

Библиотека

по

А

В

Т

О

М

А

Т

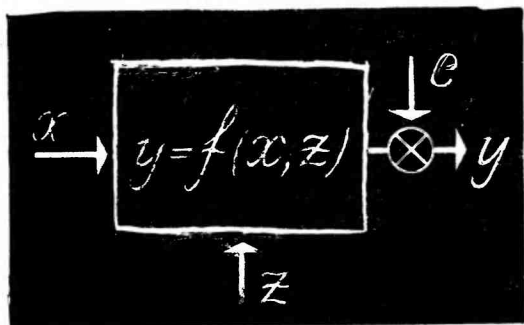
И

К

Е

В.П.Бородюк и Э.К.Лецкий

СТАТИСТИЧЕСКОЕ ОПИСАНИЕ ПРОМЫШЛЕННЫХ ОБЪЕКТОВ



ИЗДАТЕЛЬСТВО «ЭНЕРГИЯ»

В. П. БОРОДЮК И Э. К. ЛЕЦКИЙ

СТАТИСТИЧЕСКОЕ
ОПИСАНИЕ
ПРОМЫШЛЕННЫХ
ОБЪЕКТОВ



«ЭНЕРГИЯ»

МОСКВА 1971

6П2.12
Б 82
УДК 62-50

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

И. В. Антик, Г. Т. Артамонов, А. И. Бертинов, А. А. Воронов,
В. С. Малов, Л. М. Закс, В. Э. Низе, О. В. Слежановский,
Б. С. Сотсков, Ф. Е. Темников, М. Г. Чиликин, А. С. Шаталов

Бородюк В. П. и Лецкий Э. К.

Б 82 Статистическое описание промышленных объектов. М., «Энергия», 1971.

112 с. с ил. (Б-ка по автоматике. Вып. 453).

В книге рассмотрены вопросы применения методов математической статистики для построения статистического математического описания сложных объектов. Основное внимание уделено методологическим особенностям использования так называемых «пассивных» методов идентификации в условиях реального промышленного объекта.

Книга рассчитана на инженеров, занятых решением практических задач автоматического управления.

3-3-13

413-70

6П2.12

ПРЕДИСЛОВИЕ

Проблема математического описания промышленного объекта приобрела особую остроту в связи с использованием управляющих вычислительных машин для автоматизации производственных процессов. По существу математическое описание объекта — это тот фундамент, на котором строится система автоматического управления. Этим объясняется пристальный интерес, проявляемый специалистами различных отраслей промышленности к проблеме построения работоспособного математического описания.

Существует несколько путей решения указанной задачи. Авторы остановились на статистических методах, использующих в качестве исходного материала данные, собранные по способу «пассивного» эксперимента, т. е. в режиме нормальной эксплуатации объекта. По мнению авторов, такой подход предпочтителен, когда речь идет о построении математического описания действующего промышленного объекта.

Основное внимание в книге уделено особенностям использования некоторых традиционных методов математической статистики в условиях промышленного объекта.

Книга рассчитана на специалистов, работающих в области автоматизации производственных процессов, знакомых с основными понятиями математической статистики.

Данное в конце книги приложение преследует две цели: напомнить читателю некоторые сведения из матричной алгебры, статистики случайных величин и случайных процессов и ввести его в круг понятий, используемых в работе.

Авторы выражают признательность доктору техн. наук, проф. Г. К. Кругу и канд. техн. наук Г. Ф. Филаретову, советы и замечания которых были очень полезны при подготовке материалов книги.

Главы 1 и 2 написаны В. П. Бородюком, гл. 3 и 4 написаны Э. К. Лецким.

Все критические замечания и пожелания авторы примут с благодарностью.

Авторы

ВВЕДЕНИЕ

Эффективное управление промышленным объектом возможно в том случае, когда основные закономерности, присущие объекту, представлены в виде математического описания. Построение математического описания (идентификация) объекта является первым этапом синтеза системы автоматического управления. Полученная модель позволяет разработать алгоритм управления и определить структуру и оптимальные характеристики управляющего устройства. В некоторых случаях такая модель может быть использована непосредственно в системе управления.

В общем виде математическое описание представляет собой совокупность уравнений и ограничивающих условий, которые в количественной форме описывают статические и динамические связи между технологическими переменными объекта.

Современные технологические процессы представляют собой сложные объекты, входные и выходные переменные которых, а также параметры, характеризующие внутреннее состояние объекта, зависят от многочисленных, зачастую трудно уловимых факторов. Поэтому получение работоспособного математического описания представляет собой сложную исследовательскую задачу.

Существуют два подхода к ее решению: детерминистический и статистический.

В основе детерминистического подхода лежат теоретические представления о природе исследуемого процесса. Искомые уравнения связи выводятся из рассмотрения физико-химических закономерностей явлений, протекающих в объекте. Для этого используются алгебраические методы, дифференциальные уравнения, уравнения в конечных разностях. Так, законы сохранения энергии и массы формулируются в алгебраической форме с помощью уравнений теплового и материального балансов. Кинетика реакции и процесс теплопередачи хорошо описываются дифференциальными уравнениями. Процессы перегонки можно представить уравнениями в конечных разностях. Однако никакое теоретическое рассмотрение не в состоянии учесть все многообразие реально действующих факторов, а потому теоретическое математическое описание в значительной степени утрачивает силу при переходе к реальным условиям промышленного объекта.

В последние годы интенсивное развитие получил статистический подход, позволяющий извлекать нужные результаты при неполной информации о механизме процесса. Статистический подход основан на обработке экспериментальных данных, собранных непосредственно на действующем объекте. Для обработки данных используются методы теории вероятностей и математической статистики.

