N(p)$ – информация Фишера.

Значение S показывает, насколько исследователь доверяет АИ по отношению к опытным данным.

Значения параметров априорного бета-распределения можно определить из системы уравнений:

$$\begin{cases} \frac{a}{a+b} = \hat{p}_A; \\ I_{a,b}^B(\hat{p}) = S \cdot I_N(\hat{p}). \end{cases}$$

В результате ее решения получим:

$$a = \hat{p}_A [\hat{p}_A (1 - \hat{p}_A) \cdot I_N(p) S - 1];$$

$$b = a \frac{1 - \hat{p}_A}{\hat{p}_A}.$$

Отметим, что выбор вида априорного распределения должен основываться на разумном сочетании субъективных знаний и объективных данных. Существует опасность злоупотребления неформальным знанием, что в итоге приведет к ошибочному выбору априорного распределения.

1.3.2. Определение апостериорного распределения

Выше было отмечено, что для определения апостериорного распределения используется формула Байеса, в которую входят априорное распределение и вероятность результата испытаний.

При биномиальной схеме эксперимента вероятность результата

$$p(m, N) = (1 - p)^m p^{N-m}.$$

Получим апостериорные распределения при рассмотренных выше априорных распределениях.

Априорная информация в виде априорного бета-распределения

Применив теорему Байеса, получим апостериорное распределение, которое тоже является бета-распределением:

$$f(p) = \frac{p^{a+N-m-1} (1-p)^{b+m-1}}{\int_0^1 p^{a+N-m-1} (1-p)^{b+m-1} dp}.$$

Для примера на рис. 1.2 показаны априорное бета-распределение с параметрами $a = 3, b = 5$ (что соответствует $\hat{p}_A = 0,4$) и апостериорное распределение при результатах испытаний $m = 3, N = 4$.

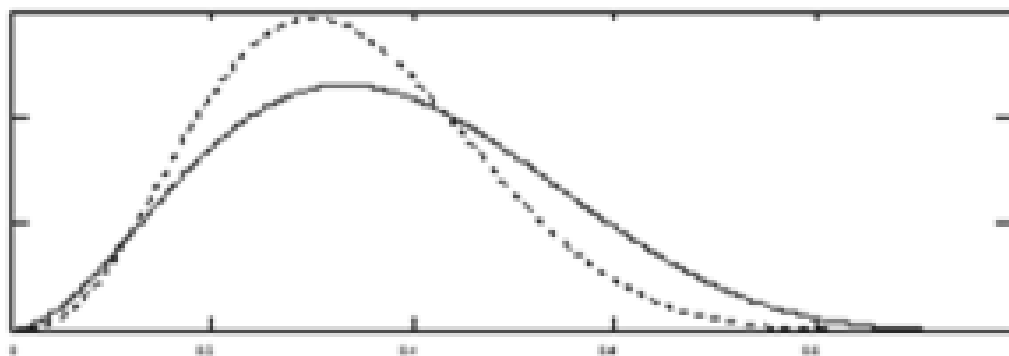


Рис. 1.2 Априорное (сплошная кривая) и апостериорное (пунктирная кривая) бета-распределения

Априорная информация в виде равномерного распределения

Апостериорное распределение в этом случае будет иметь следующий вид:

$$f(p) = \frac{(1-p)^m p^{N-m}}{\int_{p_H}^{p_B} (1-p)^m p^{N-m} dp},$$

где p_H, p_B – соответственно нижняя и верхняя границы априорного распределения.

Для примера на рис.1.3 показаны априорное равномерное распределение с границами $[0.2; 0.7]$ и апостериорное распределение при результатах испытаний $m = 1, N = 4$.

Отличие апостериорного распределения от априорного в обоих случаях вызвано результатом испытаний, и отличие тем больше, чем сильнее отличается экспериментальная оценка от априорной.

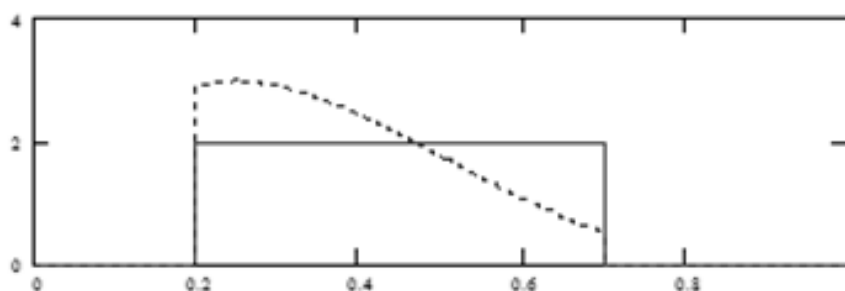


Рис. 1.3 Равномерное априорное (сплошная кривая) и апостериорное (пунктирная кривая) распределения

1.3.3. Расчет байесовской оценки при биномиальной схеме испытаний

При проведении статистического эксперимента неизбежны потери, возникающие вследствие замены истинного значения параметра его оценкой, выражающиеся функцией потерь $\psi(\hat{p}, p)$. В зависимости от характера потерь исследователь может обосновать вид функции потерь. Рассмотрим различные ее виды и для каждой функции запишем лучшую (которая минимизирует функцию потерь) байесовскую оценку.

В теории оценивания широко используется функция потерь вида

$$\psi(p, \hat{p}) = C(p)(\hat{p} - p)^k,$$

где $C(p)$ – функция от неизвестного параметра, которая предполагается положительной и конечной.

При $k = 1$ функция потерь линейна, при ней лучшей байесовской оценкой является медиана апостериорного распределения, определяемая из уравнения

$$F(\hat{p}_B^{Me}) = \frac{B_{p_B^{Me}}(a + N - m, b + m)}{B(a + N - m, b + m)} = \frac{1}{2}. \quad (1.1)$$

Здесь $B_{p_B^{Me}}(a + N - m, b + m)$ – неполная бета-функция, где

$$B_{p_B^{Me}} = \int_0^{p_B^{Me}} (1 - p)^{a+N-m-1} p^{b+m-1} dp.$$

При $k = 2$ функция потерь является квадратичной. Известно, что при квадратичной функции потерь лучшая байесовская оценка определяется как математическое ожидание апостериорного распределения:

$$\hat{p}_B^{MO} = \frac{\int_0^1 p^{a+N-m} (1 - p)^{b+m-1} dp}{\int_0^1 p^{a+N-m-1} (1 - p)^{b+m-1} dp} = \frac{B(a + N - m + 1, b + m)}{B(a + N - m, b + m)}. \quad (1.2)$$

Получим аналитическое выражение этой оценки. Для этого выразим бета-функцию через гамма-функции:

$$B(a + N - m + 1, b + m) = \frac{\Gamma(a + N - m + 1)\Gamma(b + m)}{\Gamma(a + b + N + 1)}. \quad (1.3)$$

Подставив уравнение (1.3) в (1.2), с учетом свойств гамма-функции, получим байесовскую оценку:

$$\hat{p}_B^{MO} = \frac{a + N - m}{a + b + N}. \quad (1.4)$$

В теории оценивания также используется квадратичная функция потерь относительной ошибки вида

$$\psi(\hat{p}, p) = \left(\frac{\hat{p} - p}{p} \right)^2,$$

для которой лучшая байесовская оценка

$$\hat{p}_B^T = \frac{\int_0^1 \frac{f(p)}{p} dp}{\int_0^1 \frac{f(p)}{p^2} dp}. \quad (1.5)$$

Проведя аналогичные преобразования, можно получить

$$\hat{p}_B^T = \frac{a + N - m - 2}{a + b + N - 2}. \quad (1.6)$$

При отсутствии информации о виде функции потерь логично выбрать в качестве байесовской оценки моды (наиболее вероятное значение) апостериорного распределения:

$$\hat{p}_B^{\text{Mod}} = \frac{a - 1 + N - m}{a + b + N - 2}. \quad (1.7)$$

Оценки (1.1) – (1.7) получены в случае, когда в качестве априорного распределения выбрано бета-распределение.

В случае равномерного априорного распределения байесовские оценки для рассмотренных функций потерь, кроме (1.6), находятся аналогично, т.е. как медиана, математическое ожидание и мода апостериорного распределения.

2. ОБРАБОТКА МАССИВА ДАННЫХ

2.1. Первичная обработка исходных данных и построение гистограммы

При большом числе наблюдений простая статистическая совокупность становится неудобной формой записи статистического материала, ввиду громоздкости и малой наглядности. Для придания большей компактности и наглядности статистический материал подвергается дополнительной обработке, в частности строится статистический ряд. Статистический ряд представляется графически в виде гистограммы. По оси абсцисс гистограммы откладываются интервалы, на каждом из которых, как на основании, строится прямоугольник, высота которого пропорциональна (в выбранном масштабе) соответствующей частоте попадания конкретного значения в интервал.

Перед построением гистограммы данные сортируют (по возрастанию), определяют минимальное x_{\min} и максимальное x_{\max} значения, а также размах вариационного ряда ($x_{\max} - x_{\min}$).

Для построения гистограммы интервал изменения данных нужно разбить на участки одинаковой длины.

С одной стороны, число таких участков должно быть как можно больше, а с другой – в каждый из этих участков должно попадать как можно больше значений x_i . Компромисс между этими требованиями приводит к тому, что обычно выбирают число интервалов k для построения гистограммы как ближайшее целое к \sqrt{n} .

Используют и другие подходы к определению числа интервалов. Например, известна формула Стэрджесса: $k = 1 + 3.322 \lg(n)$. Примерный вид гистограммы изображен на рис. 2.1.

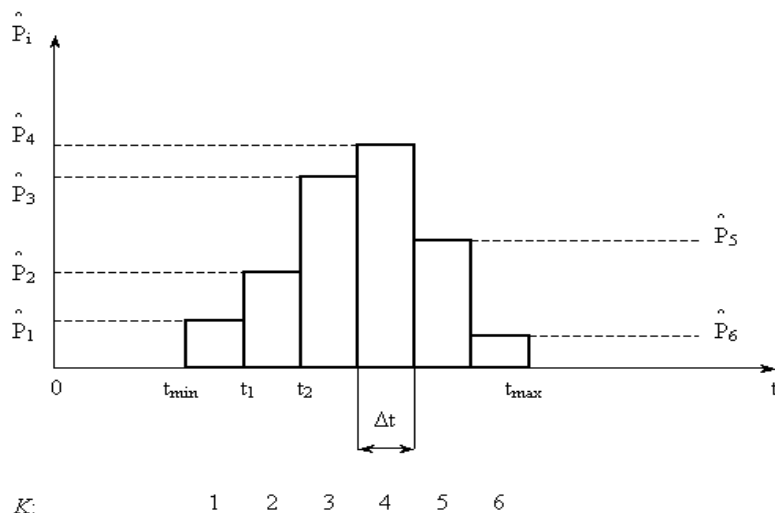


Рис 2.1. Гистограмма

2.2. Расчет статистических параметров распределения

Расчет статистических параметров распределения (точечных оценок), а именно математического ожидания \tilde{m}_x , дисперсии \tilde{D}_x , среднеквадратичного отклонения $\tilde{\sigma}_x$, асимметрии A_x и эксцесса E_x , проводят по следующим формулам:

$$\tilde{m}_x = \frac{1}{n} \sum_{i=1}^n x_i ; \quad (2.1)$$

$$\tilde{D}_x = \frac{1}{n-1} \sum_{i=1}^n (x_i - \tilde{m}_x)^2 ; \quad (2.2)$$

$$\tilde{\sigma}_x = \sqrt{\tilde{D}_x} ; \quad (2.3)$$

$$A_x = \frac{1}{(n-1)\tilde{\sigma}_x^3} \sum_{i=1}^n (x_i - \tilde{m}_x)^3 ; \quad (2.4)$$

$$E_x = \frac{1}{(n-1)\tilde{\sigma}_x^4} \sum_{i=1}^n (x_i - \tilde{m}_x)^4 - 3. \quad (2.5)$$

Асимметрия и эксцесс используются для выбора аппроксимирующей функции. Коэффициент асимметрии является характеристикой «скошенности» распределения, например, если распределение симметрично относительно математического ожидания, то $A_x = 0$.

На рис. 2.2,а распределение $f_2(t)$ имеет положительную асимметрию $A_x > 0$, а $f_3(t)$ – отрицательную $A_x < 0$.

Экссесс характеризует «крутость» (остро- или плосковершинность) распределения. Для нормального распределения $E_x = 0$.

Кривые $f(t)$, более островершинные по сравнению с нормальной, имеют $E_x > 0$, и, наоборот, более плосковершинные имеют $E_x < 0$ (рис.2.2,б).

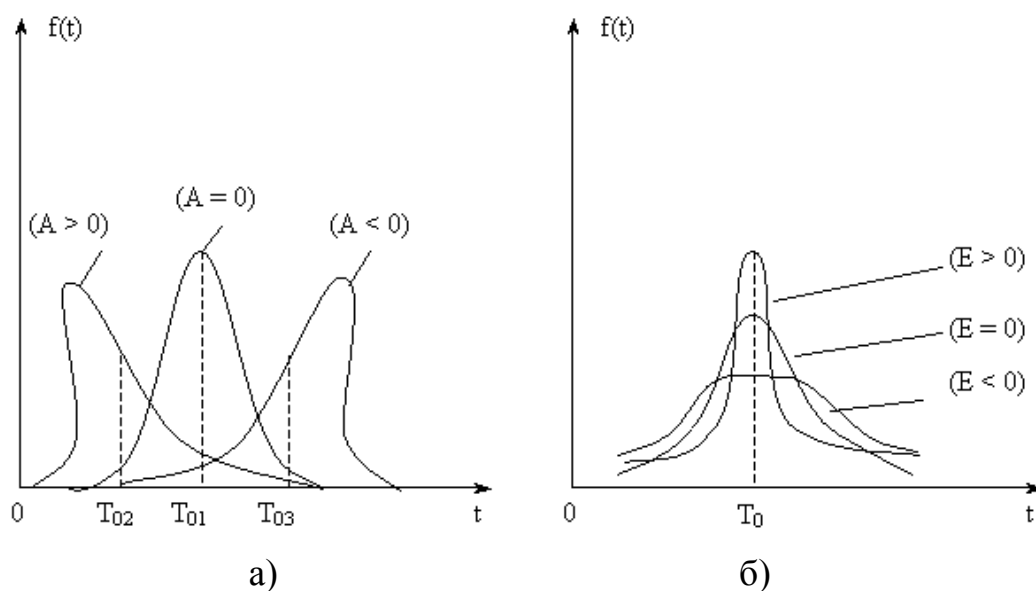


Рис. 2.2. Асимметрия и эксцесс

2.3. Подбор теоретического распределения и его параметров

Выбор закона распределения состоит в подборе аналитической функции, наилучшим образом аппроксимирующей эмпирические функции.

Выбор – в значительной мере процедура неопределенная и во многом субъективная, при этом многое зависит от априорных знаний об объекте и его свойствах, условиях работы и т.д.

Подбор теоретического распределения состоит из следующих этапов:

- 1) подбора вида распределения (закона распределения);
- 2) расчета параметров распределения (переменных, входящих в выражение для функции и плотности распределения);
- 3) проверки правильности подбора.

Таким образом, сначала подбирают вид теоретического распределения и его параметры, затем проверяют правильность подбора с помощью критериев согласия, например Колмогорова или Пирсона.

Некоторое предположение о законе распределения случайной величины или о параметрах этого закона, формулируемое на основе выборки, получило название статистической гипотезы.

Примерами статистических гипотез являются предположения: генеральная совокупность распределена по экспоненциальному закону; математические ожидания двух экспоненциально распределенных выборок равны друг другу. В первой из них высказано предположение о виде закона распределения, а во второй – о параметрах двух распределений. Гипотезы, в основе которых нет никаких допущений о конкретном виде закона распределения, называют непараметрическими, в противном случае – параметрическими.

Гипотезу, утверждающую, что различие между сравниваемыми характеристиками отсутствует, а наблюдаемые отклонения объясняются лишь случайными колебаниями в выборках, на основании которых производится сравнение, называют нулевой (основной) гипотезой и обозначают H_0 . Наряду с основной гипотезой рассматривают и альтернативную (конкурирующую, противоречащую) ей гипотезу H_1 . И если нулевая гипотеза будет отвергнута, то будет иметь место альтернативная гипотеза.

Закон теоретического распределения подбирается исходя из вида гистограммы. Например, предположим по виду гистограммы, что теоретическое распределение может быть одного из четырех видов:

- 1) нормальное;
- 2) показательное (экспоненциальное);
- 3) равномерное;
- 4) распределение Рэлея.

Параметры, входящие в выражение для функции и плотности теоретического распределения, найдём исходя из принципа максимума правдоподобия: так, чтобы вычисленные по этим параметрам математическое ожидание (для однопараметрических законов) или математическое ожидание и дисперсия (для двухпараметрических законов) совпали с выборочными.

Так, для нормального распределения параметры m и σ берем равными соответственно выборочным математическому ожиданию и дисперсии:

$$m = \tilde{m}_x; \quad \sigma = \tilde{\sigma}_x.$$

Для показательного распределения находим параметр λ :

$$\lambda = \frac{1}{\tilde{m}_x}.$$

Параметры равномерного распределения a и b будут определяться выражениями:

$$a = \tilde{m}_x - \tilde{\sigma}_x \sqrt{3}; \quad b = \tilde{m}_x + \tilde{\sigma}_x \sqrt{3}.$$

Параметр σ для распределения Рэлея

$$\sigma = \tilde{m}_x \sqrt{\frac{2}{\pi}}.$$

Приведём выражения для вычисления функции и плотности распределения.

Для нормального распределения

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}; \quad F_x(x) = \int_{-\infty}^x f_x(t) dt = \Phi\left(\frac{x-m}{\sigma}\right) + 0.5, \quad (2.6)$$

где $\Phi(u)$ – интеграл Лапласа.

Для показательного распределения

$$f_x(x) = \begin{cases} 0; & x < 0; \\ \lambda e^{-\lambda x}; & x \geq 0; \end{cases} \quad F_x(x) = \begin{cases} 0; & x < 0; \\ 1 - e^{-\lambda x}; & x \geq 0. \end{cases} \quad (2.7)$$

Для равномерного распределения

$$f_x(x) = \begin{cases} 0; & x \notin [a, b]; \\ \frac{1}{b-a}; & x \in [a, b]; \end{cases} \quad F_x(x) = \begin{cases} 0; & x \notin [a, b]; \\ \frac{x-a}{b-a}; & x \in [a, b]. \end{cases} \quad (2.8)$$

Для распределения Рэлея

$$f_x(x) = \begin{cases} 0; & x < 0; \\ \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}; & x \geq 0; \end{cases} \quad F_x(x) = \begin{cases} 0; & x < 0; \\ 1 - e^{-\frac{x^2}{2\sigma^2}}; & x \geq 0. \end{cases} \quad (2.9)$$

Для визуального сравнения и, возможно, первого отсеивания неподходящих законов распределений строят на одном графике теоретические и эмпирическую плотности распределения. Эмпирическая плотность распределения – это та же гистограмма, у которой масштаб по оси ординат изменён таким образом, чтобы площадь под кривой стала равна единице. Для этого все метки по оси ординат в гистограмме нужно разделить на nh , где n – число экспериментальных данных, а h – ширина интервала при построении гистограммы. Теоретические плотности распределения строим по формулам (2.6) – (2.9).

Ответ на вопрос, какое теоретическое распределение лучше всего согласуется с эмпирическим (нормальное, показательное, равномерное или Рэлея), можно получить с помощью критерия согласия.

Критерий согласия – это правило проверки гипотезы о том, что случайная величина, представленная своей выборкой, имеет распределение предполагаемого типа.

Существует несколько критериев согласия. Чаще всего используются критерии Пирсона и Колмогорова.

В общем случае проверка состоит в следующем. Рассчитывается критерий как некоторая мера расхождения теоретического и эмпирического распределений, причем эта мера является случайной величиной.

Чем больше мера расхождения, тем хуже согласованность эмпирического распределения с теоретическим. Если она больше установленного критического значения, то гипотезу о выборе закона распределения следует отвергнуть как мало правдоподобную. В противном случае экспериментальные данные не противоречат принятому распределению.

2.3.1. Критерий согласия Колмогорова

Критерий согласия Колмогорова применяется для проверки правильности подбора теоретического распределения. Для его применения нужно найти максимальную по модулю разность между теоретической (генеральной, предполагаемой, подобранной) функцией распределения $F_x(x)$ и выборочной (опытной, эмпирической, экспериментальной) $\tilde{F}_x(x)$:

$$D = \max_{\forall x} |F_x(x) - \tilde{F}_x(x)|, \quad (2.10)$$

Далее необходимо вычислить значение $\lambda = D\sqrt{n}$, которое следует сравнить с квантилем (λ_p) распределения Колмогорова.

Если величина λ не очень большая (не превосходит квантиля λ_p), то на уровне значимости p статистическую гипотезу можно принять.

Если же $\lambda > \lambda_p$, то теоретическое распределение подобрано неверно.

График эмпирической функции распределения $\tilde{F}_x(x)$ представляет ступенчатую линию – ломаную со ступеньками высотой $1/n$ в точках с абсциссами x_i .

Максимум разности между теоретической и эмпирической функциями распределения достигается как раз на одной из этих ступенек.

Уровень значимости обычно выбирают достаточно "жёстким" (большим), например 0.3.

Критерий Колмогорова позволяет проверить согласованность распределений по малым выборкам, он проще критерия Пирсона, поэтому его часто применяют на практике. При применении этого критерия следует учитывать, что условия применения критерия предусматривают, что теоретическая функция распределения должна быть известна полностью (известны вид функции и ее параметры). Но на практике параметры обычно неизвестны и оцениваются по ЭД. Это при-

водит к завышению значения вероятности соблюдения нулевой гипотезы, т.е. повышается риск принять в качестве правдоподобной гипотезу, которая плохо согласуется с ЭД (повышается вероятность совершить ошибку второго рода).

2.3.2. Критерий согласия Пирсона

В критерии согласия Пирсона сравниваются между собой теоретические и эмпирические числа попаданий в интервалы.

Возьмём те интервалы, по которым была построена гистограмма. Эмпирические числа попаданий в эти интервалы n_j сравниваем с теоретическим числом попаданий np_j , где p_j – вероятность попадания нашей величины в j -й интервал. Теоретическое распределение можно считать подобранным верно на уровне значимости α , если суммарная квадратичная относительная разность между теоретическим и практическим числом попаданий в каждый интервал будет не очень большой. Должно выполняться условие

$$\sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j} \leq \chi_{1-\alpha}^2(k-L), \quad (2.11)$$

где $(k-L)$ – число степеней свободы; L – число независимых условий, наложенных на частоты p_j .

Например, независимыми условиями могут быть:

- а) условие $\sum p_j = 1$;
- б) условие совпадения теоретического среднего значения и экспериментального;
- в) условие совпадения теоретической дисперсии и экспериментальной.

На практике обычно принимают, что $L = 3$. Чем больше число степеней свободы, тем больше случайная величина χ^2 подчиняется распределению Пирсона.

Мощность критерия зависит от количества разрядов и объема выборки. Критерий рекомендуется применять при $np_j \geq 5$.

Построим таблицу результатов, в которую занесём: номера интервалов (1-й столбец), границы интервалов a_j и b_j (2-й и 3-й столбцы), практическое число попаданий n_j (4-й столбец), вероятность попадания в интервал p_j (5-й столбец), теоретическое число попаданий np_j (6-й столбец) и величины $(n_j - np_j)^2 / np_j$ (7-й столбец).

Вероятность попадания в j -й интервал подсчитывается по формуле

$$p_j = F_x(b_j) - F_x(a_j).$$

Проверим выполнение условия, что все $np_j \geq 5$, и объединим те интервалы, в которых $np_j < 5$. Перестроим таблицу и подсчитаем статистику Пирсона. Сравним

полученную величину с квантилем распределения Пирсона на заданном уровне значимости.

В целом, с помощью критерий согласия, можно опровергнуть выбранную гипотезу, если же условие критерия при выбранном уровне значимости выполняется, то это не может служить доказательством правильности гипотезы, а указывает лишь на то, что гипотеза не противоречит данным эксперимента.

Пример 2.1. Результаты опроса общественного мнения говорят о том, что наибольшей популярностью покупателей пользуются мужские пальто серого цвета (30% опрошенных), затем пальто черного цвета (26% опрошенных), синего (22%) и коричневого (22%). Имеется ли существенная разница между цветами пальто с точки зрения покупателей, если считать, что результаты получены при $n = 100$ и при $n = 1000$?

Решение. Применим критерий Пирсона при уровне значимости $\alpha = 0.05$ для проверки гипотезы о принадлежности данных опроса к равномерному распределению. Число интервалов – $k = 4$.

При $n = 100$ имеем значения частот

$$n_1 = 0.3 \cdot n = 30, n_2 = 26, n_3 = 22, n_4 = 22.$$

Вероятности p_i ($i = 1, \dots, 4$) для равномерного распределения равны 0.25.

Теоретические частоты $np_i = 100 \cdot 0.25 = 25$.

Критическое значение $\chi^2_{1-\alpha}$ с числом степеней свободы $r = (4 - 2 - 1) = 1$ равно $\chi^2_{0,95} = 3,8$.

Так как значение критерия

$$\frac{(30 - 25)^2 + (26 - 25)^2 + (22 - 25)^2 + (22 - 25)^2}{25} = 1,76$$

меньше критического значения ($\chi^2_{0,95} = 3,8$), то гипотезу о принадлежности данных опроса к равномерному распределению можно принять.

При $n = 1000$ значение критерия станет равным 17,6, поэтому гипотеза о принадлежности распределения популярности пальто различного цвета к равномерному распределению должна быть отвергнута. В этом случае имеет смысл, например, упорядочить все цвета мужских пальто по уровню спроса.

3. МОДЕЛИ ВРЕМЕННЫХ РЯДОВ

3.1. Структура временного ряда

Случайный процесс является функцией $X(t, \omega)$ двух переменных: времени (t) и элементарного случайного события (ω).

Если t — подмножество действительных чисел, то это случайный процесс (СП) с непрерывным временем, если t принадлежит множеству целых чисел, то имеем случайную последовательность или временной ряд.

Таким образом, временной ряд – это последовательность чисел, которые являются значениями протекающего во времени процесса.

Существуют два эквивалентных подхода к изучению СП:

1. Если рассматривать $X(\omega)$ при фиксированных значениях t , то можно получить при каждом t случайную величину, и в этом случае СП определяется как семейство случайных величин

$$X(t, \omega) = \{X(t_1, \omega), X(t_2, \omega) \dots X(t_j, \omega)\},$$

где переменная t может принимать любые значения.

2. Фиксируется элементарное событие ω , и при каждом ω_i , в результате получается неслучайная функция $x(t)$, которая называется реализацией СП. Тогда СП определяется ансамблем реализаций

$$X(t, \omega) = \{X(t, \omega_1), X(t, \omega_2) \dots X(t, \omega_i)\}.$$

Ансамбль может содержать бесконечное число реализаций. Таким образом, любая реализация $x(t, \omega)$ является одной из дважды бесконечного множества функций, порожденных СП. В дальнейшем, как это принято в прикладной теории СП, вместо $X(t, \omega)$ будем писать $X(t)$. Написание $X(t)$ СП без указания аргумента ω оправдывается тем, что при проведении экспериментов регистрируются значения $X(t)$ и t без регистрации значений ω .

Временные ряды, как правило, возникают в результате измерения некоторого показателя. Это могут быть как показатели (характеристики) технических систем, так показатели природных, социальных, экономических и других систем, например:

- в экономике – курсы акций, валют, объемы продаж, объем производства;
- в метеорологии – температура, интенсивность осадков, сила ветра, давление, влажность и т.д.;
- в гидрологии – уровни воды в водоемах;
- в технике – параметры сигналов.

При изучении временных рядов ставятся следующие цели:

- выявление особенностей временного ряда;
- подбор модели, позволяющей решать задачи прогноза и управления.

При построении модели ряд разделяют на детерминированную (неслучайную) и случайную части.

Пусть x_1, x_2, \dots, x_n – временной ряд, тогда элемент ряда можно записать следующим образом:

$$x_t = d_t + \varepsilon_t, \quad (3.1)$$

где d_t – детерминированная (закономерная, систематическая) составляющая временного ряда, элементы которой вычисляются по определенному правилу как функция от времени; ε_t – случайный компонент временного ряда, придающий хаотичность и непредсказуемость. Для описания и анализа этого компонента используют понятия и методы теории вероятностей и математической статистики.

В большинстве приложений в детерминированной части чаще всего выделяют три составляющие:

$$d_t = tr_t + S_t + C_t,$$

где tr_t – тренд; S_t , C_t – сезонный и циклический компонент соответственно.

В последнее время стали добавлять ещё один компонент – интервенцию, под которой понимается существенное кратковременное воздействие на временной ряд (например, «черный вторник», нештатная ситуация и др.).

Трендом (*trend* – тенденция, направление) временного ряда называют изменяющийся, нециклический компонент, описывающий влияние долговременных факторов, эффект которых сказывается постепенно. К таким факторам относятся изменение демографических характеристик, рост потребления, технологическое и экономическое развитие. Как правило, действие перечисленных факторов происходит постепенно. Описать характер изменения тренда можно с помощью следующих моделей:

- линейной модели: $tr_t = b_0 + b_1 t$;
- полиномиальная модель: $tr_t = b_0 + b_1 t + b_2 t^2 + \dots + b_n t^n$ (обычно $n \leq 5$);
- экспоненциальной модели: $tr_t = \exp(b_0 + b_1 t)$;
- логистической модели: $tr_t = \frac{a}{1 + b \cdot \exp(-ct)}$;
- модели Гомперца: $\log(tr_t) = a - br^t$, где $0 < r < 1$.

Сезонный компонент временного ряда описывает поведение, изменяющееся регулярно в течение заданного периода (года, месяца, недели, дня и т.п.). Он состоит из почти повторяющихся циклов. Примеры: объем продаж перед праздниками, пики грузоперевозок в течение суток, пики перевоза пассажиров в городском транспорте, изменение нагрузки на электрическую и телефонную сети в течение суток и года и т.п.

Циклический компонент описывает длительные периоды относительного подъема и спада (от двух до десяти лет) и состоит из циклов, меняющихся по амплитуде и протяженности. Примеры: взаимодействие спроса и предложения, рост и истощение ресурсов, изменение налоговой политики, кадровая политика.

3.2. Модели случайных процессов

Одним из главных направлений исследования временных рядов является выбор класса моделей. В качестве первичного шага при выборе класса моделей можно предложить анализ имеющейся априорной информации об изучаемом процессе. Источниками такой информации могут быть литературные данные, точка зрения ведущих специалистов и т.д. Эта информация может быть уточнена и дополнена в ходе проведения предварительных экспериментов. По результатам анализа априорной информации выдвигается гипотеза о том, что данные порождены одной из моделей среди множества возможных классов. В каждом классе нужно найти наиболее подходящую модель и затем по некоторому множеству критериев проверить, какая из моделей удовлетворяет наибольшему числу критериев проверки адекватности. Кроме того, модель должна обеспечивать удовлетворительный прогноз. Конечно, такой подход требует большой вычислительной работы. Однако задачу можно упростить, рассматривая отдельно задачу выбора класса и проверки адекватности.

Для описания *случайного компонента* временного ряда используют понятие случайного процесса.

В предыдущем пункте было сформулировано общее понятие случайного процесса, для дальнейшего описания приведем упрощенное определение.

Случайным процессом (СП) $X(t)$ называют случайную функцию, значение которой при каждом t является случайной величиной.

Говорят, что случайный процесс задан, если для каждого t определена функция распределения

$$F_t(x) = P(X(t) \leq x).$$

Для каждой пары элементов t_1 и t_2 определена функция распределения

$$F_{t_1, t_2}(x_1, x_2) = P(X(t_1) \leq x_1, X(t_2) \leq x_2),$$

и в общем случае для каждого элемента t_1, t_2, \dots, t_n определена функция распределения величины $(x(t_1), x(t_2), \dots, x(t_n))$ –

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n).$$

Указанные функции распределения согласованы в том смысле, что старшие распределения определяют младшие:

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \lim_{x_{n+1} \rightarrow \infty} F_{t_1, \dots, t_n, t_{n+1}}(x_1, \dots, x_n, x_{n+1}). \quad (3.2)$$

Важнейшими классами СП являются *гауссовы*, *стационарные* и *стационарные эргодические* СП.

Гауссовы случайные процессы – это СП, у которых все конечномерные функции распределения нормальны, т. е. система величин $(X(t_1), \dots, X(t_j))$ имеет нормальное распределение при любых t . Ценность нормального распределения состоит в том, что n -мерное нормальное распределение выражается через математическое ожидание и автокорреляционную функцию (АКФ) процесса, для определения которых достаточно знания двумерной плотности вероятности.

Стационарные случайные процессы являются *стационарными в узком смысле*, если для любого действительного τ его конечномерная функция распределения не меняется при сдвиге на τ :

$$f(x_1 \dots x_n; t_1 \dots t_n) = f(x_1 \dots x_n; t_1 + \tau \dots t_n + \tau).$$

Случайным процессом называется *стационарным в широком смысле (слабо стационарным)*, если среднее значение постоянно, а АКФ зависит только от разности $t_2 - t_1 = \tau$:

$$M[X(t)] = m_x = \text{const.}$$

Реальные СП являются *нестационарными*. Стационарность – это математическая абстракция. Многие нестационарные СП на конечных интервалах времени можно аппроксимировать стационарным процессом. Вопрос о том, когда реальный процесс можно с достаточной степенью точности аппроксимировать стационарным СП, решается на основе изучения экспериментальной реализации и знания ее механизма генерации процесса.

Математически простейшей моделью является последовательность независимых случайных величин, например процесс типа «белого шума».

Белым шумом называется процесс, в котором средние значения равны нулю, сами случайные величины независимы и имеют одинаковые распределения.

Описание временного ряда функцией распределения (3.2) слишком сложно, и получить эту функцию практически невозможно. Обычно прибегают к более

простым моделям. Примерами используемых более простых и реальных моделей являются модели процессов скользящего среднего и процессов авторегрессии.

Процессом скользящего среднего (первого порядка) (сокращенно MA1, от англ. moving average – движущее среднее) называют процесс

$$x(t) = \varepsilon_t + \theta \varepsilon_{t-1} + \mu,$$

где ε_t – белый шум; μ – среднее значение; θ – некоторый числовой коэффициент. У этого процесса зависимыми являются только соседние величины $x(t-1)$ и $x(t)$.

Процессом авторегрессии (первого порядка) (сокращенно AR1, от англ. autoregression) называют процесс

$$x(t) - \mu = \varphi(x(t-1) - \mu) + \varepsilon_t,$$

где φ – постоянный коэффициент, не равный нулю. В этой модели прослеживается зависимость элементов ряда от всех предыдущих: $x(t)$ зависит от ε_t и $x(t-1)$, $x(t-1)$ зависит от ε_{t-1} и $x(t-2)$, $x(t-2)$ зависит от ε_{t-2} и $x(t-3)$.

Коэффициент φ определяет глубину зависимости процесса от предыдущих членов временного ряда. При $|\varphi| < 1 \Rightarrow$ зависимость быстро убывает. Как правило, именно такие процессы встречаются в прикладных задачах.

3.3. Численные характеристики временных рядов

Числовые характеристики временных рядов вводятся в полной аналогии с числовыми характеристиками случайных величин.

Математическое ожидание случайного процесса $X(t)$ – это функция $m(t)$ такая, что для каждого момента t значение функции $m(t)$ является математическим ожиданием случайной величины $X(t)$:

$$m(t) = M[x(t)].$$

Функцию $m(t)$ называют также средним значением процесса $X(t)$. Она используется для описания систематического изменения процесса. Например, для модели (3.1) среднее значение равно $tr_t + s_t + c_t$, т.е. под усреднением понимается усреднение случайной величины $X(t)$ при неизменном t , а не усреднение по времени.

Ковариационная функция случайного процесса – это величина

$$B(t, s) = \text{cov}(X(t), X(s)) = M[(X(t) - m(t))(X(s) - m(s))],$$

где t, s – моменты времени.

Значение ковариационной функции при $s = t$ задает дисперсию случайного процесса:

$$B(t, t) = \text{cov}(X(t), X(t)) = D[X(t)].$$

Квадратный корень из дисперсии называют стандартным отклонением случайного процесса:

$$\sigma(t) = \sqrt{\text{cov}(X(t), X(t))}.$$

Корреляционная функция случайного процесса – это величина

$$\text{corr}[X(t), X(s)] = \frac{\text{cov}[X(t), X(s)]}{\sigma(t)\sigma(s)}.$$

При фиксированных t и s корреляционная функция является коэффициентом корреляции.

Из определения стационарности, данного в подразд. 3.2, следует, что для любых t, s и τ :

$$\begin{aligned} m(t + \tau) &= m(t), \\ B(s + \tau, t + \tau) &= B(s, t). \end{aligned}$$

Примем, что $\tau = -t$, тогда справедливо, что

$$m(t) = m(0) \text{ и } B(s, t) = B(s - t, 0).$$

Отсюда следует, что у стационарного процесса функции $m(t)$ и $\sigma(t)$ постоянны, а ковариационная функция $B(s, t)$ зависит не от пары t, s как в общем случае, а от $|s - t|$. Также можно убедиться, что и корреляционная функция стационарного процесса является функцией от $|s - t|$.

Рассмотрим неравенство $t = s + k, k > 0$. В этом случае *автокорреляционная функция* стационарного процесса $X(t)$ – это корреляционная функция:

$$r(k) = \text{corr}(X(t), X(t + k)),$$

где k (целое) называют *задержкой* или *лагом*, т.е. расстоянием между членами ряда, для которых вычисляется коэффициент корреляции.

Для получения оценок указанных характеристик в общем случае требуется совокупность независимых реализаций процесса $X(t)$, но чаще всего (ввиду невозможности повторения процесса) оценки приходится определять по одной реализации случайного процесса. Доверять таким оценкам можно только для стационарных процессов.

Получим оценку среднего значения. Пусть имеем x_1, x_2, \dots, x_n – стационарный процесс.

Среднее по реализации

$$\bar{m} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Чтобы оценка была достаточно точной, число n должно быть больше периода, на котором сохраняется корреляция.

Рассмотрим выборочную автокорреляционную функцию. Образует для временного ряда x_1, x_2, \dots, x_n совокупность из $(n-1)$ пар:

$$(x_1, x_2), (x_2, x_3), \dots, (x_{n-1}, x_n).$$

Первый элемент пары (в силу стационарности) можно рассматривать как реализацию случайной величины $X(t)$, а второй – как реализацию $X(t+1)$.

Средние значения $X(t)$ и $X(t+1)$ будут соответственно

$$\bar{x}_{(1)} = \frac{1}{n-1} \sum_{t=1}^{n-1} x_t, \quad \bar{x}_{(2)} = \frac{1}{n-1} \sum_{t=2}^{n-1} x_t.$$

Выборочное значение ковариации можно определить следующим образом:

$$\frac{1}{n-1} \sum_{t=1}^{n-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)}).$$

Оценка коэффициента корреляции между $X(t)$ и $X(t+1)$ может быть записана в виде

$$\bar{r}(k=1) = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\sum_{t=1}^{n-1} (x_t - \bar{x}_{(1)})^2 \cdot \sum_{t=1}^{n-1} (x_{t+1} - \bar{x}_{(2)})^2}}.$$

При больших значениях n , учитывая, что $\bar{x}_{(1)} \approx \bar{x}_{(2)} = \bar{x}$, последнее выражение можно заменить более простым:

$$\bar{r}(k=1) = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}.$$

Аналогично можно получить оценку корреляции между $X(t)$ и $X(t+k)$ для произвольного k :

$$\bar{r}(k) = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}.$$

Функцию $\bar{r}(k)$ аргумента k ($k = 1, 2, \dots$) называют выборочной автокорреляционной функцией. В практических случаях следует учитывать, что точность ее оценки заметно снижается с ростом лага по причине уменьшения числа наблюдений, используемых при вычислении оценки. Поэтому обычно ограничиваются изучением небольшого числа первых членов автокорреляционной функции.

График выборочной корреляционной функции называют коррелограммой. Она позволяет выявлять некоторые свойства временного ряда, например, является ли он белым шумом (у белого шума $r(k \neq 0) = 0$), содержится ли в этом ряде тренд (при наличии тренда график $r(k)$ незатухающий) или сезонные колебания (в этом случае коррелограмма будет иметь всплески).

4. ПРАКТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Практический анализ временных рядов проводится с помощью специальных компьютерных программ и имеет целью построение математической модели ряда, с помощью которой можно восстановить недостающие участки ряда и осуществить прогноз его дальнейшего поведения.

Анализ обычно состоит из следующих этапов:

- построения и изучения графика;
- сведения ряда к стационарному;
- подбора модели случайной составляющей ряда.

Перечисленные этапы могут повторяться многократно. Модель считается подобранной, если остаточный компонент может считаться «белым шумом». После подбора модели оценивается дисперсия остатков (для построения доверительных интервалов прогноза) и проверяется адекватность модели.

Используются следующие основные группы методов анализа временных рядов:

- графические методы;
- методы сведения к стационарным процессам;
- методы исследования внутренних связей между элементами временных рядов.

Графические методы играют важную роль в анализе временных рядов, так как по графику можно определить, (в отличие от табличного представления временного ряда) наличие тренда и его характер, наличие сезонных и циклических компонентов, а также корреляции между соседними элементами ряда.

4.1. Выделение тренда

Основным методом выделения тренда является метод аналитического выравнивания временных рядов. Основным содержанием этого метода является расчет общей тенденции развития (тренда) как функции времени на основе математической модели, которая наилучшим образом отображает основную тенденцию развития временного ряда.

В качестве моделей обычно используют линейную, показательную, степенную и логарифмическую функции в зависимости от особенностей временного ряда. Для оценки точности трендовой модели используется коэффициент детерминации, равный отношению дисперсии теоретических данных, полученных по трендовой модели к дисперсии эмпирических данных. Считается, что трендовая модель адекватна изучаемому процессу и отражает тенденцию его развития при значениях коэффициента детерминации, близких к единице.

Математической основой аналитического выравнивания является метод наименьших квадратов. Согласно терминологии регрессионного анализа (подробно регрессионный анализ будет рассматриваться позднее) значения временного ряда x_t рассматривают как отклик (зависимую переменную), время t – как фактор, влияющий на отклик (независимую переменную):

$$x_{ti} = f(t_i, \theta) + \varepsilon_i,$$

где f – функция тренда; θ – вектор параметров; ε – независимые и одинаково распределенные случайные величины.

Согласно методу наименьших квадратов модель подбирается в соответствии со следующим критерием:

$$\sum_{i=1}^n ((x_i - f(t, \theta))^2) \rightarrow \min.$$

Этот метод дает хорошие оценки при независимых остатках.