Применительно к задачам идентификации этот подход оказался весьма плодотворным. В настоящее время накоплена обширная ли-

температура, посвященная различным статистическим методам построения математического описания. Многочисленность методов решения задачи обусловлена, с одной стороны, различиями в постановке задачи, а с другой стороны, разнообразием свойств самих объектов исследования. С точки зрения свойств объекта пригодность того или иного метода определяется такими особенностями, как линейность или нелинейность характеристик, степень выраженности динамических свойств, уровень случайных помех, наличие дрейфа, возможность внесения искусственных возмущений и т. д.

По способу накопления экспериментальных данных статистические методы можно разделить на активные и пассивные.

Активный эксперимент основан на искусственных возмущениях, которые вводятся в объект по заранее спланированной программе. Разработана методика составления оптимальных планов, позволяющая быстро вскрывать нужные эффекты, целенаправленно продвигаясь к области наилучшего режима [Л. 1—3], строить модели процесса, адекватные результатам эксперимента [Л. 4]. Однако в производственных условиях активный эксперимент имеет ряд существенных ограничений. Чтобы исследуемый эффект не «тонет» в естественном шуме, величина пробных воздействий должна быть значительной, а это сопряжено с опасностью срыва технологического режима. Кроме того, в промышленных условиях часто имеются неуправляемые переменные, которые необходимо включать в рассмотрение и использовать при построении модели (такие, как качественный состав сырья, примеси в реагентах и т. д.).

Способ пассивного наблюдения предполагает регистрацию технологических параметров в режиме нормальной работы объекта без внесения преднамеренных возмущений. Этот способ удлинит время эксперимента, однако он оказывается экономически оправданным, когда статистическому исследованию подвергается реальный промышленный объект с непрерывным производством дорогостоящего продукта, а иногда и единственно возможным. Дополнительным преимуществом этого способа является возможность использовать обширный архивный материал предприятия.

Настоящая работа посвящена вопросам экспериментального получения статистического математического описания промышленного процесса с использованием данных, собираемых в режиме нормальной работы объекта. Статистическое математическое описание, методы получения которого здесь рассматриваются, представляет собой полином (линейный или нелинейный), устанавливающий для каждого фиксированного момента времени зависимость между выходной переменной объекта и остальными контролируемыми переменными. Если эта зависимость остается одной и той же в любой момент времени, то исследуемый объект является стационарным или несклонным к дрейфу.

Вопросы использования метода множественной регрессии для построения математической модели стационарного объекта рассматриваются в гл. 2.

Если коэффициенты полиномиальной модели объекта изменяются во времени, то такой объект называется нестационарным или дрейфующим. В этом случае задача сводится к получению оценок коэффициентов полинома для определенного момента времени. Методы решения такой задачи рассматриваются в гл. 3 и 4.

Метод множественной регрессии (гл. 1), по существу сочетает два математических метода: хорошо известный метод наименьших

квадратов и метод статистической оценки параметров, получивший развитие благодаря основополагающим работам Фишера и Стюдента.

Хотя теория метода множественной регрессии разработана давно, сложность и громоздкость вычислительной процедуры долгое время служили препятствием для проникновения его в практику инженерных исследований, и лишь внедрение электронных вычислительных машин позволило привлечь этот метод к решению сложных практических задач.

Анализ многочисленных публикаций по практическому применению метода множественной регрессии (см., например, [Л. 2, 3]) показывает, что в большинстве случаев для построения статистического математического описания промышленных объектов непрерывного действия исследователи пользуются регрессионной моделью с независимыми нормально распределенными ошибками и равноточными измерениями.

Эта модель позволяет провести полный статистический анализ эмпирического уравнения регрессии (построение доверительных интервалов для оценок, проверка различных статистических гипотез, определение надежности предсказания и т. д.). Такой анализ является основой для технологической интерпретации и практического использования математического описания исследуемого объекта.

В некоторых специальных случаях применяются более сложные регрессионные модели [Л. 39, 50, 51].

К настоящему времени опубликовано значительное число работ, в которых множественный регрессионный анализ применялся для нахождения математического описания производственных процессов. Наряду с работами, содержащими положительный опыт применения метода регрессионного анализа, отмечались случаи, когда достичь желаемой точности математического описания не удавалось. Последнее обстоятельство послужило поводом для скептического отношения некоторых авторов к методу множественной регрессии как средству идентификации промышленных объектов. Основная причина неудач заключается в механическом приложении классической схемы регрессионного анализа к решению задачи идентификации промышленного объекта без учета специфических особенностей производственных условий. Эти особенности весьма существенны и их нельзя игнорировать без риска получить отрицательный результат. К ним относятся:

- непрерывный характер изменения технологических переменных;
- сравнительно низкая точность промышленных методов контроля и регистрации данных;

- узкие диапазоны изменения технологических переменных, зачастую соизмеримые с погрешностью контрольно-измерительной аппаратуры;

- значительные запаздывания в объекте, объясняющиеся отчасти транспортным перемещением реагентов, а отчасти инерционным характером самих процессов [Л. 19];

- влияние сезонных и суточных изменений окружающей среды;
- неконтролируемое изменение качественного состава сырья и переклечение режимов при переходе с одного сорта продукции на другой и т. д.

Многообразие и разнородный характер перечисленных факторов требуют разработки специальных мер, обеспечивающих работоспособность метода множественной регрессии в условиях промыш-

ленного объекта, а иногда применения иных методов решения (например, в случае дрейфующих объектов).