Основные модели тренда были приведены в подразд. 3.1. Рассмотрим самую простую модель – линейную, в этом случае тренд может быть задан прямой линией:

$$tr_t = \theta_0 + \theta_1 t = f(t, \theta_0, \theta_1).$$

По методу наименьших квадратов можно получить оценки коэффициентов модели (независимой переменной является время t):

$$\begin{aligned} \hat{\theta}_0 &= \bar{x} = \frac{1}{n} \sum_{t=1}^n x_t; \\ \hat{\theta}_1 &= \frac{\sum_{t=1}^n (x_t - \bar{x})(t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2}, \end{aligned}$$

где

$$\bar{t} = \frac{t(t+1)}{2}.$$

4.2. Выделение сезонных эффектов на фоне тренда

Предположим, что ряд состоит из тренда, сезонной и случайной составляющих:

$$x_t = tr_t + s_t + \varepsilon_t.$$

Пусть p – период последовательности сезонного компонента s_t , так что $s_t = s_{t+p}$ для любого момента t .

Требуется оценить значения s_t по наблюдениям x_t при известном значении периода (обычно по графику визуально нетрудно определить период сезонного компонента).

Сначала необходимо оценить тренд. В рамках описываемой задачи предполагаем, что тренд выделен и получена некая оценка тренда как некоторой гладкой функции.

Для каждого сезона i ($1 \leq i \leq p$) рассмотрим все относящиеся к нему разности $x_i - \widehat{tr}_i, x_{i+p} - \widehat{tr}_{i+p}, \dots, x_{i+mp} - \widehat{tr}_{i+mp}$, характеризующие влияние сезонных изменений.

Поскольку сказываются случайные составляющие, то для определения сезонного влияния разности надо усреднить.

В качестве простейшей модели можно использовать простое среднее:

$$\widehat{s}_i = \frac{1}{m+1} \sum_{l=0}^m (x_{i+lp} - \widehat{tr}_{i+lp}), \quad i = 1, 2, \dots, p.$$

В качестве другой оценки можно взять взвешенной среднее или медиану.

После получения оценки сезонных эффектов в модели можно провести удаление этих эффектов из рассматриваемого ряда путем вычитания их из начальных значений ряда. Такая процедура удаления s_i из начального ряда называется сезонным выравниванием (коррекцией, декомпозицией).

На практике считается, что оценки сезонных эффектов недостаточно точны, если число периодов в исследуемом сезонном временном ряде меньше пяти. Это означает, например, что при рассмотрении месячных данных для достаточно точной оценки сезонных эффектов необходимы, как минимум, наблюдения за 5–6 лет.

4.3. Метод скользящего среднего

При наличии циклического компонента подход к расчету сезонных эффектов другой, так как при рассмотренном варианте необходимо оценить не только тренд, но и циклическую составляющую. Проще всего одновременно учесть

тренд и циклическую составляющую с помощью метода скользящего среднего. Этот метод также полезен, если не ясна модель тренда.

Метод скользящего среднего заключается в переходе от начальных значений ряда к их средним значениям на заранее известном интервале времени, который скользит вдоль ряда. Каждый член ряда заменяется простым или взвешенным средним m (ширина окна) соседних членов.

Ряд получается более гладким и показывает тенденцию его поведения, т.е. эта процедура отбрасывает сезонный компонент, но для этого интервал нужно сделать кратным периоду сезонности. Таким образом, если интервал сделать кратным периоду сезонной составляющей, то она устраняется.

Простое скользящее среднее определяется как среднее арифметическое значение по следующей формуле:

$$x_t = \frac{1}{m} \sum_{i=t-p}^{t+p} x_i,$$

где x_i – фактическое значение i -го уровня; m – число уровней, входящих в интервал сглаживания ($m = 2p + 1$); x_t – текущий уровень ряда динамики; i – порядковый номер уровня в интервале сглаживания; p – при нечетном m имеет значение $p = (m - 1) / 2$.

Интервал сглаживания, т.е. число входящих в него уровней m , определяют по следующему правилу: когда необходимо сгладить незначительные, беспорядочные колебания, интервал сглаживания берут большим, если требуется сохранить более незначительные колебания и освободиться от периодически повторяющихся выбросов, то интервал сглаживания уменьшают.

Пример 4.1. Дан временной ряд (рис.4.1).

Требуется определить интервал сглаживания.

Решение. Из исходных данных видно, что период сезонности равен трем, поэтому если стараться делать интервал сглаживания кратным периоду сезонности, то подходящим значением будет $m=3$.

По методу скользящего среднего разделим весь временной отрезок на интервалы и на них вычислим среднее значение.

Для первого, второго и третьего интервалов:

$$(1+4+2)/3=2,33;$$

для второго, третьего и четвертого интервалов:

$$(4+2+5)/3=3,67;$$

для третьего, четвертого и пятого интервалов:

$$(2+5+1)/3=2,67;$$

для четвертого, пятого и шестого интервалов:

$$(5+1+2)/3=2,67;$$

для пятого, шестого и седьмого интервалов:

$$(1+2+5)/3=2,67.$$

По полученным средним значениям построим сглаженный ряд (на рис. 4.1 изображен серым цветом).

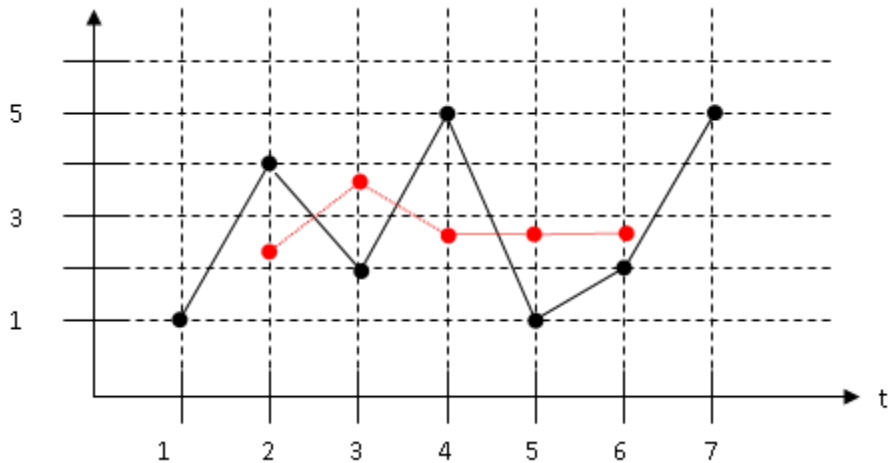


Рис 4.1. Пример скользящего среднего

Пример 4.2. На рис. 4.2 изображен фрагмент трафика телекоммуникационного канала.

Требуется определить интервал сглаживания.

Решение. График характеризует интенсивность получения информации по данному каналу. Из графика видно, что в отдельные дни недели происходит уменьшение загрузки канала, а в другие дни загрузка повышается. Кроме того, вероятно, имеет место плавный рост объема к концу месяца. Таким образом, можно предположить, что временной ряд имеет тренд и сезонный компонент с периодом сезонности (P), равным 7 дней. Тогда значение интервала сглаживания можно принять равным периоду, т.е. $m = 7$.

Первая точка сглаженного ряда получается с учетом того, что $i = 4$ (порядковый номер уровня в интервале сглаживания):

$$\hat{x}_{t1} = \frac{1}{7}(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7),$$

Вторая точка сглаженного ряда ($i = 5$)

$$\hat{x}_{t2} = \frac{1}{7}(x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8)$$

и т.д.

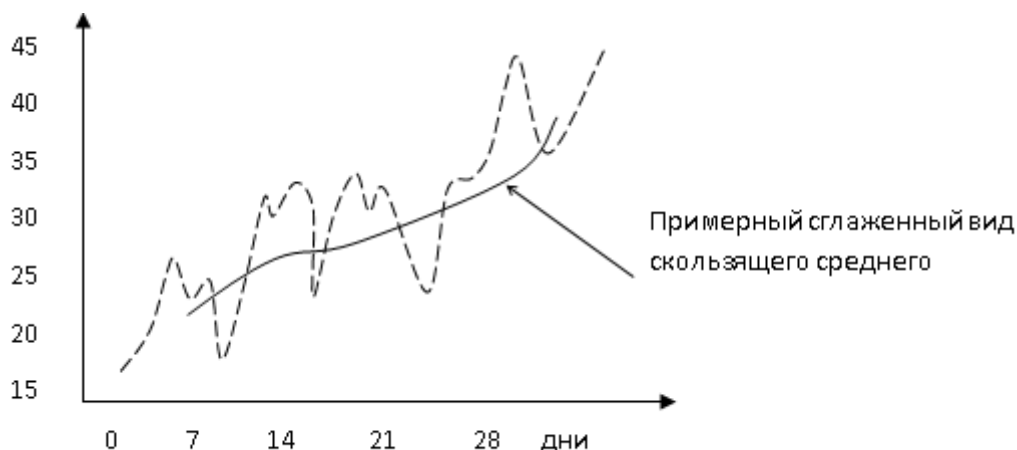


Рис. 4.1. Загрузка информационного канала

5. ИССЛЕДОВАНИЕ СТРУКТУРЫ СТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА

5.1. Цели и методы анализа

Периодическая зависимость представляет собой общий тип компонентов временного ряда. Каждое наблюдение очень похоже на соседнее, и имеется повторяющаяся периодическая составляющая, это означает, что каждое наблюдение также похоже на наблюдение, имевшееся в то же самое время период назад.

В общем, периодическая зависимость может быть формально определена как корреляционная зависимость порядка k между каждым i -м элементом ряда и $(i-k)$ -м элементом. Ее можно измерить с помощью автокорреляции (т.е. корреляции между самими членами ряда).

Величину k обычно называют лагом (иногда используют эквивалентные термины: “сдвиг”, “запаздывание”). Если ошибка измерения не слишком большая, то периодичность можно определить визуально, рассматривая поведение членов ряда через каждые k временных единиц.

После удаления детерминированной составляющей временной ряд должен свестись к стационарному процессу. Следующим шагом должно быть исследование ряда остатков, т.е. изучение ряда, полученного из исходного временного ряда после исключения детерминированного компонента.

Цели исследования:

- проверка стационарности остатков (при ее нарушении детерминированная часть нуждается в уточнении);
- уточнение оценки дисперсии временного ряда (важно для прогнозирования);

– подбор моделей стационарного ряда, которые отражают зависимость между элементами (на базе модели можно осуществить прогноз будущего поведения ряда).

Обычно в качестве модели стационарных временных рядов чаще всего используются процессы авторегрессии, скользящего среднего и их комбинации.

Проверка стационарности ряда остатков осуществляется с помощью *коррелограммы* – графика выборочной автокорреляционной функции.

Коррелограмма (автокоррелограмма) показывает графически автокорреляционную функцию (АКФ), т.е. коэффициенты автокорреляции для последовательности лагов из определенного диапазона.

На коррелограмме обычно отмечается диапазон в размере двух стандартных ошибок на каждом лаге, однако обычно величина автокорреляции более интересна, чем ее надежность, потому что интерес в основном представляют очень сильные, а следовательно, и значимые автокорреляции.

При изучении коррелограмм следует помнить, что автокорреляции последовательных лагов формально зависимы между собой. Например, если первый член ряда тесно связан со вторым, а второй – с третьим, то первый элемент должен также каким-то образом зависеть от третьего и т.д. Это приводит к тому, что периодическая зависимость может существенно измениться после удаления автокорреляций первого порядка, (т.е. после взятия разности с лагом 1).

5.2. Интерпретация графика коррелограммы

Последовательность коэффициентов корреляции r_k , где $k = 1, 2, \dots, n$, как функция интервала k между наблюдениями называется автокорреляционной функцией (АКФ). Вид выборочной автокорреляционной функции тесно связан со структурой ряда.

Рассмотрим, как выглядят коррелограммы для некоторых классов нестационарных временных рядов. Следует отметить, что коррелограмма практически не несет информации о статистической зависимости или независимости членов временного ряда, однако она может отражать причины нарушения стационарности.

На рис. 5.1 представлен график АКФ, характеризующегося умеренным трендом и неявно выраженной сезонностью.

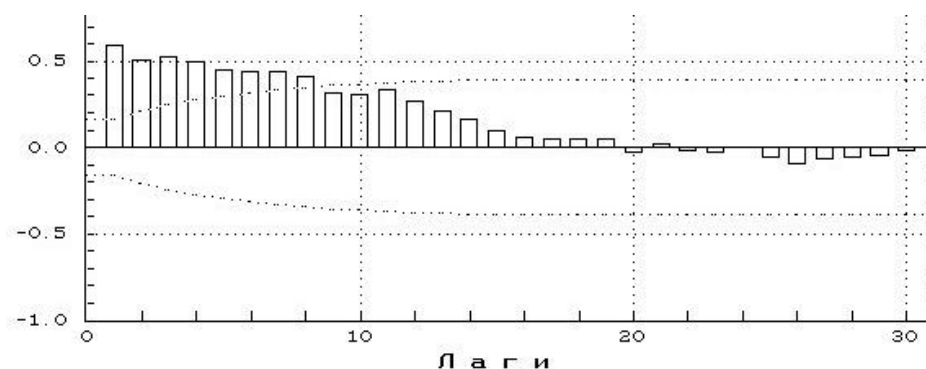


Рис 5.1. Умеренный тренд и неясно выраженная сезонность

Рис. 5.2 демонстрирует АКФ ряда, характеризующегося выраженной сезонной детерминантой. Если есть сезонная составляющая, то коррелограмма имеет периодические всплески, соответствующие периоду сезонных колебаний. Это дает возможность установить предполагаемый период сезонности.

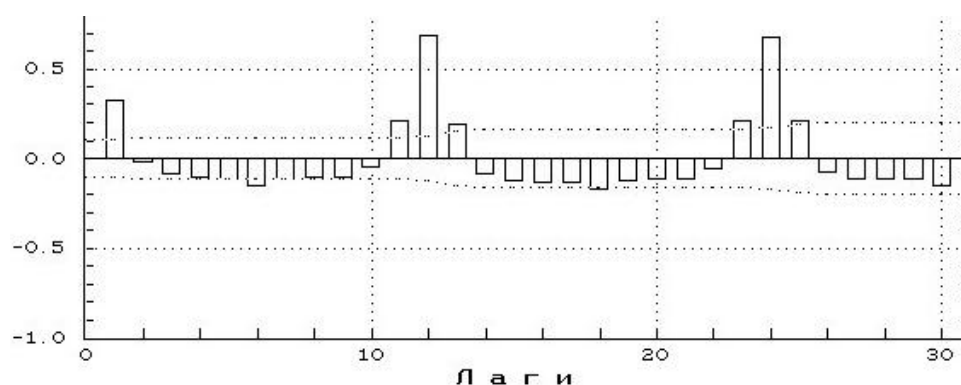


Рис 5.2. Наличие сезонной составляющей

Практически незатухающий график АКФ ряда (рис. 5.3) свидетельствует о наличии отчетливого тренда.

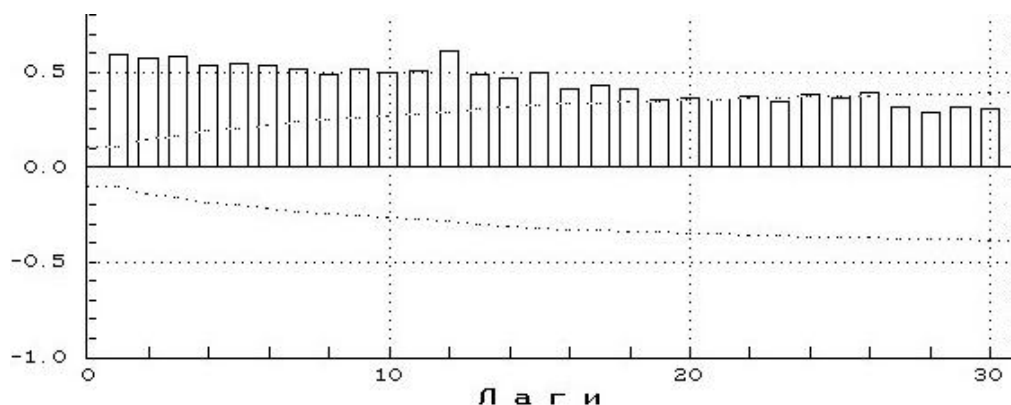


Рис 5.3. Наличие отчетливого тренда

Рассмотрим коррелограммы стационарных случайных процессов. В случае стационарных временных рядов коррелограмма показывает коррелированность значений ряда при различных расстояниях между ними.

5.2.1. Коррелограмма белого шума

Для белого шума автокорреляционная функция r_k равна нулю для всех $k \neq 0$. При ограниченной длительности временного ряда коррелограмма не будет нулевой (при $k \neq 0$), но значения r_k практически не будут выходить за 95%-ный доверительный интервал, который на рис. 5.4 изображен пунктирной линией.

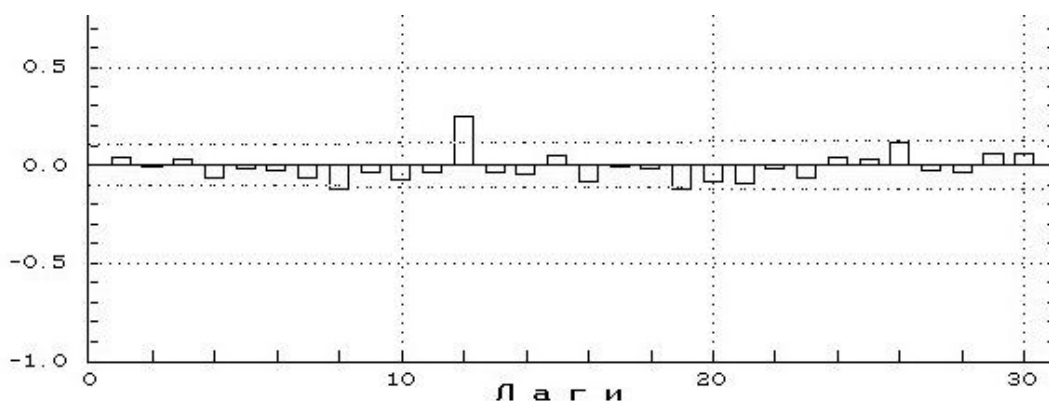


Рис.5.4. Выборочная коррелограмма «белого шума»

Если выборочные оценки автокорреляционной функции попадают в указанные интервалы, то можно предположить, что значения процесса являются белым шумом. При наличии относительно небольшого числа наблюдений одно или несколько значений могут выходить из указанных пределов.

5.2.2. Коррелограмма процессов скользящего среднего

Графики многих стационарных процессов являются более гладкими, чем график белого шума. Это связано с наличием положительной корреляции между соседними членами ряда. При отрицательной корреляции ряд получается более изломанным, чем траектория белого шума. Простейшим примером процессов, у которых зависимы одно или несколько соседних значений, являются процессы скользящего среднего.

Пусть $\varepsilon_1, \dots, \varepsilon_n$ – гауссовский белый шум.

Обозначим $X(t)$ процесс скользящего среднего первого порядка (MA1):

$$X(t) = \varepsilon_t + \theta \varepsilon_{t-1}.$$

У данного процесса зависят между собой только соседние значения $X(t)$ и $X(t-1)$. Причем их переменная (автокорреляционная функция $r(1) = r_1$)

$$r_1 = \frac{M[x(t) \cdot x(t-1)]}{\sigma[x(t)]\sigma[x(t-1)]} = \frac{M[(\varepsilon_1 + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})]}{\sqrt{\sigma^2 + \theta^2\sigma^2}\sqrt{\sigma^2 + \theta^2\sigma^2}} =$$

$$\frac{M[\varepsilon_t\varepsilon_{t-1}] + M[\varepsilon_t\theta\varepsilon_{t-2}] + M[\theta\varepsilon_{t-1}^2] + M[\theta\varepsilon_{t-1}\theta\varepsilon_{t-2}]}{\sigma^2 + \theta^2\sigma^2} = \frac{\theta\sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta}{1 + \theta^2}.$$

Нетрудно получить, что $r_2 = r_3 = \dots = 0$.

По выборочному значению r_1 можно найти оценку θ и тем самым задать модель.

Для процесса скользящего среднего второго порядка отличаются от нуля только r_1 и r_2 . Для процессов порядка q отличны от нуля только первые q значений r_1, r_2, \dots, r_q . Это дает возможность определить модели скользящего среднего:

$$x(t) = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} - \text{2-го порядка};$$

$$x(t) = \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q} - \text{q-го порядка};$$

$$\begin{aligned} \sigma_k = B(k) &= \text{cov}(x(t), x(t+k)) = M[x(t) \cdot x(t+k)] = \dots \\ &= \sigma^2(\theta_k + \theta_1\theta_{k-1} + \theta_2\theta_{k-2} + \dots + \theta_q\theta_{k-q}); \\ \sigma_0 = D[x(t)] &= \sigma^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2). \end{aligned}$$

Тогда

$$r_k = \frac{\sigma_k}{\sigma_0}.$$

Эти уравнения позволяют определить оценку θ .

Если значение q больше пяти, то модель скользящего среднего считается многопараметрической, при этом точность оценок параметров ухудшается, и эти модели становятся малоприменимыми и лучше попытаться описать ряд с помощью модели авторегрессии.

5.2.3. Коррелограмма процессов авторегрессии

Пусть $\varepsilon_1, \dots, \varepsilon_n$ – белый шум.

Модель авторегрессии первого порядка (AR(1)) при нулевом среднем имеет следующий вид:

$$x(t) = \varphi \cdot x(t-1) + \varepsilon_t.$$

Определим для этой модели r_k :

$$\sigma_0 = \frac{\sigma^2}{1 - \varphi^2};$$

$$\sigma_1 = M[x(t)x(t+1)] = M[x(t)(\varphi x(t) + \varepsilon_{t+1})] = \varphi M[x(t)^2] + M[x(t)\varepsilon_{t+1}] = \varphi \cdot \sigma_0;$$

$$\begin{aligned}\varepsilon_2 &= M[x(t) \cdot x(t-2)] = M[x(t)(\varphi \cdot x(t+1) + \varepsilon_{t+2})] = \\ &= M[x(t)(\varphi^2 \cdot x(t) + \varphi \varepsilon_{t+1} + \varepsilon_{t+2})] = \varphi^2 \varepsilon_0\end{aligned}$$

и т.д. для любого k

$$\varepsilon_k = \varphi^k \varepsilon_0.$$

Тогда

$$r_k = \frac{\varepsilon_k}{\varepsilon_0} = \varphi^k.$$

Таким образом, из последнего равенства следует, что автокорреляционная функция убывает с ростом лага k .

При $0 < \varphi < 1$ коррелограмма отражает показательное убывание корреляций с возрастанием интервала между наблюдениями; при $-1 < \varphi < 0$ коррелограмма имеет характер затухающей косинусоиды.

Сравним характерное поведение коррелограмм стационарного процесса AR(1), например при $\varphi = \pm 0.8$ (рис. 5.5).

По самому ряду x_1, x_2, \dots, x_n можно построить выборочную автокорреляционную функцию, в частности r_1 , и тем самым получить оценку параметра φ , которая будет равна выборочной функции r_1 .

Таким образом, по виду коррелограммы и значению выборочной функции r_1 можно задать модель авторегрессии первого порядка.

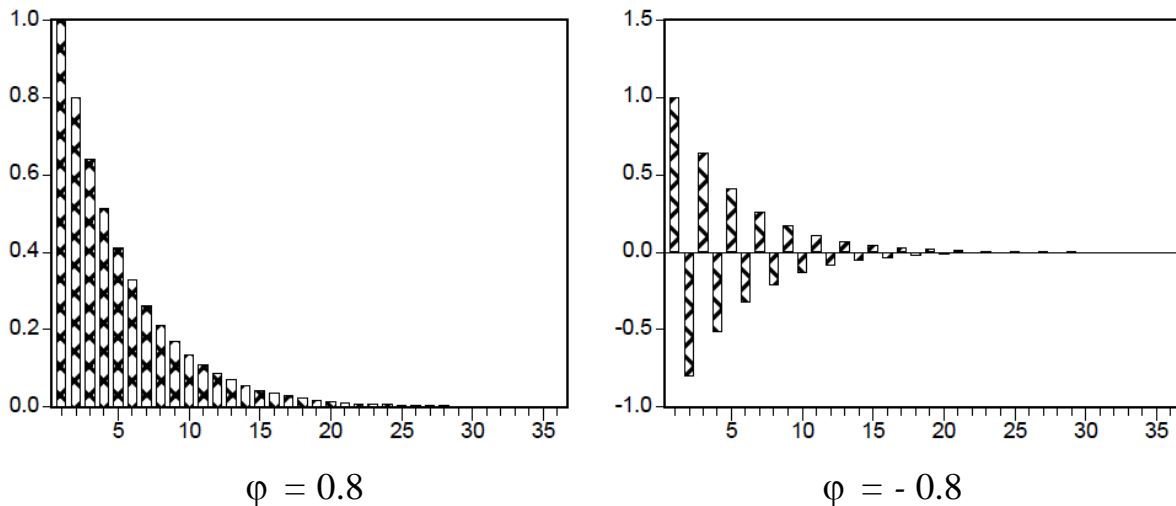


Рис. 5.5. Коррелограммы процесса AR(1)

Процесс $X(t)$ называют процессом авторегрессии второго порядка (AR(2)), если выполняется соотношение

$$X(t) = \varphi_1 X(t-1) + \varphi_2 X(t-2) + \varepsilon_t,$$

где ε_t – белый шум; φ_1, φ_2 – некоторые константы.

Текущее значение процесса AR(2) в момент времени t формируется как линейная комбинация его значений в предыдущие моменты $(t-1)$ и $(t-2)$ и независимой от них случайной величины ε_t .

Чтобы процесс был стационарным, необходимо и достаточно, чтобы выполнялась следующая система неравенств:

$$\begin{cases} \varphi_1 + \varphi_2 < 1; \\ \varphi_1 - \varphi_2 > -1; \\ \varphi_2 > -1. \end{cases}$$

Эти неравенства определяют область возможных значений φ_1 и φ_2 . Графически данную область можно представить в виде треугольника.

Определим для рассматриваемой модели r_k :

$$\begin{aligned} \varphi_1 &= M[X(t)X(t-1)] = M[X(t-1)(\varphi_1 \cdot X(t-1) + \varphi_2 X(t-2) + \varepsilon_t)] = \varphi_1 \varphi_0 + \varphi_2 \varphi_1; \\ \varphi_2 &= \varphi_1 \varphi_1 + \varphi_2 \varphi_0, \end{aligned}$$

Тогда

$$\begin{cases} r_1 = \varphi_1 + \varphi_2 r_1; \\ r_2 = \varphi_1 r_1 + \varphi_2. \end{cases} \quad (5.1)$$

Полученную систему (5.1) называют системой уравнений Юла–Уолкера. Эти уравнения связывают параметры процесса AR(2) со значениями его автокорреляционной функции:

$$\begin{cases} \varphi_1 = \frac{r_1 - r_1 \cdot r_2}{1 - r_1^2}; \\ \varphi_2 = \frac{r_2 - r_1^2}{1 - r_1^2}. \end{cases}$$

В общем случае для любого k можно получить

$$r_k = \varphi_1 r_{k-1} + \varphi_2 r_{k-2}.$$

Исследование уравнений показывает, что автокорреляционная функция процесса является суммой затухающих экспонент и затухающих синусоидальных волн. Причем это верно для процесса авторегрессии более высокого порядка.

Пример 5.1. Рассмотрим процесс авторегрессии AR(2):

$$X(t) = 4.375 + 0.25X(t-1) - 0.125X(t-2) + \varepsilon_t.$$

Решение. Для построения коррелограммы воспользуемся уравнениями Юла – Уокера.

Имеем

$$r_K = 0,25 \cdot r_{K-1} + 0,125 \cdot r_{K-2}.$$

По определению, $r_0 = 1$.

Для r_1 имеем

$$r_1 = 0,25 \cdot r_0 + 0,125 \cdot r_{-1} = 0,25 - 0,125 \cdot r_1.$$

Отсюда находим:

$$r_1 = 0,25 / (1 + 0,125) = 0,22.$$

Далее последовательно определяем:

$$r_2 = 0,25 \cdot r_1 - 0,125 \cdot r_0 = 0,25 \cdot 0,22 - 0,125 = -0,069,$$

$$r_3 = -0,045, r_4 = -0,003, r_5 = 0,005$$

и т.д.

Из полученных значений видно, что корреляции даже между соседними наблюдениями очень малы, и можно ожидать, что поведение траекторий такого ряда не очень существенно отличается от поведения реализаций процесса белого шума.

Теоретическая коррелограмма рассматриваемого процесса приведена на рис. 5.6.

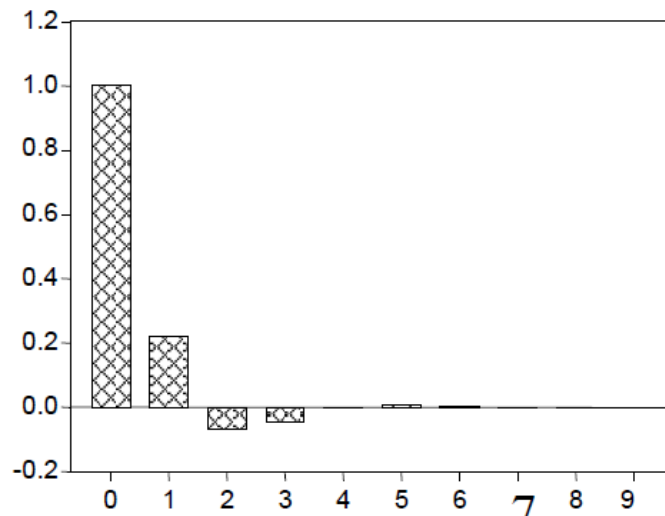


Рис. 5.6. Итоговая коррелограмма

Обобщим полученные результаты для процессов более высоких порядков.

Случайный процесс $X(t)$ с нулевым средним называется процессом авторегрессии порядка (AR(p)) если для него выполняется следующее соотношение:

$$X(t) = \varphi_1 X(t-1) + \varphi_2 X(t-2) + \dots + \varphi_p X(t-p) + \varepsilon_t.$$

Можно убедиться, что

$$r_k = \varphi_1 r_{k-1} + \varphi_2 r_{k-2} + \dots + \varphi_p r_{k-p}.$$

Если порядок процесса авторегрессии равен p , то коэффициенты φ определяются из следующей системы уравнений:

$$\begin{cases} r_1 = \varphi_1 + \varphi_2 r_1 + \dots + \varphi_p r_{p-1}; \\ r_2 = \varphi_1 r_1 + \varphi_2 + \dots + \varphi_p r_{p-2}; \\ \vdots \\ r_p = \varphi_1 r_{p-1} + \varphi_2 r_{p-2} + \dots + \varphi_p. \end{cases}$$

Трудность состоит в том, что по коррелограмме трудно определить порядок. Программы по исследованию рядов помогают пользователю в определении порядка следующим образом. Решается система Юла–Уолкера для разных периодов и строится график функции $\varphi_p(p)$, $p = 1, 2, \dots$. Эта функция называется частной автокорреляционной функцией (ЧАКФ). Главное свойство этой функции состоит в том, что если процесс соответствует модели AR(P), то элементы ЧАКФ равны нулю при $q > p$.

Процесс подбора модели продолжается до тех пор, пока после удаления из стационарного ряда смоделированного ряда не получится «белый шум».

Кроме вышерассмотренных моделей используют комбинированную модель авторегрессии скользящего среднего (ARMA (p, q)):

$$X(t) = \sum_{i=1}^p \varphi_i X(t-i) + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}.$$

Для анализа временных рядов могут быть использованы следующие программы: Forecast Expert, Медозавр, Олимп, Эвреста. Можно использовать и универсальные: SPSS, STADIA, STATGRAPHICS, STATISTICA, SISTAT.

6. РЕГРЕССИОННЫЕ МОДЕЛИ

Достаточно часто встречаются случаи, когда сложно, а иногда и невозможно теоретически определить зависимость функции выхода от внутренних параметров. В таких случаях решение задачи возможно лишь с помощью эксперимента.

Показатель свойства, для которого нужно построить модель, называют выходным (наблюдаемым) параметром или откликом. Параметры, от которых зависит этот отклик, называются факторами.

Эксперимент, в котором наблюдатель только измеряет значения факторов, называется пассивным.

Эксперимент, в котором наблюдатель задает значения факторов, называется активным.

При активном эксперименте можно составить такие планы, которые упрощают процедуру определения коэффициентов модели. Исходная формула для определения коэффициентов одна та же, но при определенных планах она существенно упрощается настолько, что коэффициенты можно считать вручную. Составлением таких планов занимается теория планирования эксперимента.

6.1. Определение коэффициентов модели по экспериментальным данным

Рассмотрим порядок получения модели при проведении пассивного многофакторного эксперимента.

Задача состоит в определении зависимости выходного параметра от факторов:

$$y \approx f(x_1, x_2, \dots, x_n).$$

Знак приближенного равенства стоит по причине возможного влияния других факторов, не вошедших в функцию f . В таком случае можно записать:

$$y = f(x_1, x_2, \dots, x_n) + \varepsilon,$$

где ε – случайная составляющая.

Стараются строить такие модели, чтобы выполнялось равенство

$$M[y] = f(x_1, x_2, \dots, x_n).$$

Модель, в которой данное равенство выполняется, называется *регрессионной*, а функция f – *функцией регрессии*.

Здесь будем пользоваться понятием «*регрессионная модель*» и в случае, если строгого равенства не получается, иметь в виду стремление к нему.

В общем виде, без каких-либо предположений функцию регрессии определить не удастся. Предполагается, что она имеет вид

$$y \approx b_0 + b_1 f_1(x_i) + b_2 f_2(x_i) + \dots + b_n f_n(x_i) = \sum_{j=0}^n b_j f_j(x_i),$$

где f_j – функции от факторов, причем $f_0 = 1$; b_j – неизвестные коэффициенты.

Если ничего о функциях f_j не известно, то обычно полагают, что достаточно учесть факторы, их квадраты и произведения между собой.

Так, при двух факторах модель может иметь следующий вид:

$$y \approx b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2.$$

Характер зависимостей $f_j(x_i)$ может быть известен, например: $f_1(x_1) = \sin(ax_1)$, $f_2(x_2) = e^{x_2^2}$, $f_3(x_3) = \ln(x_3)$ и т.д.

В итоге задача сводится к определению неизвестных коэффициентов b_j .

Результаты эксперимента сводят в таблицу (табл.6.1).

Наблюдаемыми величинами являются факторы и выходной параметр y .

Функции f_j вычисляются по известной зависимости от x_i .

Коэффициенты b_j должны быть такими, чтобы модельные значения y (вычисляемые по модели) мало отличались от наблюдаемых.

Таблица 6.1

Опыт	Факторы				Функции f_j				y
	x_1	x_2	...	x_n	f_1	f_2	...	f_n	
1	x_{11}	x_{12}		x_{1n}	f_{11}	f_{12}		f_{1n}	y_1
2	x_{21}	x_{22}		x_{2n}	f_{21}	f_{22}		f_{2n}	y_2
3	x_{31}	x_{32}		x_{3n}	f_{31}	f_{32}		f_{3n}	y_3
•				•					
•				•					
•				•					
N	x_{N1}	x_{N2}		x_{Nn}	f_{N1}	f_{N2}		f_{Nn}	y_N

Наиболее часто для определения коэффициентов b_j используют критерий минимума суммы квадратов невязок:

$$R = \sum_{i=1}^N \underbrace{\left(y_i - \sum_{j=0}^n \overbrace{b_j f_{ij}}^{\text{оценки}} \right)}_{\text{невязка}}^2 \rightarrow \min .$$

Значения коэффициентов b_j , удовлетворяющих заданному критерию, можно найти из уравнений:

$$\frac{\partial R}{\partial b_0} = 0; \quad \frac{\partial R}{\partial b_1} = 0; \quad \dots \quad \frac{\partial R}{\partial b_n} = 0.$$

В результате дифференцирования получим систему уравнений:

$$\begin{cases} \sum_{i=1}^N \left(y_i - \sum_{j=0}^n b_j f_{ij} \right) f_{i0} = 0; & \sum_{j=0}^n b_j \sum_{i=1}^N f_{ij} f_{i0} = \sum_{i=1}^N f_{i0} y_i; \\ \sum_{i=1}^N \left(y_i - \sum_{j=0}^n b_j f_{ij} \right) f_{i1} = 0; & \dots \\ \sum_{i=1}^N \left(y_i - \sum_{j=0}^n b_j f_{ij} \right) f_{in} = 0; & \sum_{j=0}^n b_j \sum_{i=1}^N f_{ij} f_{in} = \sum_{i=1}^N f_{in} y_i. \end{cases}$$

Введем обозначения:

$$\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} = B \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = Y \begin{pmatrix} f_{10} & f_{11} & \dots & f_{1n} \\ f_{20} & f_{21} & \dots & f_{2n} \\ \vdots & \vdots & \dots & \vdots \\ f_{N0} & f_{N1} & \dots & f_{Nn} \end{pmatrix} = (F_0 F_1 \dots F_n) = F.$$

В новых обозначениях последнюю систему уравнений можно записать в виде

$$\begin{cases} \sum_{j=0}^n b_j \cdot F_0^T \cdot F_j = F_0^T Y; \\ \dots \\ \sum_{j=0}^n b_j \cdot F_n^T \cdot F_j = F_n^T Y; \end{cases} \quad \begin{cases} F_0^T \sum_{j=0}^n b_j \cdot F_j = Y F_0^T; \\ \dots \\ F_n^T \sum_{j=0}^n b_j \cdot F_j = Y F_n^T; \end{cases} \quad \begin{cases} F_0^T \cdot F \cdot B = Y \cdot F_0^T; \\ \dots \\ F_n^T \cdot F \cdot B = Y \cdot F_n^T, \end{cases},$$

или в матричном виде $F^T F B = F^T Y$. Отсюда получим вектор искомых коэффициентов

$$B = (F^T F)^{-1} F^T Y,$$

где $F^T F$ – информационная матрица.

Пример. 6.1. Требуется определить коэффициенты модели

$$y \approx b_0 + b_1 x = b_0 f_0 + b_1 f_1$$

при результатах наблюдений, приведенных в табл. 6.2.

Таблица 6.2

x	f_0	f_1	y
0	1	0	1
0	1	0	2
1	1	1	3
1	1	1	4

Решение. Исходя из таблицы исходных данных, матрица значений базовых функций, ее транспонированный вид и информационная матрица будут равны соответственно:

$$F = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}; \quad F^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix};$$

$$F^T \cdot F = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}.$$

Проведя несложные матричные операции, получим:

$$(F^T \cdot F)^{-1} = \frac{1}{|F^T \cdot F|} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1 \end{pmatrix};$$

$$F^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 10 \\ 7 \end{pmatrix}.$$

В итоге вектор искомых коэффициентов будет иметь вид:

$$B = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1 \end{pmatrix} \begin{pmatrix} 10 \\ 7 \end{pmatrix} = \begin{pmatrix} 1,5 \\ 2 \end{pmatrix}.$$

Таким образом, искомая модель имеет вид $y = 1,5 + 2x$.

6.2. Проверка адекватности модели

Очевидно, что после получения модели необходимо определить, насколько полученная зависимость отражает реальную систему или исследуемое явление, другими словами, необходимо проверить адекватность модели. При этом нужно учитывать влияние случайных возмущений на выходную величину y . В случае отсутствия таких возмущений опыты полностью повторяемы, т. е. при одном и том же наборе факторов выходная величина должна получиться одной и той же.

Отклонение выходной величины (если есть) объясняется наличием случайных выбросов (промахов), появившихся по причине грубых ошибок при измерении. В этом случае адекватность модели можно проверить по отклонениям модельных значений от реальных, причем проверить модель необходимо и в других точках измерений, при других значениях факторов (рис. 6.1).

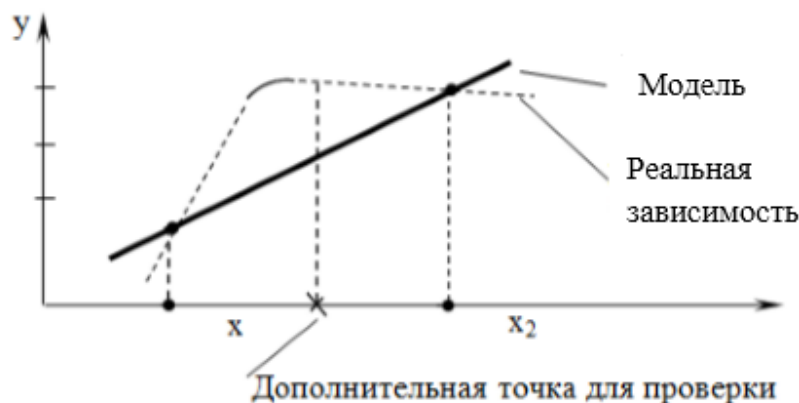


Рис. 6.1. Проверка адекватности

При наличии существенных отклонений обычно повторяют опыты при соответствующих значениях факторов, а также задаются другими значениями факторов из заданного диапазона и сравнивают модельные и реальные значения при этих факторах.

При подтверждении существенных отклонений модель необходимо изменить, причем эксперимент повторять не обязательно и можно использовать для ее построения уже имеющиеся данные. Для принятия решения об адекватности надо иметь в виду, что число опытов должно быть больше числа неизвестных, в противном случае можно принять ошибочное решение. Очевидно, что тем больше число опытов тем лучше для проверки адекватности.

При наличии случайных возмущений для проверки адекватности требуется неоднократное повторение каждого опыта. При этом определяется среднее значение выходной величины для каждого опыта. Отличие средних значений от модельных в какой-то степени характеризует адекватность, но простое сравнение еще не дает основания для принятия решения. Наиболее адекватной является модель, проходящая через математические ожидания $M[y_i]$ т. е. регрессионная модель (рис. 6.2).

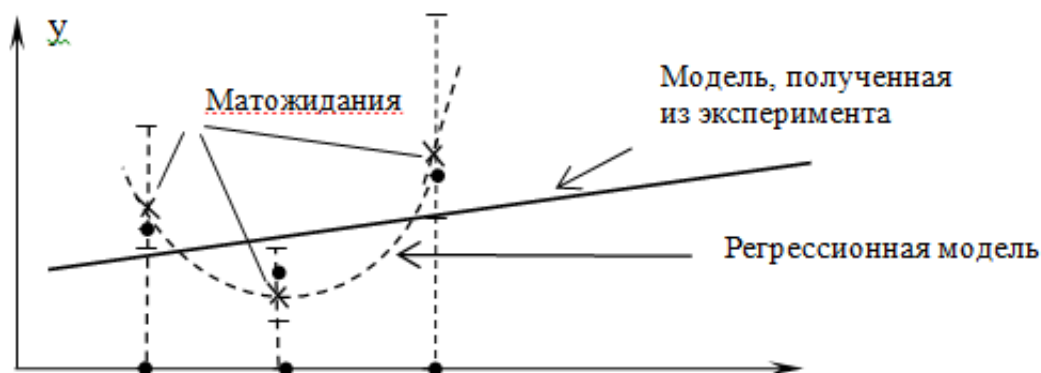


Рис. 6.2. Регрессионная модель

Обычно от модели не требуется, чтобы она точно проходила через математические ожидания, и при проверке адекватности можно ограничиться ответом на вопрос: есть ли основание утверждать, что модель не проходит через математические ожидания выходной величины в точках измерений? Если оснований нет, то модель считается достаточно адекватной. В противном случае модель необходимо пересмотреть.