При неконтролируемом дрейфе характеристик объекта найденную модель необходимо постоянно корректировать, используя вновь поступающую информацию. Чтобы решать эту задачу наилучшим образом, необходимо знать свойства процесса изменения коэффициентов модели (например, корреляционные функции дрейфующих параметров). Тогда оптимальный алгоритм построения оценок дрейфующих коэффициентов может быть найден аналитически путем

использования методов теории фильтрации [Л. 5]. В практических задачах, связанных с описанием промышленных объектов, информация о свойствах дрейфа почти всегда отсутствует. Поэтому в настоящей работе рассматриваются методы решения, в которых оптимальные свойства алгоритма достигаются в рамках заданной структуры путем наилучшего выбора значений параметров алгоритма. К ним относятся метод текущего регрессионного анализа, последовательные процедуры метода стохастической аппроксимации, метод кусочно-линейной аппроксимации. В гл. 3 и 4 эти методы описаны, и приведены алгоритмы решения задач их оптимальной настройки по конечным выборкам наблюдений, а также необходимые формулы для аналитической настройки в случаях, когда известны статистические свойства процесса изменения дрейфующих параметров.

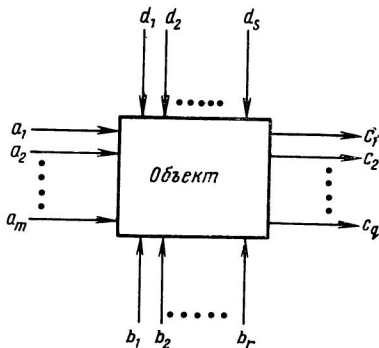


Рис. 1-1. Структурная схема сложного объекта.

описаны, и приведены алгоритмы решения задач их оптимальной настройки по конечным выборкам наблюдений, а также необходимые формулы для аналитической настройки в случаях, когда известны статистические свойства процесса изменения дрейфующих параметров.

Сфера применения методов экспериментальных исследований чрезвычайно широка. Исследуемыми объектами могут быть флотационная ванна, агломерационная фабрика, конверторная установка производства стали, промышленные печи (обжига, прокаточные, маршевые), различные аппараты химической технологии (реакторы, установки полимеризации и ректификации), объекты цементной, бумажной, пищевой и других отраслей промышленности.

Вне зависимости от физико-химической сущности явлений, происходящих в исследуемом процессе, всю совокупность параметров, определяющих текущее состояние объекта, можно разбить на четыре группы (рис. 1-1):

1. Группа $A = (a_1, a_2, \dots, a_m)$. В эту группу объединены параметры, характеризующие качество и количество исходных продуктов. Последние представляют собой первичное сырье или продукцию предыдущего звена технологической цепи. К этой группе параметров относятся данные лабораторных испытаний, результаты химических анализов, показания анализирующих приборов и т. д. Значения каждого параметра ограничены технологическим регламентом процесса:

$$a_{j\min} \leq a_j \leq a_{j\max}; \quad j=1, 2, \dots, m.$$

2. Группа $B = (b_1, b_2, \dots, b_r)$. Вторую группу образуют параметры, характеризующие те управляющие воздействия, при помощи которых оператор поддерживает заданный технологический режим. Сюда относятся показания расходомеров, положения уставок регуляторов, положения регулирующих заслонок и т. д. Значения управляющих воздействий лимитируются техническими ограничениями:

$$b_{j\min} \leq b_j \leq b_{j\max}; \quad j = 1, 2, \dots, r.$$

3. Параметры группы $C = (c_1, c_2, \dots, c_q)$ характеризуют те обобщенные технико-экономические показатели, которыми оцениваются качество и экономическая эффективность работы объекта. Эти показатели являются определяющими для технологов при выборе технологического режима и управления объектом. Обычно задача управления сложным объектом формулируется следующим образом: в данной технологической ситуации, определяемой значением параметров группы A , найти такие значения параметров группы B , лежащие внутри соответствующих ограничений, при которых параметры $c_1, c_2, \dots, c_q$ принимают свои экстремальные значения или не выходят за пределы заданного интервала (например, максимум выхода синтезируемого продукта или минимум его себестоимости при удовлетворительном качестве продукции).

Все параметры перечисленных групп можно измерять в процессе работы объекта, и поэтому будем их называть контролируемыми.

4. Четвертую группу $D = (d_1, d_2, \dots, d_s)$ образуют неконтролируемые параметры. Они характеризуют действие возмущающих факторов, таких, как суточные и сезонные изменения окружающей среды, присутствие случайных примесей в исходных продуктах, старение оборудования и т. п., а также влияние тех переменных процесса, которые недоступны количественному измерению. Наличие неконтролируемых переменных, медленно изменяющихся во времени, вызывает дрейф характеристик объекта.

В зависимости от требуемой детализации рассмотренная структурная модель сложного объекта может воспроизводить различные участки производства — от цеха до отдельной установки или агрегата.

Задача построения статического математического описания объекта состоит в нахождении зависимости между каждой из выходных переменных объекта и остальными контролируемыми переменными. Эту зависимость будем искать на основе данных пассивного эксперимента, представляющих результаты измерения значений переменных объекта в определенные моменты времени в режиме его нормальной эксплуатации.

ГЛАВА ПЕРВАЯ

ЛИНЕЙНАЯ МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

1-1. ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ МОДЕЛИ

Рассмотрим математическую модель вида

$$y = \beta_1 x_1 + \dots + \beta_k x_k + e, \quad (1-1)$$

где $\beta_1, \dots, \beta_k$ — неизвестные постоянные параметры, которые в дальнейшем будем называть регрессионными коэффициентами; $x_1, \dots, x_k$ — наблюдаемые значения некоторых переменных; y — наблюдаемая случайная величина; e — неизвестная случайная ошибка.

Задача состоит в том, чтобы по данным выборки, содержащей $n > k$ наблюдаемых значений переменных $y, x_1, \dots, x_k$, найти статистические оценки для неизвестных параметров $\beta_1, \dots, \beta_k$ и дисперсии σ^2 случайной ошибки e .