На поставленный вопрос можно получить ответ при известных дисперсиях выходного параметра относительно математических ожиданий в точках измерений и относительно модели.

Если дисперсии равны, то и точки, относительно которых определяются, дисперсии, близки.

Таким образом, расхождение дисперсий и позволяет ответить на поставленный вопрос, а значит, принять решение об адекватности. Проблема состоит в том, что сами дисперсии не известны, а по опытным данным можно определить только их оценки.

Таким образом, вопрос свелся к следующему: есть ли основание по известным оценкам дисперсий, утверждать, что дисперсии не совпадают?

На поставленный вопрос дается ответ в математической статистике в рамках теории проверки гипотез (при некотором допущении относительно распределения случайной величины y_i). Эта задача решается следующим образом:

- 1) выдвигается гипотеза о том, что дисперсии совпадают;
- 2) рассматривается статистика, равная отношению оценок дисперсий, и для него строится плотность распределения (для нормального распределения случайной величины y_i это распределение Фишера) при условии, что гипотеза справедлива;
- 3) задается уровень значимости α – вероятность того, что гипотеза, при ее верности, ошибочно отвергается. Этой вероятности в области возможных значений отношения оценок дисперсий соответствует так называемая критическая область. Гипотеза отвергается при попадании в критическую область, так как это область больших расхождений (обычно в числителе ставят ту оценку, которая больше);
- 4) по значению отношения оценок дисперсий гипотеза принимается или отвергается:

$$\frac{S_{\text{ост}}^2}{S_{\text{вос}}^2} < F_{1-\alpha}(f_{\text{ост}}, f_{\text{вос}}),$$

где $F_{1-\alpha}(f_{\text{ост}}, f_{\text{вос}})$ – квантиль распределения Фишера.

Таким образом, если неравенство выполняется, то принимается решение об адекватности математической модели (точнее, нет оснований отвергнуть полученную модель).

Оценки дисперсий определяются нижеследующим образом.

Оценка дисперсии относительно математических ожиданий $M[y_i]$ называется *дисперсией воспроизводимости*:

$$S_{\text{вос}}^2 = \frac{1}{f_{\text{вос}}} \sum_{i=1}^N \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2,$$

где y_{ij} – значение выходного параметра в i -й точке при j -м его повторении;

\bar{y}_i – среднее значение выходного параметра в i -й точке, $\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$; m_i – число повторений в i -й точке; N – число различных опытов; $f_{\text{вос}}$ – число степеней свободы.

Число степеней свободы определяется как разность общего числа измерений и числа опытов:

$$f_{\text{вос}} = \sum_{i=1}^N m_i - N.$$

До определения средних значений число степеней свободы равнялось общему числу измерений, затем для определения средних значений их пришлось связать N уравнениями. Осталось $f_{\text{вос}}$ – число степеней свободы.

Дисперсия воспроизводимости характеризует произвольное изменение тех факторов, которые оказывают влияние на выходной параметр, но при испытаниях не измеряются.

Оценка дисперсии относительно модели называется *остаточной дисперсией*:

$$S_{\text{ост}} = \frac{1}{f_{\text{ост}}} \sum_{i=1}^N \sum_{j=1}^{m_i} (y_{ij} - y_{M_i})^2;$$

где $f_{\text{ост}}$ – число степеней свободы, равное разности общего числа измерений и числа неизвестных параметров модели ($n+1$, т. е. столько должно быть уравнений для определения параметров, которые связывают опыты):

$$f_{\text{ост}} = \sum_{i=1}^N m_i - n - 1.$$

Остаточная дисперсия характеризует точность построения математической модели.

Пример. 6.2. Пусть $N = 5$, $k = 3$, $m_1 = \dots = m_5 = 3$. Требуется построить математическую модель.

Решение. Выдвигается гипотеза о виде модели:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2.$$

Значения оценок дисперсии воспроизводимости и остаточной дисперсии равны соответственно:

$$S_{\text{вос}} = \frac{\sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_i)^2}{5 \cdot 3 - 5};$$

$$S_{\text{ост}} = \frac{\sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - y_{M_i})^2}{5 \cdot 3 - 7}.$$

При $\alpha = 0,05$ $F_{1-\alpha}(8,10) = 3,07$.

Если в результате практических расчетов условие

$$\frac{S^2_{\text{ост}}}{S^2_{\text{вос}}} < 3,07$$

выполнится, то принимается решение об адекватности математической модели при выбранном уровне значимости.

7. ПЛАНИРОВАНИЕ АКТИВНЫХ ФАКТОРНЫХ ЭКСПЕРИМЕНТОВ

7.1. Определение коэффициентов модели при ортогональных планах

Планом называют задание количества опытов, числа повторений каждого опыта и задание значений факторов в каждом опыте.

Общим требованием к планам является требование неодинаковости, а точнее, непропорциональности значений факторов. В противном случае нельзя почувствовать их раздельное влияние. Математически это приведет к тому, что при одинаковых функциях от этих факторов, например $f_1 = x_1$, $f_2 = x_2$, определитель $(F^T F) = (X^T X)$ будет равен нулю, и обращаться в этом случае матрицу нельзя.

В активном факторном эксперименте функции f_i равны в основном самим факторам, и в этом случае вектор коэффициентов $B = (F^T F)^{-1} F^T Y$ эквивалентен записи через X , т. е. $B = (X^T X)^{-1} X^T Y$.

При близких значениях факторов определитель будет близок к нулю. Такие матрицы называют плохо обусловленными. Решение в этих случаях будет сильно зависеть от погрешностей задания факторов, т.е. маленькие ошибки в X будут приводить к большим ошибкам в определении коэффициентов вектора B . Поэтому значения разных факторов нужно задавать в разных сочетаниях. В ряде случаев планы могут быть составлены так, что вычисление вектора B значительно упрощается. Рассмотрим такие планы.

Обозначим G_i число различных уровней i -го фактора.

Эксперимент, в котором уровни каждого фактора комбинируют со всеми уровнями других факторов, называется **полным факторным экспериментом** (ПлФЭ).

ПлФЭ записывают в виде $G_1 \times G_2 \times \dots \times G_K$, при одинаковом количестве уровней для каждого фактора ПлФЭ представляют в виде G^K .

Эксперимент, в котором число опытов $N < G_1 \times G_2 \times \dots \times G_K$, называют **неполным (дробным) факторным экспериментом**.

Выбор числа уровней факторов обычно определяется видом математической модели. Например, для линейной модели достаточно варьировать факторы на двух уровнях. Для построения нелинейных моделей число уровней может быть увеличено.

Для построения плана применяют кодированные факторы, меняющиеся от -1 до $+1$, так что при двух уровнях фактор принимает значения -1 и $+1$.

Кодированное значение фактора получается из реального путем преобразования по формуле

$$x_i = \frac{2\chi_i - \chi_{\max} - \chi_{\min}}{\chi_{\max} - \chi_{\min}},$$

где χ_i – значение фактора x_i ; χ_{\min} и χ_{\max} – минимальное и максимальное значения фактора x_i . Покажем, как получена данная формула.

Отрезки $[x, -1]$ и $[\chi, \chi_{\min}]$ должны быть пропорциональны:

$$x - (-1) = \frac{+1 - (-1)}{\chi_{\max} - \chi_{\min}} (\chi - \chi_{\min}).$$

После очевидных преобразований можно перейти к искомому выражению:

$$x = \frac{2}{\chi_{\max} - \chi_{\min}} (\chi - \chi_{\min}) - 1 = \frac{2\chi - \chi_{\max} - \chi_{\min}}{\chi_{\max} - \chi_{\min}},$$

что и требовалось показать.

После построения модели можно перейти к прежним реальным факторам:

$$\chi = \frac{\chi_{\max} - \chi_{\min}}{2} (x + 1) + \chi_{\min}.$$

Пример. 7.1. Рассмотрим полный факторный план 2^2 , т.е. эксперимент из четырех опытов над двумя факторами.

Решение. Таблица плана имеет вид табл. 7.1.

Таблица 7.1

x_0	x_1	x_2
+1	-1	-1
+1	+1	-1
+1	-1	+1
+1	+1	+1

При добавлении третьего фактора таблица удлинится вдвое и может быть построена следующим образом: третий фактор записывается со значением -1 , а затем таблица для двух факторов копируется для значения $+1$ третьего фактора (табл. 7.2).

Таблица 7.2

X₁	X₂	X₃
−1	−1	−1
+1	−1	−1
−1	+1	−1
+1	+1	−1
−1	−1	+1
+1	−1	+1
−1	+1	+1
+1	+1	+1

Из табл. 7.2 видно, что в ней присутствуют все возможные сочетания факторов. Можно план строить так же, как и последовательность целых двоичных чисел.

В итоге получился план 2^3 . Ему соответствует матрица X (табл. 7.3).

Таблица 7.3

X₀	X₁	X₂	X₃
+1	−1	−1	−1
+1	+1	−1	−1
+1	−1	+1	−1
+1	+1	+1	−1
+1	−1	−1	+1
+1	+1	−1	+1
+1	−1	+1	+1
+1	+1	+1	+1

Обратим внимание на следующее свойство полных 2^K экспериментов: свойство ортогональности столбцов

$$x_j^T x_r = \sum_{i=1}^N x_{ij} x_{ir} = 0 ,$$

где i – номер строки; j – номер столбца.

Пользуясь свойством ортогональности, упростим формулу для получения вектора коэффициентов B .

Элементы информационной матрицы получаются перемножением столбцов X. Произведение разных столбцов из-за ортогональности равны нулю. Произведение столбца на себя равно N . Поэтому в итоге:

$$X^T X = \begin{pmatrix} N & 0 & \cdots & 0 \\ 0 & N & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & N \end{pmatrix} = NE,$$

где E – единичная матрица.

Матрица, обратная единичной матрице, есть также единичная матрица. Поэтому

$$(X^T X)^{-1} = \frac{1}{N} E.$$

Таким образом, вектор коэффициентов может быть определен по формуле

$$B = \frac{1}{N} X^T Y,$$

или в развернутом виде

$$\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} = \frac{1}{N} \begin{pmatrix} x_0^T \\ x_1^T \\ \vdots \\ x_n^T \end{pmatrix} Y.$$

Для отдельного коэффициента формулу можно представить в следующем виде:

$$b_j = \frac{1}{N} x_j^T Y,$$

где x_j^T – столбец матрицы плана.

Таким образом, для вычисления отдельного коэффициента матрицы B необходимо учитывать только один соответствующий столбец.

Пример. 7.2. Пусть нужно определить коэффициенты модели

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3; \quad Y^T = (y_1, y_2, \dots, y_8).$$

Решение. Значения коэффициентов, в соответствии с вышеизложенным, определяются следующим образом:

$$b_0 = \frac{1}{8} (y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8);$$

$$b_1 = \frac{1}{8} (-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8);$$

$$b_2 = \frac{1}{8} (-y_1 - y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8);$$

$$b_3 = \frac{1}{8} (-y_1 - y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8).$$

Полные планы типа 2^K можно использовать для получения моделей с нелинейностями как произведения факторов (учитываются взаимодействия между факторами), при этом свойство ортогональности ПлФЭ сохраняется и для определения коэффициентов можно пользоваться упрощенной формулой.

Так, например, если к плану, составленному ранее для имеющихся факторов, добавим столбцы функций x_1x_2 , x_1x_3 , x_2x_3 и $x_1x_2x_3$, то получим таблицу функций (табл. 7.4), в которой все столбцы ортогональны.

Заметим, что ортогональность при введении в модель взаимодействий (произведений) сохраняется и при большем числе уровней факторов.

Таблица 7.4

x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
+1	-1	-1	-1	+1	+1	+1	-1
+1	+1	-1	-1	-1	-1	+1	+1
+1	-1	+1	-1	-1	+1	-1	+1
+1	+1	+1	-1	+1	-1	-1	-1
+1	-1	-1	+1	+1	-1	-1	+1
+1	+1	-1	+1	-1	+1	-1	-1
+1	-1	+1	+1	-1	-1	+1	-1
+1	+1	+1	+1	+1	+1	+1	+1

При учете квадратов факторов двумя уровнями при построении не обойтись. Возведение значений любого фактора в квадрат даст +1, что приведет к совпадению столбцов с нулевым столбцом. При этом нельзя не только воспользоваться упрощенной формулой, но и решить задачу построения модели. В этом случае следует ввести три уровня: -1, 0, +1. План при таких уровнях содержит 3^K опытов (табл. 7.5).

Таблица 7.5

x_0	x_1	x_2	x_1x_2	x_1^2	x_2^2
+1	-1	-1	+1	+1	+1
+1	-1	0	0	+1	0
+1	-1	+1	-1	+1	+1
+1	0	-1	0	0	+1
+1	0	0	0	0	0
+1	0	+1	0	0	+1
+1	+1	-1	-1	+1	+1
+1	+1	0	0	+1	0
+1	+1	+1	+1	+1	+1

Квадраты факторов дают не ортогональный план, но ортогональность имеет место, если свободный член отсутствует ($b_0 = 0$). Это достигается в том случае, когда в матрице X не учитывается столбец x_0 .

7.2. Дробные факторные эксперименты

Полный факторный план имеет избыточное число опытов для определения коэффициентов. Так, например, для плана 2^K при $K=10$ число опытов будет равно 1024, для плана 3^K – 59049. В связи с этим возникает задача уменьшения числа опытов при сохранении ортогональности. Эту задачу решают путем построения плана для меньшего числа факторов, в котором значения других факторов определяются как произведения исходных. Планы при этом сокращаются в 2,4,8...или 3,9,27...раз. Их называют дробными факторными планами или дробными репликами ПлФЭ и обозначают:

- ДФЭ 2^{K-1} (полуреплика ПлФЭ 2^K);
- ДФЭ 2^{K-2} (четвертьреплика ПлФЭ 2^K);
- ДФЭ 2^{K-3} (1/8–реплика ПлФЭ 2^K).

Пример. 7.3. Требуется определить коэффициенты модели:

$$y \approx b_0 + \sum_{j=1}^6 b_j x_j.$$

Решение. Вместо полного плана $2^6 = 64$ построим полный план для трех факторов, а значения остальных определим как произведение первых трех. В результате получим план, показанный в табл.7.6.

Таблица 7.6

				x_1x_2	x_1x_3	$x_1x_2x_3$
x_0	x_1	x_2	x_3	x_4	x_5	x_6
+1	–1	–1	–1	+1	+1	–1
+1	–1	–1	+1	+1	–1	+1
+1	–1	+1	–1	–1	+1	+1
+1	–1	+1	+1	–1	–1	–1
+1	+1	–1	–1	–1	–1	+1
+1	+1	–1	+1	–1	+1	–1
+1	+1	+1	–1	+1	–1	–1
+1	+1	+1	+1	+1	+1	+1

Произведения факторов полного факторного эксперимента, используемые для определения значений других факторов, называются генерирующими соотношениями.

Искомые коэффициенты будут определяться следующим образом:

$$b_0 = \frac{1}{8}(y_1 + \dots + y_8);$$

$$b_1 = \frac{1}{8}(-y_1 - y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8)$$

и т.д.

Пример 7.4. Требуется определить коэффициенты модели:

$$y \approx b_1x_1 + b_2x_2 + b_3x_3 + b_{33}x_3^2.$$

Решение. Факторы будем варьировать на трех уровнях, при этом полный план $3^3 = 27$.

Для решения задачи используем девять опытов, построив план для двух факторов (табл. 7.7).

Таблица 7.7

		$x_1 \cdot x_2$	
x_1	x_2	x_3	x_3^2
-1	-1	+1	+1
-1	0	0	0
-1	+1	-1	+1
0	-1	0	0
0	0	0	0
0	+1	0	0
+1	-1	-1	+1
+1	0	0	0
+1	+1	+1	+1

Пример. 7.5. Требуется определить коэффициенты модели

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3; \quad Y^T = (y_1, y_2, \dots, y_8).$$

Решение. Вместо полного плана 2^3 (восемь опытов), построим полуреплику 2^2 (четыре опыта).

Третий фактор определим как произведение первых двух, в итоге таблица эксперимента примет вид табл. 7.8.

Таблица 7.8

x_1	x_2	$x_3 (x_1x_2)$
-1	-1	+1
+1	-1	-1
-1	+1	-1
+1	+1	+1

Искомые коэффициенты будут определяться следующим образом:

$$b_0 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4);$$

$$b_1 = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4);$$

$$b_2 = \frac{1}{4}(-y_1 - y_2 + y_3 + y_4);$$

$$b_3 = \frac{1}{4}(y_1 - y_2 - y_3 + y_4).$$

8. ИССЛЕДОВАНИЕ ВЛИЯНИЯ КАЧЕСТВЕННЫХ ФАКТОРОВ

8.1. Общие сведения о качественных факторах

Качественными являются факторы, уровни которых не могут быть выражены количественно.

Качественные факторы имеют такое же важное значение для анализа результатов и, следовательно, для процесса принятия решения, как монетарные или количественные факторы.

Большинство имеющихся вариантов приводят к результатам, которые нельзя выразить в количественных показателях. Они могут затрагивать человеческий фактор решений, как, например, моральное состояние, общественные отношения и качество или долгосрочные последствия, результат которых в количественном выражении нельзя оценить с приемлемой степенью точности.

Перед тем как принять взвешенное решение, необходимо учесть и качественные факторы.

Качественные факторы определяют внутренние свойства, признаки и особенности изучаемых объектов (например, качество технических средств, производительность труда, рентабельность и т.д.).

Разные качественные факторы относятся к различным решениям. Только ваш собственный здравый смысл, имеющаяся информация и ваш опыт в соответствующей области позволят вам установить качественные факторы, относящиеся к конкретному решению. Например, при страховании жизни одним из качественных факторов, относящихся к этому решению, является чувство уверенности, которое приходит с получением страхового полиса, гарантирующего в случае необходимости заботу о ваших иждивенцах.

Никакой объем научных исследований не даст вам окончательного списка всех возможных факторов. Важно рассматривать все качественные факторы, относящиеся к решению, какими бы незначительными они вам ни казались. Чем

больше соответствующей информации использует руководство, тем правильнее окажутся принятые им решения.

Приведем примеры качественных факторов (и их уровни):

- проект (уровни – варианты проекта);
- способ изготовления (уровни – различные технологии);
- учебник (уровни – варианты учебника);
- лекарства (уровни – различные лекарства для лечения одной болезни).

Факторы могут оказывать влияние на какой-то результирующий признак, т.е., например, проект, способ изготовления, оборудование могут влиять на надежность или другие показатели качества изделия, лекарство – на время выздоровления, учебник – на процесс усвоения изучаемого материала и т.д.

Предполагается, что результирующий признак может оцениваться количественно, хотя его значения могут носить как объективный характер (если результирующий признак измеряется или рассчитывается) так и субъективный (если его ценность оценивают эксперты).

Задача заключается в установлении влияния фактора на результат опытным путём. Она не является тривиальной в условиях действия других факторов, которые в совокупности на результирующий признак оказывают заметное влияние. Для её решения нужен статистический материал, полученный в результате опыта, при котором неоднократно регистрируют результаты при каждом уровне фактора. Этот материал удобно представить табл. 8.1.

Таблица 8.1

	Уровни фактора			
	1	2	К
Значения результирующего признака	y_{11}	y_{12}	...	y_{1K}
	y_{21}	y_{22}	...	y_{2K}
	y_{31}	y_{32}	...	y_{3K}
	\vdots	\vdots	...	\vdots
	y_{m1}	y_{n2}	...	y_{nK}

В табл. 8.1 n_1, n_2, \dots, n_K – объемы выборок (в общем случае разные) при каждом уровне фактора.

Прежде чем судить о влиянии качественного фактора на признак, полезно узнать, есть ли влияние вообще (может быть расхождения число случайные). Если влияния нет, то дальнейший анализ обычно не проводят. Поэтому сначала

выдвигают гипотезу об отсутствии влияния (нулевую гипотезу). Гипотеза равносильна утверждению, что все данные таблицы принадлежат одному и тому же распределению.

Для проверки гипотезы необходимо выполнить следующие действия:

- выбрать статистику, чувствительную к гипотезе;
- определить распределение статистики при нулевой гипотезе;
- задаться критической областью статистики, вероятность попадания в которую мала;
- проверить, попадает ли вычисленное значение статистики в критическую область.

При выборе распределения статистики необходимо знать распределение наблюдений. Если оно не известно, то наблюдение подвергают преобразованию, после чего распределение становится известным. Этот общий подход и рассмотрим.

8.2. Проверка гипотезы об отсутствии влияния фактора при неизвестном распределении результирующего фактора

Рассмотрим случай, когда среди значений признака нет совпадений. Это обычно справедливо при равномерном распределении. Тогда совпадения возможны только за счет неточности измерений. Чаще всего наблюдения заменяют рангами, которые располагаются равномерно.

Ранг – это номер наблюдения в упорядоченной последовательности наблюдений.

Для удобства расчетов составляется ранговая таблица (табл. 8.2).

Таблица 8.2

Факторы			
1	2	К
r_{11}	r_{12}	...	r_{1K}
r_{21}	r_{22}	...	r_{2K}
\vdots	\vdots	...	\vdots
r_{m1}	r_{n22}	...	r_{nKK}
\bar{r}_1	\bar{r}_2		\bar{r}_K

В табл. 8.2 \bar{r}_j – средние ранги столбцов, определяемые следующим образом:

$$\bar{r}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} r_{ij}.$$

Для проверки предположений при реализации выборки в качестве статистики используют сумму

$$\sum_{i=1}^n \left(\frac{\overset{x_i}{\text{наблюдаемое значение}} - \overset{x_{0i}}{\text{ожидаемое значение}}}{\sigma[\text{наблюдаемое значение}]} \right)^2.$$

Эта статистика при условии, что ожидаемые значения не вычисляются из наблюдений, а известны заранее, в среднем равняется N :

$$N = n_1 + n_2 + \dots + n_K.$$

Эта же сумма при вычислении ожидаемых значений в среднем равна числу степеней свободы, равному “ N минус число параметров”.

Если наблюдаемая величина имеет нормальное распределение, то сумма (статистика) имеет χ^2 -распределение (хи-квадрат), поскольку случайная величина χ^2 может быть представлена суммой квадратов нормированных случайных величин. Число степеней свободы χ^2 равно числу слагаемых.

Для проверки гипотезы H_0 часто используют указанную статистику. Она обозначается H и имеет следующий вид:

$$H = \sum_{j=1}^K \frac{\left(\overset{\text{средний}}{\text{ранг}} \right)^2}{\left(\sigma[\bar{r}_j] \right)^2} \cdot \left(\bar{r}_j - \frac{N+1}{2} \right)^2.$$

Статистика чувствительна к гипотезе H_0 : если гипотеза справедлива, то \bar{r}_j мало отличается от $\frac{N+1}{2}$, а если не справедлива, то отличается сильно, разность большая и статистика H_0 будет иметь большое значение.

Величину $\sigma[\bar{r}_j]$ можно выразить через $\sigma[r_{ij}]$:

$$\bar{r}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} r_{ij};$$

$$D[\bar{r}_j] = \frac{1}{n_j^2} \sum_{i=1}^{n_j} D[r_{ij}] = \frac{D[r_{ij}]}{n_j}.$$

В свою очередь, дисперсию отдельного наблюдения нетрудно найти, если учесть, что $p(r_{ij}) = \frac{1}{N}$:

$$D[r_{ij}] = \sum_{i=1}^N \frac{1}{N} \left(i - \frac{N+1}{2} \right)^2.$$

Выражение получится более простым, если дискретное равномерное распределение заменить непрерывным. Известно, что дисперсия равномерного распределения шириной Δ равна $\Delta^2/12$. Поэтому

$$D[r_{ij}] \approx \frac{N^2}{12}.$$

Более точным является выражение

$$D[r_{ij}] = \frac{N^2 - 1}{12}.$$

Таким образом,

$$H = \frac{12}{N^2 - 1} \sum_{j=1}^K n_j \left(\bar{r}_j - \frac{N+1}{2} \right)^2, \quad (8.1)$$

где n_j – число наблюдений для определения каждого фактора; $\frac{N+1}{2}$ – среднее значение ранга; N – общее число наблюдений; K – число уровней.

Величина (8.1) называется статистикой Краскела – Уоллеса.

Средние значения \bar{r}_j имеют распределение, близкое к нормальному вследствие того, что получены они как суммы. Поэтому H имеет распределение, близкое к χ^2 – распределению, с числом степеней свободы, равным $K-1$ (одна степень использована для связи $n_1 + \dots + n_K = N$). Приближение тем больше, чем больше объемы выборок n_1, \dots, n_K .

При малых объемах n_j для этой статистики имеются специальные таблицы распределения, по которым можно найти $H_{1-\alpha}$. В остальных случаях используют таблицы для распределения χ^2 .

Пример 8.1. Для выявления влияния денежного стимулирования на производительность труда шести группам из пяти человек каждая были предложены задачи одинаковой сложности (каждому в отдельности). Каждой группе давали свое вознаграждение за задачу. Результаты сведены в табл. 8.3.

Требуется проверить гипотезу об отсутствии влияния вознаграждения.

Решение. Используем статистику Краскела – Уоллеса.

Ранговая таблица будет иметь вид табл. 8.4.

Таблица 8.3

Номер сотрудника	Номера групп					
	1	2	3	4	5	6
1	10	8	12	12	24	19
2	11	10	17	15	16	18
3	9	16	14	16	22	27
4	13	13	9	16	18	25
5	7	12	16	19	20	24
Среднее	10	11,3	13,6	15,6	20	22,6

Таблица 8.4

Номер сотрудника	Номера групп					
	1	2	3	4	5	6
1	5,5	2	9	9	27,5	23,5
2	7	5,5	20	14	17	21,5
3	3,5	17	13	17	26	30
4	11,5	11,5	3,5	17	21,5	29
5	1	9	17	23,5	25	27,5
\bar{r}	5,7	9	10,5	16,1	23,4	26,3

Согласно формуле (8.1) при $N = 30$ и $K = 6$ получим $H = 21,8$.

Квантиль распределения с $K-1=5$ степенями свободы при уровне значимости 0,001 равен $\chi^2_{кр}(\alpha = 0,01) = 15,1$.

Таким образом, так как $H = 21,8 > \chi^2_{кр}(\alpha = 0,01) = 15,1$, гипотезу об отсутствии влияния вознаграждения при данном уровне значимости можно отвергнуть.

8.3. Проверка гипотезы об отсутствии влияния фактора при нормальном распределении результирующего признака

Ранее предполагалось, что случайная составляющая ε_{ij} результирующего признака имеет непрерывное распределение, величины ε_{ij} независимы и одинаково распределены. Довольно часто полагают, что это распределение нормальным распределением с нулевым средним и общей дисперсией, которая неизвестна.

Дополнительная информация о законе распределения позволяет использовать более сильные методы однофакторного анализа. Эти методы в совокупности составляют однофакторный дисперсионный анализ.

Рассмотрим статистику

$$\sum_{i=1}^{n_j} \frac{(y_{ij} - \bar{y}_j)^2}{\sigma^2}.$$

Ввиду того что наблюдаемые величины имеют нормальные распределения, эта сумма имеет распределение χ^2 с числом степеней свободы $n_j - 1$ (одна степень свободы пошла на вычисление среднего \bar{y}_j):

$$\sum_{i=1}^{n_j} \frac{(y_{ij} - \bar{y}_j)^2}{\sigma^2} = \chi_{n_j-1}^2.$$

При суммировании случайных величин χ^2 получается χ^2 -распределение, при этом числа степеней свободы складываются:

$$\sum_{j=1}^K \sum_{i=1}^{n_j} \frac{(y_{ij} - \bar{y}_j)^2}{\sigma^2} = \chi_{N-K}^2.$$

Рассмотрим статистику для среднего

$$\sum_{j=1}^K \frac{(\bar{y}_j - \bar{y})^2}{\sigma^2 / n_j} = \chi_{K-1}^2,$$

где $\frac{\sigma^2}{n_j}$ – дисперсия средней величины \bar{y}_j наблюдений j -го столбца.

Известно, что отношение $\frac{\chi_{n_1}^2 / n_1}{\chi_{n_2}^2 / n_2} = F_{n_1, n_2}$ имеет распределение Фишера.

Таким образом, получим статистику с известным законом распределения

$$F_{K-1, N-K} = \frac{\frac{1}{K-1} \sum_{j=1}^K (y_j - \bar{y})^2}{\frac{1}{N-K} \sum_{j=1}^K \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}.$$

Из последнего выражения видно, что статистика не зависит от σ . Отметим, что статистика чувствительна к гипотезе, т.е. она возрастает тем быстрее, чем больше средние групповые значения отличаются от общего среднего (если гипотеза не выполняется).

Пример 8.2. Заказчик решил выбрать изготовителя. Три изготовителя сделали по три экземпляра прибора и проверили их надёжность. Получили результаты (наработка до отказа, табл. 8.5).

Таблица 8.5

Наработка	Номера предприятий		
	1	2	3
	200	112	230
	160	135	145
	140	192	170
\bar{y}_j	167	146	178
\bar{y}	164		

Решение. Рассчитаем статистику ($K = 3, N = 9$):

$$F_{K-1, N-K} = \frac{\frac{1}{K-1} \sum_{j=1}^K (\bar{y}_j - \bar{y})^2}{\frac{1}{N-K} \sum_{j=1}^K \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2} =$$

$$= \frac{\frac{1}{J-1} [(167-164)^2 + (146-164)^2 + (178-164)^2]}{\frac{1}{9-3} \left[(200-167)^2 + (160-167)^2 + (140-167)^2 + (112-146)^2 + \right.}$$

$$\left. + (135-146)^2 + (192-146)^2 + (220-178)^2 + (145-178)^2 + (170-178)^2 \right]} = 0,58.$$

Пусть $\alpha=0,05$, при этом $F_{кр}=5,14$ (берется из таблицы).

Сравнение F и $F_{кр}$ показывает, что гипотеза справедлива и можно выбрать любого изготовителя, так как предприятие не влияет на качество, т.е. попадает в допустимую область.

Проверим гипотезу по статистике Краскела – Уоллеса.

Ранговая таблица имеет вид табл. 8.6.

Таблица 8.6

Номера предприятий		
1	2	3
8	1	9
5	2	4
3	7	6
5,3	3,3	6,3
$(N+1)/2 = 5$		

Величина H равна

$$H = \frac{12}{N^2 - 1} \sum_{j=1}^K n_j \left(r_j - \frac{N+1}{2} \right)^2 = \frac{12}{81-1} \cdot 3 \left[(5,3-5)^2 + (3,3-5)^2 + (6,3-5)^2 \right] = 2,1.$$

Для проверки гипотезы используем критерий Пирсона.

Примем, что $\alpha = 0,05$, тогда $H_{кр} = 6,0$.

Сравнение результатов дает основание принять решение о подтверждении гипотезы.

9. ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ СИСТЕМ

9.1. Классификация задач и методов идентификации

Построение математических моделей того или иного типа на основе результатов наблюдений за поведением объектов и исследование их свойств составляет основное содержание науки идентификации [1, 7, 8, 9, 15, 18, 20, 21, 22, 23].

Задача идентификации сводится, в общем случае, к определению оператора модели, преобразующего входные воздействия объекта в выходные величины [2, 15, 20, 21]. Оператор объекта является его математической формализацией, т.е. математической моделью объекта, и может быть определен в соответствующих пространствах функций. Операторы могут характеризоваться разными структурой и характеристиками, и соответственно задача идентификации объекта может иметь различные постановки:

1. Задача нахождения характеристик (параметров) объекта.

По известным наблюдаемым переменным требуется определить операторы (или параметры операторов).

2. Задача оценивания переменных состояния.

Состояние объекта характеризуется многомерной переменной состояния, вектором, однозначно определяющим все его характеристики. По известным наблюдаемым случайным сигналам при известных операторах с известными параметрами требуется определить (оценить) ненаблюдаемый случайный сигнал состояния. При этом возможны следующие постановки задачи:

а) оценивание состояния в текущий момент времени – задача фильтрации, или, собственно, оценивание;

б) оценивание состояния в будущий момент времени, сдвинутый на Δt относительно текущего момента – задача прогнозирования или экстраполяции;

в) оценивание состояния в прошлый момент времени – задача сглаживания или интерполяции.

3. Задача генерации случайных сигналов с заданными характеристиками или определения характеристик случайных сигналов.

В некоторых случаях возникает задача, при которой одновременно проводятся параметрическая идентификация и оценивание состояния (одновременная идентификация и оценивание), а также возможен ряд других частных постановок задач идентификации и оценивания [21, 22].

В большинстве работ, посвященных идентификации, выделяются следующие основные составляющие, которые нужно выполнить на этапе идентификации [12, 23]:

- сформулировать требования к данным наблюдений: как выполнить сбор экспериментальных данных, как использовать эти данные, собранные в реальных условиях проведения эксперимента [3, 12, 19];
- определить класс объектов – совокупность моделей-кандидатов, из которой впоследствии будет отобрана наилучшая модель;
- сформировать так называемую функцию потерь или риска, характеризующую адекватность объекта и настраиваемой модели, и на ее основе сформулировать критерий качества идентификации;
- выбрать способ оценки степени соответствия исследуемой модели экспериментальным данным;
- определить процедуру верификации модели: провести проверку и подтверждение адекватности модели, т.е. выяснить, в какой степени модель действительно «объясняет» поведение изучаемой системы.

При построении математических моделей существенную роль играют следующие факторы [4, 17, 18, 23]:

1. До начала проведения эксперимента необходимо определить условия, в которых будет проводиться сбор данных, решить вопросы дальнейшего конкретного использования этих данных. Эти задачи решаются на этапе планирования эксперимента путем выбора числа опытов эксперимента и условий его проведения, необходимых и достаточных для решения поставленной задачи с требуемой точностью. Этот этап непосредственно не относится к идентификации, а предвращает ее.

2. В конструктивном смысле идентификация – это определение по входным и выходным воздействиям такой модели из определенного класса моделей, которой реальная исследуемая система эквивалентна. В соответствии с этим нужно определить класс моделей, среди которых будет выбрана наиболее подходящая. На этом этапе необходимо выбрать общую структуру модели и класс уравнений, которыми предполагается описывать наблюдаемый процесс. Этот этап иногда

называется идентификацией в широком смысле или структурной идентификацией и зачастую оказывается решающим фактором. Для успешного решения задачи структурной идентификации требуется использовать априорные сведения о физических, химических или иных явлениях, происходящих в процессе, знание формальных аналитических свойств моделей, инженерные навыки и интуицию.

До настоящего времени общих формальных подходов к решению задачи структурной идентификации не существует, и этап структурной идентификации часто сводится к эвристическому заданию структуры модели на основе априорных сведений об объекте.

3. Близость полученной модели реальной исследуемой системе относительна, так как операторы объекта и модели могут быть описаны на разных языках, иметь разную структуру или количество входов, и потому понятие адекватности может быть сформулировано разными способами. Так как непосредственно оценить близость операторов объекта и модели сложно или зачастую невозможно, наиболее часто оценивается близость выходных величин объекта и модели или математического ожидания ошибок оценок параметров. Для этого вводится понятие функции потерь или риска, в дальнейшем подлежащей минимизации. Далее для выбора «наилучшей» модели из определенного класса на основании этой функции потерь формируется некоторый критерий, и в дальнейшем задача идентификации становится задачей оптимизации выбранного критерия.

4. После определения структуры модели и класса уравнений необходимо определить численные значения параметров – коэффициенты дифференциальных, разностных, интегральных уравнений или других математических конструкций линейной или нелинейной модели объекта и (или) состояний, вошедших в уравнения математической модели. Таким образом, решению подлежит задача оценивания параметров и (или) состояний по имеющимся экспериментальным данным, т.е. по значениям измеряемых переменных. Данная задача называется задачей параметрической идентификации или идентификацией в узком смысле. При оценивании параметров приходится решать задачу минимизации некоторых функциональных зависимостей от измеряемых величин (обычно от разности выходных сигналов модели и объекта) и от неизмеряемых величин – параметров и состояний. Для решения этой задачи необходимо разработать алгоритм идентификации, который на основе доступных для наблюдения входных и выходных величин определял бы параметры настраиваемой модели, минимизирующие погрешность модельного описания в соответствии с выбранным функционалом качества.

5. Переход от этапа построения модели к последующему ее использованию требует оценку качества полученной модели, т.е. проверку адекватности модели объекту. Вследствие того что абсолютная эквивалентность модели объекту принципиально недостижима, основным условием подтверждения адекватности модели является возможность использования полученной модели для решения той задачи, ради которой эта модель строилась. Поэтому адекватность предполагает воспроизведение моделью с необходимой полнотой всех свойств объекта, существенных для целей данного исследования. Степень адекватности модели и объекта обычно оценивают путем сравнения их выходных сигналов при подаче одинаковых входных воздействий на объект и его модель. Это сравнение предпочтительно производить на основе новой информации, отличной от данных, которые использовались в процессе идентификации объекта.

Ключевым этапом идентификации является идентификационный эксперимент, который состоит в следующем. На входы объекта и модели подается внешнее воздействие $u(t)$. В реальных условиях взаимодействия объекта со средой сигналы наблюдения за объектом искажены случайными возмущениями, определяемыми спецификой функционирования самого объекта, погрешностями методов и средств измерений и неконтролируемыми воздействиями внешней среды. При использовании такой схемы наблюдений полагается, что результаты измерений входного сигнала являются действительным входным сигналом, а все внутренние и внешние возмущения, отклонения измеренных значений от истинных воздействий характеризуются обобщенной помехой $\eta(t)$. Обычно в результате эксперимента получают наблюдения входа и выхода, т.е. реализации случайных функций $u(t)$ и $y(t)$. Поскольку объект связывает вход $u(t)$ с выходом $y(t)$, то эту связь выходной величины с входной формально можно представить некоторым оператором f :

$$y(t) = f(u(t), \eta(t), b). \quad (9.1)$$

В соответствии с зависимостью (9.1) выходная величина объекта зависит от внешнего воздействия $u(t)$, помехи $\eta(t)$ и от неизвестного вектора параметров b , значения которых непосредственному наблюдению недоступны.

На основании сведений об объекте формируется модель, под которой понимается некоторый оператор f , преобразующий наблюдаемое входное воздействие $u(t)$ в ее реакцию $y_M(t)$.

Модель описывается уравнениями, подобными уравнениям объекта (9.1) и содержащими информацию об измеряемых входных и выходных величинах, причем полагается, что помехи не меняют вида модели. Коэффициенты этих уравнений являются параметрами модели.

Выходная величина модели зависит от параметров b , которые рассчитываются на основе алгоритма, обрабатывающего вектор всех наблюдений. Для нахождения вектора параметров b необходимо определить оптимальный, в смысле подобия объекту, способ корректировки модели. При таком подходе задача идентификации заключается в построении модельного оператора f из некоторого класса операторов (задача структурной идентификации) и определении по наблюдениям $u(t)$ и $y(t)$ вектора параметров b (параметрическая идентификация), такого, чтобы выходной сигнал модели был бы наиболее близок к выходному сигналу объекта.

На основе сравнения искаженного помехой $\eta(t)$ выходного сигнала объекта с выходным сигналом модели находится невязка – разность выходных величин объекта и модели. Для оценки соответствия модели объекту вводится функция потерь (функция невязки), в любой момент времени зависящая от выходов объекта и модели и не зависящая от оператора, и на ее основе формулируется критерий качества идентификации, характеризующий адекватность модели реальному объекту.

Минимизация функционала идентификации, соответствующая улучшению качества идентификации, осуществляется путем надлежащего выбора структуры модели и изменением значений ее параметров. Процедура изменения реализуется алгоритмом идентификации.

Существуют разные способы оценивания параметров, различающиеся между собой по используемому критерию оптимальности и имеющейся априорной информации. В определенной степени выбор критерия оптимальности субъективен, а процедура оценивания существенно зависит от принятого критерия.

В подавляющем большинстве случаев критерий качества идентификации выбирается квадратичным, в виде интегрального значения квадрата невязки или среднего значения квадрата невязки.

Величины $y(t)$, $y_M(t)$ и функция потерь (функция невязки) рассматриваются как временные функции, определенные на интервале наблюдений.

В некоторых задачах идентификации применяются модульные функции невязки, еще реже используются функционалы качества, отличные от квадратичных и модульных [5, 22]. Кроме рассмотренных критериев, усредняющих потери на некотором интервале наблюдений, возможны формулировки критериев, усредняющих потери по множеству реализаций [1, 14]. Кроме того, при решении задач идентификации вектора параметров по имеющимся выборкам измерений

сигналов могут использоваться статистические критерии [16, 17]: максимального правдоподобия, максимума апостериорной плотности распределения вероятности.

Методы оценивания параметров моделей объектов, в общем случае, можно разделить на два класса подходов в зависимости от способа реализации процедуры оценивания. К первому типу относятся подходы на основе использования явных математических выражений, ко второму – реализации процедур оценивания с использованием настраиваемой модели.

При реализации методов оценивания первого типа математическая модель задается в виде явных математических соотношений, содержащих набор подлежащих определению числовых параметров. На объекте проводятся специальные идентификационные эксперименты по сбору массивов входных $u(t)$ и выходных $y(t)$ данных. Далее проводится обработка результатов полученных экспериментальных данных с целью минимизации выбранного функционала идентификации. Оптимальные процедуры оценивания параметров b в этом случае сводятся к разрешению соотношений идентификации относительно b . Оценивание параметров в данном случае осуществляется при помощи ретроспективных алгоритмов идентификации, когда решение получается в результате обработки всего массива данных путем выполнения конечного числа элементарных операций и не может быть получено как результат промежуточных вычислений. Такая процедура оценивания, с инженерной точки зрения, относится к методам идентификации вне контура регулирования и не позволяет обрабатывать поступающие наблюдения последовательно, в режиме нормальной эксплуатации.

При реализации методов оценивания второго типа используется принцип подстройки модели к объекту по признакам близости поведения. В этом случае моделируется структура математических соотношений, параметры b которой изменяются таким образом, чтобы характеристики модели были близки к характеристикам исследуемой системы. При таком подходе устройство настройки изменяет параметры настраиваемой модели b на основе алгоритмов идентификации, минимизируя тем самым функцию ошибок. Решение получается в принципе, как результат бесконечного числа таких операций, при этом каждый промежуточный результат представляет приближенное решение. Этот тип реализации относится к методам идентификации в замкнутом контуре и позволяет проводить оперативную идентификацию в режиме нормального функционирования объекта.