Модель (1-1) считается линейной в том смысле, что оцениваемые параметры $\beta_1, \dots, \beta_k$ входят в нее линейно относительно случайной переменной y . В то же время эта модель может быть нелинейной относительно переменных $x_1, \dots, x_k$. Например, некоторая переменная x_j в (1-1) может обозначать произведение переменных ($x_j = v z, x_j = x_i x_q, x_j = x_j^2$ и т. д.) или выражать более сложную функцию ($x_j = \log v, x_j = e^{x_q}$ и т. д.).

В этом параграфе излагаются основные допущения, рассматриваются некоторые способы построения точечных оценок $\hat{\sigma}^2$ и $\hat{\beta}_1, \dots, \hat{\beta}_k$ неизвестных параметров, исследуются важнейшие свойства полученных оценок.

Плотность распределения $\phi(y, x_1, \dots, x_k, \beta_1, \dots, \beta_k)$ случайной переменной y зависит от k известных величин $x_1, \dots, x_k$ и k неизвестных параметров $\beta_1, \dots, \beta_k$. Будем называть i -м опытом набор фиксированных значений переменных $x_{i1}, \dots, x_{ik}$. Тогда выборка образуется следующим образом. Для первого опыта $x_{11}, \dots, x_{k1}$ фиксируется значение y_1 случайной величины y из распределения $\phi_1(y, x_{11}, \dots, x_{k1}, \beta_1, \dots, \beta_k)$. Затем выбирается второй опыт $(x_{12}, \dots, x_{k2})$ и фиксируется значение y_2 из распределения $\phi_2(y, x_{12}, \dots, x_{k2}, \beta_1, \dots, \beta_k)$. Процедура продолжается до тех пор, пока не будет получено n значений переменной y для соответствующих n опытов. Значения $x_{i1}, \dots, x_{ki}$ в опытах можно считать отражением поведения некоторых переменных $x_1, \dots, x_k$, случайных или неслучайных. В дальнейшем переменные $x_1, \dots, x_k$ будем называть независимыми, а переменную y зависимую.

Для выборки размером n модель (1-1) можно представить в матричной форме:

$$Y = XB + E, \quad (1-2)$$

где

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}; \quad X = \begin{pmatrix} x_{11} & \dots & x_{k1} \\ \dots & \dots & \dots \\ x_{1n} & \dots & x_{kn} \end{pmatrix};$$

$$B = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}; \quad E = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}.$$

Матрицы X и Y будем называть матрицами наблюдений и будем полагать, что матрица X не содержит линейно связанных столбцов и имеет $\text{rang}(X) = k$.

Введем следующие основные допущения:

1) ошибки e_i в отдельных опытах не коррелированы между собой и не зависят от значений $x_1, \dots, x_k$ и $\beta_1, \dots, \beta_k$;

2) ошибки e_i в отдельных опытах имеют нормальное распределение с нулевым математическим ожиданием и одинаковой дисперсией σ^2 .

Сформулированные допущения можно выразить в виде

$$E \sim N_n(0, \sigma^2 I_n), \quad (1-3)$$

где N_n — обозначение n -мерного нормального закона распределения; I_n — единичная матрица размером $(n \times n)$; 0 — нуль-вектор математических ожиданий вектора E ; $\sigma^2 I_n$ — ковариационная матрица вектора E .

Используя свойства ошибки (1-3), выясним характер и параметры распределения случайной величины y .

Из (1-2) следует, что для фиксированной матрицы X вектор Y имеет нормальное распределение, поскольку он линейно связан с нормально распределенным вектором случайных величин E . Математическое ожидание Y

$$M[Y] = M[XB] + M[E] = XB.$$

Ковариационная матрица вектора Y имеет вид:

$$\text{cov}[Y] = M[(Y - M[Y])(Y - M[Y])^T] = \sigma^2 I_n.$$

Полученные результаты можно выразить в виде

$$Y \sim N_n(XB, \sigma^2 I_n). \quad (1-4)$$

Из (1-4) следует, что в каждом i -м опыте случайная величина y_i имеет нормальное распределение с математическим ожиданием $x_{i1}\beta_1 + \dots + x_{ik}\beta_k$ и одинаковой для всех опытов дисперсией σ^2 , случайные величины y_i и y_q ($i \neq q$) в отдельных опытах не коррелированы между собой.

Дисперсия σ^2 случайной величины y_i в отдельном i -м опыте равна дисперсии ошибки, и ее не следует путать с дисперсией переменной y относительно своего среднего σ_y^2 , выборочное [значение которой вы-

числяется по формуле

$$\hat{\sigma}_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2, \quad (1-5)$$

где

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Поскольку ошибки в отдельных опытах не коррелированы и нормально распределены, то для определения искоемых оценок $\hat{\beta}_1, \dots, \hat{\beta}_k$ и $\hat{\sigma}^2$ удобно воспользоваться методом максимума правдоподобия (см. приложение). Функция правдоподобия для вектора ошибок $\mathbf{E}$ имеет вид:

$$\begin{aligned} L(\mathbf{E}, \mathbf{B}, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\mathbf{E}^T \mathbf{E}}{2\sigma^2}\right) = \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{XB})^T (\mathbf{Y} - \mathbf{XB})\right]. \end{aligned} \quad (1-6)$$

Прологарифмируем обе части уравнения (1-6):

$$\begin{aligned} \ln L(\mathbf{E}, \mathbf{B}, \sigma^2) &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \\ &- \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{XB})^T (\mathbf{Y} - \mathbf{XB}). \end{aligned} \quad (1-7)$$

Для получения оценок максимального правдоподобия параметров $\beta_1, \dots, \beta_k$ и σ^2 приравняем нулю производные по этим параметрам от функции $\ln L(\mathbf{E}, \mathbf{B}, \sigma^2)$ (см. приложение):

$$\frac{\partial}{\partial \mathbf{B}} [\ln L(\mathbf{E}, \mathbf{B}, \sigma^2)] = \frac{1}{2\sigma^2} (2\mathbf{X}^T \mathbf{Y} - 2\mathbf{X}^T \mathbf{XB}) = 0; \quad (1-8)$$

$$\frac{\partial}{\partial \sigma^2} [\ln L(\mathbf{E}, \mathbf{B}, \sigma^2)] = -\frac{n}{2\sigma^2} + \frac{(\mathbf{Y} - \mathbf{XB})^T (\mathbf{Y} - \mathbf{XB})}{2\sigma^4} = 0. \quad (1-9)$$

Пусть $\hat{\mathbf{B}}$ и s^2 — решения уравнений (1-8) и (1-9). Тогда можно записать:

$$\mathbf{X}^T \mathbf{X} \hat{\mathbf{B}} = \mathbf{X}^T \mathbf{Y}; \quad (1-10)$$

$$s^2 = \frac{1}{n} (\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}})^T (\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}}). \quad (1-11)$$

Выражение (1-10) называется системой нормальных уравнений. Выраженная через элементы матриц наблюдений $\mathbf{X}$ и $\mathbf{Y}$ система

Этот результат будет использован в дальнейших выводах.

Исследуем несмещенность и характер распределения полученных оценок (1-11) и (1-13). Учитывая, что $S^{-1}X^T$ представляет собой матрицу неслучайных величин, для B можно записать:

$$M[\hat{B}] = M[S^{-1}X^TY] = S^{-1}X^T M[XB + E] = B, \quad (1-17)$$

т. е. $\hat{B}$ есть несмещенная оценка вектора регрессионных коэффициентов B .

Для определения смещения оценки s^2 сделаем в (1-11) подстановку $\hat{B} = S^{-1}X^TY$ и $Y = XB + E$. Тогда несложно получить

$$M[s^2] = \frac{1}{n} M[E^T(I_n - XS^{-1}X^T)E]. \quad (1-18)$$

Если A — некоторая матрица размером $(n \times n)$ с элементами a_{ij} ($i=1, \dots, n; j=1, \dots, n$), то квадратичную форму E^TAE можно представить в виде

$$E^TAE = \sum_{i=1}^n a_{ii}e_i^2 + \sum_{\substack{i=1 \\ i \neq j}}^n a_{ij}e_ie_j.$$

Тогда с учетом допущений для ошибок e_i имеем:

$$M[E^TAE] = \sigma^2 \sum_i a_{ii} = \sigma^2 \text{sp } A. \quad (1-19)$$

На основании результатов (1-19) и (1-16) для (1-18) получаем:

$$M[s^2] = \frac{1}{n} \sigma^2 \text{sp}[I_n - XS^{-1}X^T] = \frac{n-k}{n} \sigma^2 = \sigma^2 - \frac{k}{n} \sigma^2. \quad (1-20)$$

Результат (1-20) означает, что величина s^2 является смещенной оценкой σ^2 . Из (1-20) и (1-11) легко найти несмещенную оценку σ^2 , которую обозначим $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{n}{n-k} s^2 = \frac{(Y - X\hat{B})^T(Y - X\hat{B})}{n-k}. \quad (1-21)$$

Отметим, что оценки $\hat{B}$ и $\hat{\sigma}^2$, полученные методом максимума правдоподобия, помимо несмещенности, обладают и рядом других важных свойств, таких, как состоятельность, эффективность, достаточность [Л. 7].

Исследуем характер распределений оценок $\hat{B}$ и $\hat{\sigma}^2$.

Известно, что линейная функция нормально распределенной случайной величины тоже имеет нормальное распределение. Поэтому из (1-13) следует, что вектор оценок $\hat{B}$ имеет нормальный закон распределения, поскольку $(k \times n)$ матрица $S^{-1}X^T$ составлена

из фиксированных значений. Будем искать параметры распределения оценок $\hat{B}$: вектор математических ожиданий и ковариационную матрицу.

Из (1-17) следует, что вектором математических ожиданий оценок $\hat{B}$ является вектор B истинных значений регрессионных коэффициентов.

Найдем ковариационную матрицу для $\hat{B}$:

$$\text{cov} [\hat{B}] = M [(\hat{B} - B)(\hat{B} - B)^T]. \quad (1-22)$$

Произведем в (1-22) последовательную подстановку $\hat{B} = S^{-1}X^T Y$ и $Y = XB + E$. Тогда получим:

$$\begin{aligned} \text{cov} [\hat{B}] &= M [(S^{-1}X^T Y - B)(S^{-1}X^T Y - B)^T] = \\ &= M \{[S^{-1}X^T (XB + E) - B][S^{-1}X^T (XB + E) - B]^T\} = \\ &= M[(S^{-1}X^T E)(S^{-1}X^T E)^T] = M[S^{-1}X^T E E^T X S^{-1}] = S^{-1}X^T M[E E^T] X S^{-1} = \\ &= \sigma^2 S^{-1} X^T X S^{-1} = \sigma^2 S^{-1}, \end{aligned} \quad (1-23)$$

т. е.

$$\hat{B} \sim N_k(B, \sigma^2 S^{-1}).$$

Из (1-23) следует, что маргинальное распределение оценки $\hat{\beta}_j$ подчинено нормальному закону с параметрами $(\beta_j, \sigma^2 c_{jj})$:

$$\hat{\beta}_j \sim N_1(\beta_j, \sigma^2 c_{jj}), \quad (1-24)$$

где c_{jj} — соответствующий диагональный элемент обратной матрицы $C = S^{-1}$. Ковариация двух любых оценок $\hat{\beta}_j$ и $\hat{\beta}_q$:

$$\text{cov} [\hat{\beta}_j \hat{\beta}_q] = \sigma^2 c_{jq}, \quad (1-25)$$

где c_{jq} — соответствующий элемент обратной матрицы. Следует отметить одно важное обстоятельство. Поскольку оценки $\hat{\beta}_1, \dots, \hat{\beta}_k$ взаимозависимы, введение дополнительной переменной в уравнение регрессии или исключение переменной связано с полным пересчетом всех найденных оценок.