Следует отметить [18], что с появлением цифровых вычислительных устройств стало удобнее реализовывать используемые функции (вычисление

критерия, автоматическую настройку и др.) алгоритмически, что приводит к стиранию четких границ между различными способами идентификации. Основным признаком, указывающим на применение методов идентификации с настраиваемой моделью, следует считать наличие обратной связи.

Среди возможных алгоритмов идентификации широкое распространение получили рекуррентный метод наименьших квадратов, а также метод стохастической аппроксимации (МСА). Методу наименьших квадратов соответствует минимизация квадратичного критерия. Этот метод приводит к решению системы линейных уравнений, и поэтому оптимальное решение b может быть выражено в явной аналитической форме [10, 20]. МСА характеризуется простотой и универсальностью и позволяет не ограничиваться квадратичными критериями идентификации, а формировать разнообразные как линейные, так и нелинейные алгоритмы идентификации [2,8,19].

Классификация методов идентификации определяется в первую очередь формой представления математических моделей – обыкновенных дифференциальных, разностных уравнений, уравнений свертки и т.д. При этом ни один из методов идентификации не является универсальным для идентификации всех видов математических моделей, а используется в отдельных областях применения.

Методы идентификации можно классифицировать также по типу используемых входных сигналов при проведении идентификационного эксперимента – способу тестирования. По этому признаку различают активные и пассивные методы идентификации. В активных методах на вход объекта подаются специально сформированные воздействия – тестовые сигналы детерминированного или случайного характера [3]. Достоинствами этого подхода являются минимальные требования к априорным сведениям об объекте, целенаправленный характер идентификации, и, как следствие, уменьшение временных и материальных затрат на проведение эксперимента.

При использовании пассивных методов объект находится в условиях нормального функционирования и параметры модели отыскиваются по результатам статистической обработки наблюдений. Преимуществом этого подхода является отсутствие необходимости проводить специальные исследования объекта, достаточно лишь измерение наблюдаемых сигналов в режиме рабочего функционирования объекта с последующим расчетом параметров модели. Недостатками такого подхода являются значительные временные затраты на сбор и необходимую статистическую обработку данных и жесткие требования к частотному спектру входного воздействия – он не должен быть меньше полосы частот динамической характеристики идентифицируемого объекта.

По характеру используемых сигналов различают детерминированные и статистические методы. При проведении активной идентификации на основе детерминированных сигналов возможно применение детерминированных методов идентификации. В реальных условиях сигналы всегда подвержены действию помех и сильно зашумлены, и детерминированные алгоритмы необходимо дополнять статистическим усреднением (сглаживанием) получаемых результатов [1, 2, 3].

По признаку временных затрат методы делятся на оперативные и ретроспективные. При оперативной идентификации обеспечивается текущее отслеживание меняющихся характеристик объекта. На основе рекуррентных алгоритмов, реализуемых в темпе, близком к скорости протекания процесса, оценки параметров моделей уточняются в реальном времени на каждом шаге поступления новых измерений. При ретроспективной идентификации вначале собирается весь массив данных и оценки характеристик или параметров получаются после обработки этого массива.

9.2. Определение параметров статической системы по полной выборке

Задача определения параметров динамической системы (ДС) в общем случае сводится к задаче определения параметров статической системы (СС). СС имеет одно состояние, т.е. выход целиком определяется входами в тот же момент времени (рис.9.1).

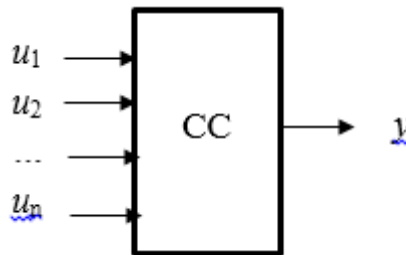


Рис. 9.1. Схема эксперимента определения параметров статической системы

Требуется определить коэффициент пропорциональности функции

$$y = v_1 u_1 + v_2 u_2 + \dots + v_n u_n = U^T B = B^T U, \quad (9.2)$$

где $U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ – вектор управления; $B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$ – вектор параметров.

Если нужно учесть постоянную составляющую, то можно положить, например, что $u_1=1$.

Функция (9.1) может описывать работу системы приблизительно, это может быть линеаризованная зависимость. Здесь требуется найти наилучшее приближение нелинейной зависимости к линейной по результатам наблюдения за поведением. Ясно, что число точек измерений должно быть больше числа параметров.

Нетрудно видеть, что задача аналогична задаче определения вектора коэффициентов b модели

$$y = b_0 + b_1 f_1 + \dots + b_n f_n$$

в теории планирования эксперимента.

В последней задаче вместо функций f_i нужно подставлять удовлетворяющие воздействия.

Выполним измерения в N точках.

Результаты наблюдения можно представить таблицей вида табл.9.1.

Таблица 9.1

k	$u_1(k)$	$u_2(k)$		$u_n(k)$	$y(k)$
1	$u_1(1)$	$u_2(1)$		$u_n(1)$	$y(1)$
2	$u_1(2)$	$u_2(2)$		$u_n(2)$	$y(2)$
3	$u_1(3)$	$u_2(3)$		$u_n(3)$	$y(3)$
.
.
.
N	$u_1(N)$	$u_2(N)$		$u_n(N)$	$y(N)$

В табл. 9.1 k – номер такта (момента времени).

Параметры b_j определяются из критерия

$$\sum_{k=1}^N (y(k) - \sum_{j=1}^n b_j u_j(k))^2 = \min. \quad (9.3)$$

Обозначим матрицей управления U – матрицу

$$U = \begin{pmatrix} u_1(1) & u_2(1) & \dots & u_n(1) \\ u_1(2) & u_2(2) & \dots & u_n(2) \\ \dots & & \dots & \\ u_1(N) & u_2(N) & \dots & u_n(N) \end{pmatrix} = \begin{pmatrix} u^T(1) \\ u^T(2) \\ \vdots \\ u^T(N) \end{pmatrix},$$

а вектор значений выхода Y матрицей:

$$Y = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix}.$$

Уравнение (9.3) преобразуется к уравнению

$$(Y - U^T b)^T (Y - U^T b) = \min.$$

Его решением является

$$B = (U^T U)^{-1} U^T Y.$$

9.3. Определение параметров динамической системы по полной выборке

Перейдём к динамическим системам (рис. 9.2).

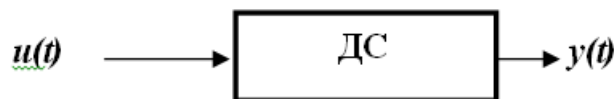


Рис. 9.2. Схема идентификационного эксперимента определения параметров динамической системы

Пусть ДС описывается дифференциальным уравнением

$$\ddot{y} + a_1 \dot{y} + a_0 y = c u.$$

Пусть оно соответствует колебательному звену с передаточной функцией

$$W(p) = \frac{K}{T^2 p^2 + 2\xi T p + 1},$$

причём

$$a_1 = \frac{2\xi}{T}; \quad a_0 = \frac{1}{T^2}; \quad c = \frac{K}{T^2}.$$

От дифференциального уравнения перейдём к разностному уравнению, учитывая, что

$$\dot{y}(t) \cong \frac{y(t + \Delta t) - y(t)}{\Delta t}; \quad \ddot{y}(t) \cong \frac{\dot{y}(t + \Delta t) - \dot{y}(t)}{\Delta t}.$$

После подстановки в дифференциальное уравнение выражений параметров и несложных преобразований (замены t на дискретное время i , причем моменты i следуют с интервалом Δt) получим

$$\begin{aligned} y(i) &= c\Delta t^2 u(i-2) + (a_1\Delta t - 1 - a_0\Delta t^2) y(i-2) + (2 - a_1\Delta t) y(i-1) = \\ &= b_1 u(i-2) + b_2 y(i-2) + b_3 y(i-1). \end{aligned}$$

Введя переменные

$$u_1(i)=u(i-2),$$

$$u_2(i)=y(i-2),$$

$$u_3(i)=y(i-1),$$

получим

$$y(i)=b_1u_1(i)+ b_2u_2(i) + b_3u_3(i),$$

или

$$y=b_1u_1+b_2u_2+b_3u_3 .$$

Этой зависимости соответствует схема, изображенная на рис. 9.3.

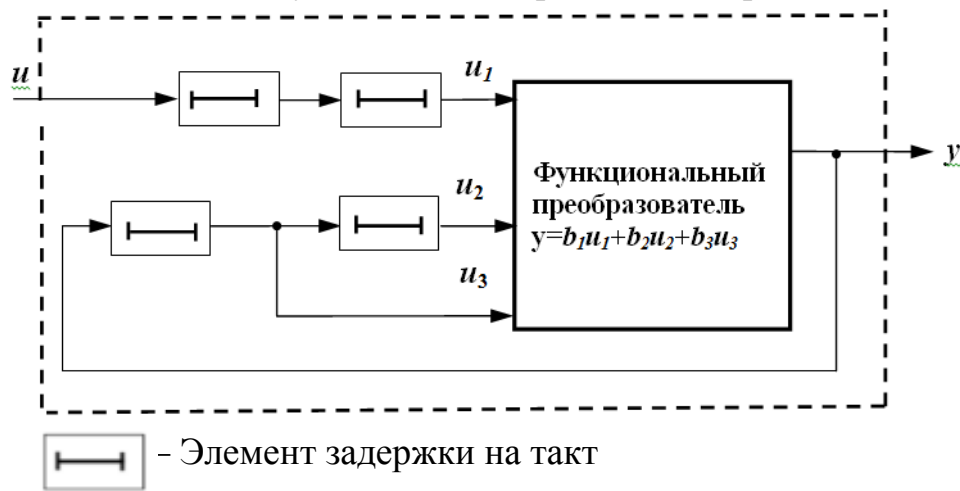


Рис. 9.3.Развернутая схема идентификационного эксперимента определения параметров динамической системы

Для того чтобы оценить точность определения параметров, нужно пересчитать их в параметры колебательного звена.

Имеем систему уравнений

$$\begin{cases} b_1 = c\Delta t^2 = \frac{K}{T^2} \Delta t^2; \\ b_2 = a_1\Delta t - 1 - a_0\Delta t^2 = \frac{2\xi}{T} \Delta t - 1 - \frac{1}{T^2} \Delta t^2; \\ b_3 = 2 - a_1\Delta t = 2 - \frac{2\xi}{T} \Delta t, \end{cases}$$

и её решение:

$$T = \frac{\Delta t}{\sqrt{1-b_2-b_3}},$$

$$K = \frac{b_1}{1-b_2-b_3},$$

$$\xi = \frac{2 - b_3}{2\sqrt{1 - b_2 - b_3}}.$$

Наблюдаемыми являются только сигналы на входе u и выходе y . Сигналы u_1 , u_2 , u_3 полностью определяются u и y , но сдвинуты от них по времени. Пусть u и y наблюдаются в моменты $i: = 1, 2, 3, \dots, N$.

Наблюдение для идентификации начнём с момента времени $i=3$, после того, как поданный сигнал $u(1)$ дойдёт до функционального преобразователя. Матрица управлений будет иметь вид

$$U = \begin{pmatrix} u_1(3) & u_2(3) & u_3(3) \\ u_1(4) & u_2(4) & u_3(4) \\ \dots & \dots & \dots \\ u_1(N) & u_2(N) & u_3(N) \end{pmatrix} = \begin{pmatrix} u(1) & y(1) & y(2) \\ u(2) & y(2) & y(3) \\ \dots & \dots & \dots \\ u(N-2) & y(N-2) & y(N-1) \end{pmatrix},$$

вектор $Y = \begin{pmatrix} y(3) \\ y(4) \\ \vdots \\ y(N) \end{pmatrix}$, тогда

$$B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = (U^T U)^{-1} U^T Y. \quad (9.4)$$

Пример 9.1. Требуется определить коэффициент b функции $y = bu$, т.е. системы с одним входом, если известны матрицы:

$$U = \begin{pmatrix} u(1) \\ u(2) \\ \vdots \\ u(N) \end{pmatrix}; \quad Y = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix}.$$

Решение. Вычислим отдельные составляющие формулы (9.4) для определения параметра:

$$U^T U = \sum_{i=1}^N u^2(i);$$

$$(U^T U)^{-1} = \frac{1}{\sum_{i=1}^N u^2(i)};$$

$$U^T Y = \sum_{i=1}^N u(i)y(i).$$

Тогда в соответствии с общей формулой (9.4) искомый параметр

$$b = \frac{\sum_{i=1}^N u(i)y(i)}{\sum_{i=1}^N u^2(i)}. \quad (9.5)$$

Пример 9.2. Требуется определить параметры апериодического звена (динамическая система), т.е. параметры K и T передаточной функции $W(p) = \frac{K}{Tp + 1}$,

по результатам наблюдений за входом и выходом звена.

Наблюдение ведётся за входом u и выходом y в моменты времени

$$\begin{aligned} t: & \quad t, \quad t+\Delta t, \quad t+2\Delta t, \dots, t+(N-1)\Delta t; \\ i: & \quad 1, \quad 2, \quad 3, \dots, \quad N. \end{aligned}$$

Представляющие интерес параметры (K и T) определяются через коэффициенты разностного уравнения, которые определяются по результатам наблюдения за поведением системы. Найдём связь параметров модели (K и T) с коэффициентами разностного уравнения.

Из передаточной функции

$$W(p) = \frac{K}{Tp + 1}$$

следует, что

$$(Tp + 1)y = Ku,$$

или

$$T \dot{y} + y = Ku.$$

Разделим левую и правую части на параметр T :

$$\dot{y} + \frac{1}{T}y = \frac{K}{T}u.$$

Введем новые обозначения:

$$\dot{y} = cy - ay.$$

После представления производной через конечные приращения

$$\frac{y(t + \Delta t) - y(t)}{\Delta t} = cu(t) - ay(t)$$

получим

$$y(t + \Delta t) = c\Delta t u(t) - a\Delta t y(t) + y(t).$$

Или со сдвигом на Δt имеем $y(t) = c\Delta t u(t - \Delta t) + (1 - a\Delta t)y(t - \Delta t)$.

Перейдем к дискретному времени и получим

$$y(i) = b_1 u(i-1) + b_2 y(i-1).$$

Введя переменные

$$u_1(i) = u(i-1),$$

$$u_2(i) = y(i-1),$$

получим каноническое выражение (9.2)

$$y(i) = b_1 u_1(i) + b_2 u_2(i).$$

Из системы двух уравнений:

$$\begin{cases} b_1 = c\Delta t = \frac{K}{T} \Delta t; \\ b_2 = 1 - a\Delta t = 1 - \frac{1}{T} \Delta t; \end{cases}$$

следует, что

$$T = \frac{\Delta t}{1 - b_2};$$

$$K = \frac{b_1 T}{\Delta t} = \frac{b_1}{1 - b_2}.$$

Структурная схема модели процесса идентификации представлена на рис. 9.4.

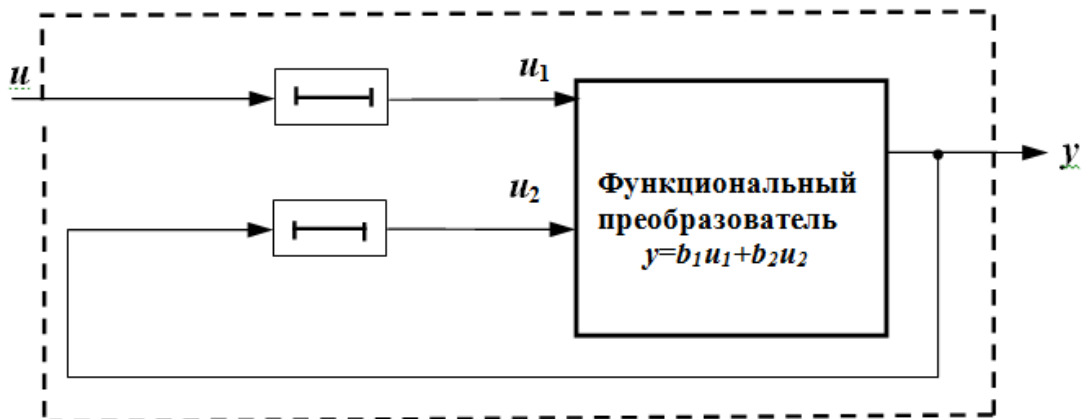


Рис. 9.4. Структурная схема модели процесса

Определение параметров b_1 и b_2 на основе множества наблюдений в соответствии с выражением (9.4) представим следующим образом:

$$U = \begin{pmatrix} u_1(2) & u_2(2) \\ u_1(3) & u_2(3) \\ \vdots & \vdots \\ u_1(N) & u_2(N) \end{pmatrix} = \begin{pmatrix} u(1) & y(1) \\ u(2) & y(2) \\ \vdots & \vdots \\ u(N-1) & y(N-1) \end{pmatrix};$$

$$Y = \begin{pmatrix} y(2) \\ y(3) \\ \vdots \\ y(N) \end{pmatrix}.$$

Тогда произведение

$$U^T U = \begin{pmatrix} u(1) & u(2) & \cdots & u(N-1) \\ y(1) & y(2) & \cdots & y(N-1) \end{pmatrix} \begin{pmatrix} u(1) & y(1) \\ u(2) & y(2) \\ \vdots & \vdots \\ u(N-1) & y(N-1) \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_{i=1}^{N-1} u^2(i) & \sum_{i=1}^{N-1} u(i)y(i) \\ \sum_{i=1}^{N-1} u(i)y(i) & \sum_{i=1}^{N-1} y^2(i) \end{pmatrix};$$

а обратная матрица

$$(U^T U)^{-1} = \frac{1}{\Delta} \begin{vmatrix} \sum_{i=1}^{N-1} y^2(i) & -\sum_{i=1}^{N-1} u(i)y(i) \\ -\sum_{i=1}^{N-1} u(i)y(i) & \sum_{i=1}^{N-1} u^2(i) \end{vmatrix},$$

где

$$\Delta = \sum_{i=1}^{N-1} u^2(i) \cdot \sum_{i=1}^{N-1} y^2(i) - \left(\sum_{i=1}^{N-1} u(i)y(i) \right)^2.$$

В свою очередь,

$$U^T Y = \begin{vmatrix} \sum_{i=1}^{N-1} u(i)y(i+1) \\ \sum_{i=1}^{N-1} y(i)y(i+1) \end{vmatrix}.$$

В итоге

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = (U^T U)^{-1} U^T Y = \frac{1}{\Delta} \begin{pmatrix} \sum_{i=1}^{N-1} y^2(i) \sum_{i=1}^{N-1} u(i) y(i+1) - \sum_{i=1}^{N-1} u(i) y(i) \cdot \sum_{i=1}^{N-1} y(i) y(i+1) \\ - \sum_{i=1}^{N-1} u(i) y(i) \sum_{i=1}^{N-1} u(i) y(i+1) + \sum_{i=1}^{N-1} u^2(i) \cdot \sum_{i=1}^{N-1} y(i) y(i+1) \end{pmatrix}.$$

Таким образом, получены коэффициенты b_1 и b_2 .

Но, так как

$$b_1 = c \Delta t = \frac{K}{T} \Delta t ;$$

$$b_2 = 1 - a \cdot \Delta t = 1 - \frac{1}{T} \Delta t ,$$

то из решения этих уравнений относительно искомых параметров получаем

$$T = \frac{\Delta t}{1 - b_2} ,$$

$$K = \frac{b_1 T}{\Delta t} = \frac{b_1}{1 - b_2} .$$

Так получаются усреднённые оценки параметров T и K апериодического звена.

9.4. Последовательная идентификация

В системах управления очень часто приходится постоянно отслеживать изменения модели управляемого объекта. При последовательной идентификации параметры модели уточняются после каждого измерения по рекуррентным, более простым, формулам. Это позволяет не только упростить расчёты, но, главное, и повысить оперативность.

Формулы для последовательной идентификации получим путём обобщения формул, которые выведены для простого примера однопараметрической статической системы

$$y = bu,$$

для которой получено значение b методом наименьших квадратов.

Для последовательной идентификации важно подчеркнуть, что полученное значение b – это оценка, т.е. приближённое значение, и она будет уточняться постепенно. Перепишем формулу (9.5) с учётом этого:

$$\hat{b}(N) = \frac{\sum_{i=1}^N u(i) y(i)}{\sum_{i=1}^N u^2(i)} .$$

Из последней формулы следует, что

$$\hat{b}(1) = \frac{y(1)}{u(1)},$$

$$\begin{aligned}\hat{b}(2) &= \frac{u(1)y(1) + u(2)y(2)}{u^2(1) + u^2(2)} = \hat{b}(1) + \frac{u(1)y(1) + u(2)y(2)}{u^2(1) + u^2(2)} - \hat{b}(1) = \\ &= \hat{b}(1) + \frac{u(1)y(1) + u(2)y(2) - u^2(1)\hat{b}(1) - u^2(2)\hat{b}(1)}{u^2(1) + u^2(2)} = \\ &= \hat{b}(1) + \frac{u(1)y(1) + u(2)y(2) - u(1)y(1) - u^2(2)\hat{b}(1)}{u^2(1) + u^2(2)} = \\ &= \hat{b}(1) + \frac{u(2)y(2) - u^2(2)\hat{b}(1)}{u^2(1) + u^2(2)} = \\ &= \hat{b}(1) + p(2)u(2)\left(y(2) - u(2)\hat{b}(1)\right).\end{aligned}$$

$$p(2) = \frac{1}{u^2(1) + u^2(2)},$$

причём

$$\frac{1}{P(2)} = u^2(1) + u^2(2) = \frac{1}{P(1)} + u^2(2).$$

Эту формулу можно обобщить для произвольного момента времени i :

$$\hat{b}(i) = \hat{b}(i-1) + p(i)u(i)\left(y(i) - u(i)\hat{b}(i-1)\right) \quad (9.6)$$

где

$$\frac{1}{p(i)} = \sum_{j=1}^i u^2(j) = \frac{1}{p(i-1)} + u^2(i).$$

Таким образом, оценка параметра на i -м шаге определяется как сумма оценки этого параметра на предыдущем шаге и поправки, равной взвешенной невязке. По смыслу влияние поправки на оценку должно постепенно уменьшаться за счёт уменьшения:

$$p(i) = \frac{1}{\sum_{j=1}^i u^2(j)}.$$

Аналогичные формулы можно получить для многомерного случая, когда у системы несколько входов:

$$y = b_1 u_1 + b_2 u_2 + \dots + b_n u_n = U^T B;$$

$$U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}; \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Эти формулы будут аналогичны (9.6), но в матричном виде:

$$\begin{aligned} \hat{B}(i) &= \hat{B}(i-1) + P(i)U(i)(y(i) - U^T(i)\hat{B}(i-1)); \\ P^{-1}(i) &= P^{-1}(i-1) + U(i)U^T(i), \end{aligned} \quad (9.7)$$

где

$$U(i) = \begin{pmatrix} u_1(i) \\ \vdots \\ u_n(i) \end{pmatrix}; \quad \hat{B}(i) = \begin{pmatrix} \hat{b}_1(i) \\ \vdots \\ \hat{b}_n(i) \end{pmatrix},$$

$P(i)$ – матрица, имеющая размерность произведения $U(i)U^T(i)$,

$$(U(i)_{[n,1]}U^T(i)_{[1,n]} = U(i)U^T(i)_{[n,n]}),$$

т.е. это квадратная матрица порядка n – числа параметров.

Рекуррентными формулами можно пользоваться по-разному:

1. Выполнить n измерений, определить $\hat{B}(n)$ и далее уточнять оценку по мере накопления опытных данных.

2. Задаться начальным вектором $\hat{B}(0)$ и уточнять его.

Начальное значение $\hat{B}(0)$ может быть произвольным, но лучше задавать его по возможности более близким к реальным значениям параметров. Тогда процесс оценивания \hat{B} сходится быстрее.

В начале процесса весовой коэффициент $P(i)U(i)$ должен быть большим или элементы $P^{-1}(i)$ должны быть малыми.

Обычно задают $P_{(0)}^{-1} = \frac{1}{\varepsilon} E$, где $\varepsilon = 10, \dots, \max$.

Задавать $P^{-1}(0) = 0$ нельзя, иначе $\hat{B}(1) = \infty, \hat{B}(2) = \infty$ и т.д.

9.5. Оценивание разброса коэффициентов модели

Ввиду случайности наблюдаемой выходной величины y (за счёт влияния неучтённых факторов) и ошибок измерений и вычислений коэффициенты b_j также случайны. Они, по сути дела, являются оценками истинных коэффициентов, и это часто в литературе подчёркивают обозначением \hat{b} и \hat{B} .

Разброс векторов оценивают не просто дисперсиями их элементов, а более общими показателями, учитывающими и зависимость элементов – **ковариациями**. В совокупности ковариации образуют так называемую ковариационную матрицу.

Для вектора $B = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ \vdots \\ b_n \end{pmatrix}$ ковариационная матрица обозначается $D[B]$ (как дис-

персия вектора B) и определяется следующим образом:

$$D[B] = \begin{pmatrix} \text{cov}(b_0, b_0) & \text{cov}(b_0, b_1) & \cdots & \text{cov}(b_0, b_n) \\ \text{cov}(b_1, b_0) & \text{cov}(b_1, b_1) & \cdots & \text{cov}(b_1, b_n) \\ \dots\dots\dots & \vdots & & \vdots \\ \text{cov}(b_n, b_0) & \text{cov}(b_n, b_1) & & \text{cov}(b_n, b_n) \end{pmatrix}. \quad (9.8)$$

Диагональные элементы – это дисперсии, так как $\text{cov}(x, x) = D[x]$, т.е.

$$\text{cov}(b_0, b_0) = D[b_0], \text{cov}(b_1, b_1) = D[b_1], \dots, \text{cov}(b_n, b_n) = D[b_n].$$

По определению, $\text{cov}(b_i, b_j) = M[(b_i - M[b_i])(b_j - M[b_j])]$ – математическое ожидание от произведения центрированных случайных величин.

Следовательно, ковариационная матрица

$$D[B] = M[(B - M[B])(B - M[B])^T] =$$

$$= M \begin{bmatrix} b_0 - M[b_0] \\ b_1 - M[b_1] \\ \vdots \\ b_n - M[b_n] \end{bmatrix} (b_0 - M[b_0], b_1 - M[b_1], \dots, b_n - M[b_n]).$$

Если \mathbf{A} и \mathbf{C} – матрицы постоянного коэффициента, а \mathbf{Z} – матрица случайных параметров, то

$$M[\mathbf{AZC}] = \mathbf{A}M[\mathbf{Z}] \cdot \mathbf{C}.$$

Поэтому для определения $\mathbf{D}[\mathbf{B}]$ надо знать $\mathbf{D}[\mathbf{Y}]$.

В качестве примера рассмотрим случай независимых наблюдений (некоррелированных значений y в опытах). Пусть

$$D[Y] = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 E_{[N]},$$

т.е. разброс y в опытах один и тот же.

Если план ортогональный, то

$$D[B] = \frac{1}{N} \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix},$$

т.е. коэффициенты не коррелированы. В этом случае коэффициенты имеют ошибку, в \sqrt{N} раз меньшую, чем ошибка измерения y .

9.6. Идентификация параметров динамических систем в режиме принудительных информационных колебаний

При эксплуатации динамических систем некоторая часть h из общего числа m их параметров q_k ($k = 1(1)m$) может считаться стационарной или известной по результатам ранее проведенного контроля (например, статического контроля коэффициентов передачи блоков или контроля отдельных узлов путем декомпозиции системы). В этом случае задача параметрической динамической идентификации сводится к определению оставшейся части $(m - h)$ параметров, характеризующих, например, динамические свойства системы или недоступные для контроля узлы системы (например, ввиду отсутствия датчиков, измеряющих параметры сигналов на входе и выходе данного узла).

Основу предлагаемого подхода к решению задачи составляют объективно существующие связи искомых $m - h$ параметров системы с параметрами автоколебаний при включении в цепь обратной связи двухпозиционного нелинейного звена с симметричной петлей гистерезиса. Аналитические соотношения, определяющие эти связи, устанавливаются при теоретическом исследовании замкнутой нелинейной системы методами припасовывания или гармонической линеаризации.

Отличительной особенностью подхода, по сравнению с известными подходами идентификации колебательных систем [1], является то, что с помощью нелинейного двухпозиционного звена с регулируемым гистерезисом удастся получить систему уравнений, нелинейных относительно идентифицируемых параметров. Это позволяет расширить возможности идентификации ДС в режиме колебаний при большом числе идентифицируемых параметров ($m - h \geq 3$).

При использовании метода гармонической линеаризации параметрами, характеризующими процессы функционирования нелинейных стохастических систем, являются амплитуда a , частота ω и постоянная составляющая y^0 гармонического сигнала на входе нелинейного звена [3,4].

Рассмотрим в общем виде методику составления и решения уравнений, которые связывают искомые параметры идентифицируемых динамических систем с допустимыми для измерений параметрами процессов их функционирования на примере некоторой линейной одномерной стохастической системы.

Уравнения состояния, наблюдения и управления системы при включении в цепь обратной связи нелинейного звена с уровнем переключения b_j будут представляться в виде:

$$\dot{Z} = F(Z, Q, \eta, x, t), \quad (9.9)$$

$$Y = \varphi(Z, Q, \chi, t), \quad (9.10)$$

$$X = \Gamma(Y, C, b_j), \quad (9.11)$$

где

$Z \in \Omega_Z \subset R^v$ – вектор состояния; $Q \in \Omega_Q \subset R^m$ – вектор параметров системы; $X \in \Omega_X \subset R^v$ – вектор управления; $\eta \in \Omega_\eta \subset R^\mu$ – вектор возмущений; $Y \in \Omega_Y \subset R^\sigma$ – вектор измерений состояния; $\chi \in \Omega_\chi \subset R^\sigma$ – вектор шумов измерений; $t \in \Omega_t = [t_0, t_\varepsilon]$ – текущее время; $F(\dots): R^v \times R^m \times R^\mu \times R^\sigma \times R^1 \rightarrow R^v$; $\varphi(\dots): R^v \times R^m \times R^\sigma \times R^1 \rightarrow R^\sigma$; $\Gamma(\dots): R^\sigma \times R^1 \times R^1 \rightarrow R^v$.

Функционирование системы (9.9) – (9.11) представляет собой автоколебательный процесс.

Так как при контроле наблюдаются параметры автоколебаний, то к уравнениям (9.9) – (9.11) необходимо добавить уравнение наблюдения этих параметров

$$\Pi_{\langle l \rangle j} = G_{\langle l \rangle j}(Z, Q, X, \eta, \chi, \xi, b_j, C, t), \quad (9.12)$$

где $\Pi_{\langle l \rangle j}$ – вектор параметров автоколебаний размера $\ell = 3$,

$$\Pi_{\langle l \rangle j} = (a_j, \omega_j, y_j^o)^T,$$

ξ – вектор шумов измерений параметров автоколебаний размера ℓ ,

$$\xi = (\xi_\alpha, \xi_\omega, \xi_y^o)^T,$$

$$G(\dots): R^v \times R^m \times R^u \times R^\mu \times R^\sigma \times R^\ell \times R^\ell \times R^\ell \times R^\ell \rightarrow R^\ell.$$

В указанной постановке задачу идентификации можно сформулировать следующим образом. При известном векторе значений уровней переключения нелинейного звена $B = (b_1, b_2, \dots, b_f)^T$ ($f \geq \frac{m-h}{2}$) требуется найти вектор параметров системы $Q_{\langle \lambda \rangle}$ ($\lambda = m - h$), который осуществляет перевод системы из состояния установившихся колебаний $\Pi_j(t_0)$ в состояние колебаний, определяемых $\Pi_j(t_k)$:

$$\Pi_j(t_0) = \Pi_{0j} \frac{G_j(\dots, Q, \dots, b_j, \dots)}{j=1(1)f} \Pi_j(t_k) = \Pi_j(t_k) = \Pi_{jk}. \quad (9.13)$$

В общем случае вектор функций G_j не задан явно. Однако с учетом того, что ДС при уровне переключения b_j имеет единственный стохастический предельный цикл, в установившемся режиме автоколебаний можно записать:

$$\Pi_{\langle l \rangle j} = G_{\langle l \rangle j \text{уст}}(Z, Q, X, \eta, \chi, \xi, b_j, c, t). \quad (9.14)$$

Поскольку вид функций $G_{j \text{уст}}$ при их точном задании является громоздким, то воспользуемся методом гармонической линеаризации (МГЛ) для получения приближенной оценки вектора $Q_{\langle \lambda \rangle}$. При этом будем полагать, что в ходе наблюдения определенного числа периодов автоколебаний замкнутой ДС влияние нестационарных шумов η, χ, ξ на амплитуду и частоту будет скомпенсировано за счет усреднения по интервалу наблюдения. С учетом этого допущения для установившегося режима автоколебаний согласно МГЛ характеристическое уравнение замкнутой системы можно записать в следующем виде:

$$Q_j(a_j, \omega_j, Q, b_j, c) = 0. \quad (9.15)$$

Здесь Q_j – функция характеристического полинома линеаризованной замкнутой системы, в которой нелинейное звено с гистерезисом представляется в виде [3]

$$\begin{aligned} \Gamma_j(b_j, c, a_j, \omega_j, p) &= \gamma(a_j, b_j, c) + \gamma'(a_j, b_j, c) p / \omega_j = \\ &= \frac{4c}{\pi a_j} \sqrt{1 - \frac{b_j^2}{a_j^2}} - \frac{4cb_j}{\pi a_j^2 \omega_j} p, \end{aligned} \quad (9.16)$$

где p – оператор Лапласа.

Из уравнения (9.18) выделим два компонента характеристического уравнения – вещественный и мнимый:

$$\begin{cases} \operatorname{Re}_j(a_j, \omega_j, Q, b_j, c) = 0; \\ \operatorname{Im}_j(a_j, \omega_j, Q, b_j, c) = 0. \end{cases} \quad (9.17)$$

Изменяя ширину гистерезисной петли нелинейного звена, можно дополнить систему (9.17) новыми уравнениями баланса до общего их числа, равного числу идентифицируемых параметров системы $\lambda = m - h$:

$$\Psi_{\langle \lambda \rangle}(A_{\langle f \rangle}, \Omega_{\langle f \rangle}, Q_{\langle m \rangle}, B_{\langle f \rangle}, c) = 0, \quad (9.18)$$

где $\Psi_{\langle \lambda \rangle}$ – вектор функций Re и Im , $\Psi_{\langle \lambda \rangle} = (\psi_1, \psi_2, \dots, \psi_\lambda)^T$; $A_{\langle f \rangle}$ – вектор амплитуд автоколебаний, $A_{\langle f \rangle} = (a_1, a_2, \dots, a_f)^T$; $\Omega_{\langle f \rangle}$ – вектор частот автоколебаний, $\Omega_{\langle f \rangle} = (\omega_1, \omega_2, \dots, \omega_f)^T$; $B_{\langle f \rangle}$ – вектор уровней переключения выходного сигнала нелинейного звена, $B_{\langle f \rangle} = (b_1, b_2, \dots, b_f)^T$; f – число экспериментов, $f \geq \frac{\lambda}{2}$.

По результатам измерений $A_{\langle f \rangle}, \Omega_{\langle f \rangle}$ и заданному $B_{\langle f \rangle}$, а также по априорной информации о h параметрах системы при решении уравнений (9.18) получим значения λ неизвестных параметров q_1, \dots, q_λ .

Отметим, что система уравнений (9.18) в общем виде представляет собой систему нелинейных трансцендентных уравнений. Для решения такой системы необходимо применять итерационные методы решения нелинейных уравнений. Наиболее общим методом решения систем уравнений вида (9.18) является метод наискорейшего спуска [10]. В отличие от метода Ньютона он всегда сходится к решению, если начальное приближение лежит в области его притяжения, т.е. не расположено на «обратном склоне» поверхности

$$U(Q) = \sum_{i=1}^{\lambda} \rho_i [\psi_i(Q)]^2 = 0,$$

где ρ – некоторые положительные числа (чаще всего единицы).

Метод наискорейшего спуска целесообразно применять в рассматриваемой задаче параметрической идентификации, так как наличие априорной информации о приближенных значениях (номинальных или полученных при предыдущей проверке технического состояния) искомых параметров $q_j (j = 1, \dots, \lambda)$ исключает возможность получения решений, дающих локальный экстремум функции $U(Q)$.

Вычисления в методе наискорейшего спуска сводятся к построению последовательности групп чисел $q_{1\mu}, q_{2\mu}, \dots, q_{\lambda\mu}$ ($\mu = 1, 2, 3, \dots$), исходя из группы $q_{10}, q_{20}, \dots, q_{\lambda 0}$, представляющей собой совокупность начальных значений параметров.

Вычисления производятся по формулам

$$Q_{\mu+1} = Q_{\mu} - \tau_{\mu} W_{\mu}^T \Psi_{\mu}, \quad (9.19)$$

где W_{μ} – матрица Якоби вектор-функции $\Psi_{\lambda}(Q_{\mu})$,

$$W_{\mu} = \frac{d\Psi_{\mu}}{dQ} \Big|_{Q=Q_{\mu}} = \begin{vmatrix} \frac{d\Psi_{1\mu}}{dq_{1\mu}} & \frac{d\Psi_{1\mu}}{dq_{2\mu}} & \dots & \frac{d\Psi_{1\mu}}{dq_{\lambda\mu}} \\ \frac{d\Psi_{2\mu}}{dq_{1\mu}} & \frac{d\Psi_{2\mu}}{dq_{2\mu}} & \dots & \frac{d\Psi_{2\mu}}{dq_{\lambda\mu}} \\ \dots & \dots & \dots & \dots \\ \frac{d\Psi_{\lambda\mu}}{dq_{1\mu}} & \frac{d\Psi_{\lambda\mu}}{dq_{2\mu}} & \dots & \frac{d\Psi_{\lambda\mu}}{dq_{\lambda\mu}} \end{vmatrix};$$

τ_{μ} – коэффициент итерации метода наискорейшего спуска:

$$\tau_{\mu} = \frac{1}{2} \frac{(\Psi_{\mu}, W_{\mu} W_{\mu}^T \Psi_{\mu})}{(W_{\mu} W_{\mu}^T \Psi_{\mu}, W_{\mu} W_{\mu}^T \Psi_{\mu})}. \quad (9.20)$$

Индекс μ в выражениях (9.19), (9.20) означает вычисление соответствующих векторов и матриц при $q_j = q_{j\mu}$ ($j = 1(1)\lambda$).

Из формул (9.19) и (9.20) видно еще одно преимущество метода наискорейшего спуска – он не требует осуществления трудоемкой операции обращения матриц, из-за которой часто в процесс идентификации вносятся дополнительные погрешности, связанные с ограниченной разрядностью вычислительных средств.

Для составления систем уравнений идентификации могут применяться другие методы описания колебаний, например метод припасовывания [2, 6]. Характерной особенностью идентификации на основе информационных колебаний является возможность воспользоваться простым арифметическим усреднением результатов измерений на определенном количестве наблюдаемых периодов колебаний, что позволят получить приемлемые оценки параметров в условиях нестационарных шумов и возмущений, в том числе окрашенных (с ненулевым математическим ожиданием).

СПИСОК ЛИТЕРАТУРЫ

1. *Арсеньев В.Н., Барановский А.М., Мануйлов Ю.С.* Методы и алгоритмы исследования и разработки автоматических систем управления. – МО СССР, 1989. – 186 с.
2. *Барановский А.М.* Активная идентификация систем стабилизации// Изв.Вуз; Приборостроение, 1997. – №7. – С. 31–34.
3. *Барановский А.М.* Алгоритм устойчивой идентификации параметров динамических систем в режиме принудительных информационных колебаний// Сб. алгоритмов и программ типовых задач/ под ред. И.А. Кудряшова. –МО РФ, 2006. – Вып. 26. – С. 3–15.
4. *Барановский А.М.* Параметрическая идентификация типовых динамических звеньев космических средств в автогенераторном режиме// Проблемные вопросы проектирования и эксплуатации бортовых и наземных систем управления ракетно-космической техники РВСН: тез. докл. – МО РФ, 1999. – С. 88–89.
5. *Барановский А.М., Яфракс М.Ф.* Об одном подходе к организации функционального контроля и диагностики // Вопросы анализа и синтеза систем управления, контроля и диагностики. – МО СССР, 1990. – С. 4–30.
6. *Бесекерский В.А.* Цифровые автоматические системы. – М.: Наука, 1976. – 576 с.
7. *Бокс Д., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. – М.: Мир, 1974. – Вып. 1. – 406 с.
8. *Грон Д.* Методы идентификации систем. – М.: Мир, 1979. – 302 с.
9. *Алексеев А.А., Кораблев Ю.А., Шестопалов М.Ю.* Идентификация и диагностика систем: учеб. для студентов высш. учеб. заведений. – М.: Издательский центр «Академия», 2009. – 352 с.
10. *Корн Г., Корн М.* Справочник по математике для научных работников и инженеров. – М.: Мир, 1982. – 831 с.
11. *Ланкастер П.* Теория матриц. – М.: Наука, 1978. – 280 с.
12. *Ли Р.* Оптимальные оценки, определение характеристик и управление. – М.: Наука, 1966. – 190 с.
13. *Линник Ю.В.* Метод наименьших квадратов и основы теории обработки наблюдений. – 2-е изд., доп. и испр. – М.: Физматиздат, 1962. – 349 с.
14. *Льюнг Л.* Идентификация систем. Теория для пользователя. – М.: Наука, 1991. – 432 с.
15. Математическая теория планирования эксперимента/ под ред. С.М. Ермакова. – М.: Наука, 1983. – 392 с.

16. *Огарков М.А.* Методы статистического оценивания параметров случайных процессов. – М.: Энергоатомиздат, 1990. – 208 с.
17. *Райбман Н.С.* Что такое идентификация? – М.: Наука, 1970. – 118 с.
18. *Сейдж Э.П., Мелса Дж.Л.* Идентификация систем управления. – М.: Наука, 1974. – 248 с.
19. *Сейдж Э.П., Мелса Дж.Л.* Теория оценивания и ее применение в связи и управлении. – М.: Связь, 1976. – 496 с.
20. Справочник по теории автоматического управления/ под ред. А.А. Красовского – М.: Наука. Гл. ред. физ.-мат. лит., 1987. – 712 с.
21. Теория управления/ А.А. Алексеев, Д.Х. Имаев, Н.Н. Кузьмин, В.Б. Яковлев – СПб.: Изд – во СПбГЭТУ "ЛЭТИ", 1999. – 435 с.
22. *Штейнберг Ш.Е.* Идентификация в системах управления. – М.: Энергоатомиздат, 1987. – 80 с.
23. *Эйкхофф П.* Основы идентификации систем управления. – М.: Мир, 1975. – 686 с.