Используя формулы (1-15), выражение (1-21) можно переписать в виде

$$\hat{\sigma}^2 = \frac{1}{n-k} Y^T (I_n - X S^{-1} X^T) Y.$$

Согласно (1-16) $\text{sp}(I_n - X S^{-1} X^T) = n - k$, поэтому нетрудно показать, что величина

$$u = \frac{\hat{\sigma}^2 (n-k)}{\sigma^2} \quad (1-26)$$

имеет нецентральное $(\chi')^2$ -распределение с $(n-k)$ степенями свободы и параметром нецентральности:

$$\lambda = \frac{1}{2\sigma^2} B^T X^T (I_n - X S^{-1} X^T) X B.$$

Но

$$(\mathbf{I}_n - \mathbf{X}\mathbf{S}^{-1}\mathbf{X}^T) \mathbf{X} = \mathbf{X} - \mathbf{X}\mathbf{S}^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{0}. \quad (1-27)$$

Следовательно, величина (1-26) имеет центральное χ^2 -распределение с $(n-k)$ степенями свободы

$$\frac{\hat{\sigma}^2 (n-k)}{\sigma^2} \sim \chi^2 (n-k). \quad (1-28)$$

Не приводя доказательства, отметим, что величина $\hat{\sigma}^2$ и вектор $\hat{\mathbf{B}}$ статистически независимы [Л. 8].

Можно показать, что найденные методом наибольшего правдоподобия оценки (1-13) и (1-21) совпадают с оценками, полученными методом наименьших квадратов (МНК).

Согласно МНК искомые оценки должны соответствовать минимуму суммы квадратов ошибок:

$$\sum_{i=1}^n e_i^2 = \mathbf{E}^T \mathbf{E} = (\mathbf{Y} - \mathbf{X}\mathbf{B})^T (\mathbf{Y} - \mathbf{X}\mathbf{B}). \quad (1-29)$$

Значения вектора $\mathbf{B}$, минимизирующие выражение (1-29), определяются из условия

$$\frac{\partial}{\partial \mathbf{B}} (\mathbf{E}^T \mathbf{E}) = 0. \quad (1-30)$$

Взяв производную по $\mathbf{B}$ от правой части выражения (1-29), получим условие (1-30) в виде

$$2\mathbf{X}^T \mathbf{Y} - 2\mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{0},$$

откуда легко получить оценку вектора $\mathbf{B}$

$$\hat{\mathbf{B}} = \mathbf{S}^{-1} \mathbf{X}^T \mathbf{Y}. \quad (1-31)$$

Используя найденные методом наименьших квадратов оценки (1-31), несложно показать, что несмещенная оценка для σ^2 совпадает с оценкой (1-21).

Основные свойства МНК-оценок $\hat{\beta}_1, \dots, \hat{\beta}_k$ определяются теоремой Гаусса — Маркова, смысл которой можно сформулировать следующим образом. Если в матричном уравнении $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$ матрица $\mathbf{X}$ имеет ранг $k \leq n$, а вектор случайных ошибок удовлетворяет условиям

$$M[\mathbf{E}] = \mathbf{0}; \quad M\{\mathbf{E}\mathbf{E}^T\} = \sigma^2 \mathbf{I}_n,$$

то наилучшие в смысле минимума дисперсии линейные относительно y несмещенные оценки параметров $\beta_1, \dots, \beta_k$ получаются при наименьшей сумме квадратов ошибок. Следует отметить, что условие нормальности распределения ошибок в теореме не оговаривается, и ошибка может быть распределена произвольно. Введение дополнительного условия нормальности (1-3) позволяет провести полный статистический анализ оценок (1-13) и (1-21), а именно: исследовать вид и параметры их распределения, построить доверительные интервалы для истинных значений $\beta_1, \dots, \beta_k$ и σ^2 , осуществить проверку основных гипотез:

Если обе части каждого уравнения системы (1-32) разделить на $(n-1)$, то система нормальных уравнений принимает вид:

[illegible]

где $\hat{d}_{jq}$ и $\hat{d}_{jy}$ — выборочные ковариации соответствующих переменных:

$$\left. \begin{aligned} \hat{d}_{jq} &= \frac{1}{n-1} \sum_{i=1}^n (x_{ji} - \bar{x}_j) (x_{qi} - \bar{x}_q); \quad \hat{d}_{jj} = \hat{\sigma}_j^2; \\ \hat{d}_{jy} &= \frac{1}{n-1} \sum_{i=1}^n (x_{ji} - \bar{x}_j) (y_i - \bar{y}); \quad \hat{d}_{yy} = \hat{\sigma}_y^2. \end{aligned} \right\} \quad (1-39)$$

Рассмотрим определитель

$$\hat{D} = \begin{vmatrix} \hat{a}_{yy} & \hat{a}_{y1} & \cdots & \hat{a}_{yh} \\ \hat{a}_{1y} & \hat{a}_{11} & \cdots & \hat{a}_{1h} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{a}_{hy} & \hat{a}_{h1} & \cdots & \hat{a}_{hh} \end{vmatrix}. \quad (1-40)$$

Обозначим $\hat{D}_{jq}$ минор определителя $\hat{D}$, полученный вычеркиванием в $\hat{D}$ j -й строки и q -го столбца. Тогда, используя правило Крамера, решение системы (1-38) можно записать в виде

$$\hat{b}_j = -\frac{\hat{D}_{yj}}{\hat{D}_{yy}}; j = 1, \dots, k. \quad (1-41)$$

Переменные $y, x_1, \dots, x_k$ можно привести к единому масштабу измерения, стандартизуя все значения в матрицах наблюдений X и Y по формуле

$$y^*_{it} = \frac{y_{it} - \bar{y}}{\hat{\sigma}_y}; \quad x^*_{jt} = \frac{x_{jt} - \bar{x}_j}{\hat{\sigma}_j}; \quad j = 1, \dots, k; \quad i = 1, \dots, n. \quad (1-42)$$

Преобразование (1-42) означает, что стандартизованные переменные y^* , x^*_1 , ..., x^*_k имеют одинаковые выборочные дисперсии,

2-136
17

равные единице, и нулевые средние значения. Используя y^* , $x^*_1, \dots, x^*_k$, вместо определителя $\widehat{D}$ получаем определитель

$$\widehat{\rho} = \begin{vmatrix} 1 & \widehat{r}_{y1} & \dots & \widehat{r}_{yk} \\ \widehat{r}_{1y} & 1 & \dots & \widehat{r}_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{r}_{ky} & \widehat{r}_{k1} & \dots & 1 \end{vmatrix},$$

каждый элемент которого представляет собой выборочный коэффициент корреляции между соответствующими переменными:

$$\widehat{r}_{jq} = \frac{1}{(n-1) \widehat{\sigma}_j \widehat{\sigma}_q} \sum_{i=1}^n (x_{ji} - \bar{x}_j) (x_{qi} - \bar{x}_q);$$

$$\widehat{r}_{yj} = \frac{1}{(n-1) \widehat{\sigma}_j \widehat{\sigma}_y} \sum_{i=1}^n (y_i - \bar{y}) (x_{ji} - \bar{x}_j).$$

Оценки регрессионных коэффициентов для стандартизованного масштаба измерения переменных можно вычислить по формуле

$$\widehat{b}^*_j = - \frac{\widehat{\rho}_{yj}}{\widehat{\rho}_{yy}}.$$

Можно показать [Л. 8], что $\widehat{D}$ и $\widehat{\rho}$ связаны соотношением

$$\widehat{D} = \widehat{\sigma}_y^2 \widehat{\sigma}_1^2 \dots \widehat{\sigma}_k^2 \widehat{\rho}. \quad (1-43)$$

Используя (1-43), несложно определить зависимость между оценками регрессионных коэффициентов, вычисленными для натурального и стандартизованного масштабов измерения переменных:

$$\widehat{b}^*_j = \frac{\widehat{\sigma}_j}{\widehat{\sigma}_y} \widehat{b}_j. \quad (1-44)$$

Уравнение регрессии для стандартизованных переменных не имеет свободного члена

$$\widehat{y}^* = \widehat{b}^*_{.1} x^*_{.1} + \dots + \widehat{b}^*_{.k} x^*_{.k}$$

и удобно для оценки сравнительного влияния независимых переменных $x_1, \dots, x_k$ на зависимую переменную y . Направление эффекта, вызванного изменением x_j , определяется знаком перед коэффициентом $\widehat{b}^*_j$, а величина эффекта пропорциональна абсолютному значению коэффициента $\widehat{b}^*_j$.

Между остаточной дисперсией $\hat{\sigma}^2$ и выборочной дисперсией $\hat{\sigma}_y^2$ существует зависимость

$$\hat{\sigma}^2 = \frac{n-1}{n-p} \hat{\sigma}_y^2 (1 - \hat{R}_{y \cdot x_1, x_2, \dots, x_k}^2), \quad (1.45)$$

где $\hat{R}_{y \cdot x_1, \dots, x_k}$ — выборочный коэффициент множественной корреляции между переменными y и $x_1, \dots, x_k$.

Для вычисления $\hat{R}_{y \cdot x_1, \dots, x_k}$ можно воспользоваться формулами

$$\hat{R}_{y \cdot x_1, \dots, x_k}^2 = 1 - \frac{\hat{D}}{\hat{\sigma}_y^2 \hat{D}_{yy}}; \quad (1.46)$$

$$\hat{R}_{y \cdot x_1, \dots, x_k}^2 = \sum_{j=1}^k \hat{r}_{yj} \hat{b}_j^*. \quad (1.47)$$

Необходимо подчеркнуть, что величина $\hat{R}_{y \cdot x_1, \dots, x_k}$, введенная нами для модели множественной регрессии, лишь формально совпадает с величиной коэффициента множественной корреляции, который употребляется в многомерном корреляционном анализе [Л. 8] как мера статистической связи между случайными переменными $y, x_1, \dots, x_k$, объединенными совместным $(k+1)$ -мерным нормальным распределением.

Пример 1-1. Пусть модель (1-1) имеет вид:

$$y = \beta_1 x_1 + \beta_2 x_2 + e.$$

Тогда матрица X содержит два $(n \times 1)$ -столбца X_1 и X_2 . В матричной форме можно записать:

$$Y = XB + E = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + E.$$

Следовательно,

$$\begin{aligned} \|X^T X\| &= \begin{bmatrix} X_1^T & X_1 & X_2^T & X_2 \end{bmatrix} = \begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 \\ \Sigma x_2 x_1 & \Sigma x_2^2 \end{bmatrix}; \\ \|X^T Y\| &= \begin{bmatrix} X_1^T & Y \\ X_2^T & Y \end{bmatrix} = \begin{bmatrix} \Sigma x_1 y \\ \Sigma x_2 y \end{bmatrix}. \end{aligned}$$

Оценки $\hat{\beta}_1$ и $\hat{\beta}_2$:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 \\ \Sigma x_2 x_1 & \Sigma x_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma x_1 y \\ \Sigma x_2 y \end{bmatrix}.$$

Обратная матрица $\|X^T X\|^{-1} = S^{-1} = C$:

$$\|X^T X\|^{-1} = \frac{1}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2} \begin{bmatrix} \Sigma x_2^2 & -\Sigma x_1 x_2 \\ -\Sigma x_2 x_1 & \Sigma x_1^2 \end{bmatrix}.$$

Дисперсия оценки $\hat{\beta}_1$

$$D[\hat{\beta}_1] = \sigma^2 c_{11} = \frac{\sigma^2 \Sigma x_2^2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2}.$$

Для дисперсии оценки $\hat{\beta}_2$ в последнем выражении нужно заменить индекс 1 на 2, 2 на 1.

Ковариация оценок $\hat{\beta}_1$ и $\hat{\beta}_2$ равна:

$$\text{cov}[\hat{\beta}_1, \hat{\beta}_2] = \sigma^2 c_{12} = \frac{-\sigma^2 \Sigma x_1 x_2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2},$$

откуда видно, что $\hat{\beta}_1$ и $\hat{\beta}_2$ будут некоррелированы тогда, когда

$\sum_{i=1}^n x_{1i} x_{2i} = 0$, т. е. когда векторы X_1 и X_2 ортогональны.

Остаточная дисперсия

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2}{n - 2}.$$

Пример 1-2. Пусть в предыдущем примере все значения переменной x_1 равны единице: $X_1 \equiv 1$.

Тогда

$$|X^T X| = \begin{vmatrix} n & \Sigma x_2 \\ \Sigma x_2 & \Sigma x_2^2 \end{vmatrix}; \quad X^T Y = \begin{pmatrix} \Sigma y \\ \Sigma x_2 y \end{pmatrix}.$$

Обратная матрица $C = |X^T X|^{-1}$:

$$C = \frac{\begin{vmatrix} \Sigma x_2^2 & -\Sigma x_2 \\ -\Sigma x_2 & n \end{vmatrix}}{n \Sigma x_2^2 - (\Sigma x_2)^2} = \frac{\begin{vmatrix} \frac{1}{n} \Sigma x_2^2 & -\bar{x}_2 \\ -\bar{x}_2 & 1 \end{vmatrix}}{\Sigma x_2^2 - \frac{1}{n} (\Sigma x_2)^2} = \frac{\begin{vmatrix} \frac{1}{n} \Sigma x_2^2 & -\bar{x}_2 \\ -\bar{x}_2 & 1 \end{vmatrix}}{\Sigma (x_2 - \bar{x}_2)^2}.$$

Оценки $\hat{\beta}_1$ и $\hat{\beta}_2$ находятся из равенства

$$\begin{aligned} \hat{B} &= \begin{vmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{vmatrix} = (X^T X)^{-1} X^T Y = \\ &= \frac{1}{\Sigma (x_2 - \bar{x}_2)^2} \begin{pmatrix} \bar{y} \Sigma x_2^2 - \bar{x}_2 \Sigma x_2 y \\ -\bar{x}_2 \Sigma y + \Sigma x_2 y \end{pmatrix}, \end{aligned}$$

т. е.

$$\hat{\beta}_1 = \frac{\bar{y} \Sigma x_2^2 - \bar{x}_2 \Sigma x_2 y}{\Sigma (x_2 - \bar{x}_2)^2};$$

$$\hat{\beta}_2 = \frac{\Sigma x_2 y - \bar{x}_2 \Sigma y}{\Sigma (x_2 - \bar{x}_2)^2} = \frac{\Sigma x_2 y - n \bar{x}_2 \bar{y}}{\Sigma (x_2 - \bar{x}_2)^2}.$$

Подставив последнее выражение в формулу для $\hat{\beta}_1$, получим:

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}_2.$$

Дисперсия и ковариация оценок $\hat{\beta}_1$ и $\hat{\beta}_2$:

$$D[\hat{\beta}_1] = \frac{\sigma^2 \Sigma x_2^2}{n \Sigma (x_2 - \bar{x}_2)^2} = \sigma^2 c_{11};$$

$$D[\hat{\beta}_2] = \frac{\sigma^2}{\Sigma (x_2 - \bar{x}_2)^2} = \sigma^2 c_{22};$$

$$\text{cov}[\hat{\beta}_1, \hat{\beta}_2] = \frac{-\sigma^2 \bar{x}_2}{\Sigma (x_2 - \bar{x}_2)^2} = \sigma^2 c_{12}.$$

Последние два выражения позволяют сделать следующие выводы:

1) дисперсия оценки $\hat{\beta}_2$ и ковариация оценок $\hat{\beta}_1$ и $\hat{\beta}_2$ тем меньше, чем больше диапазон изменения переменной x_2 в матрице наблюдений:

2) оценки $\hat{\beta}_1$ и $\hat{\beta}_2$ некоррелированы, если $\bar{x}_2 = 0$.

Чтобы получить представление о том, как влияет на точность эмпирического уравнения $\hat{Y} = X\hat{\beta}$ величина дисперсии ошибки σ^2 , рассмотрим следующий пример.

Пример 1-3. Для $n=50$ значений переменных x_1, x_2, x_3 были рассчитаны 50 значений y по уравнению

$$y = 2x_1 + 3x_2 - x_3.$$

Выборочная дисперсия $\hat{\sigma}_y^2$ равна $\hat{\sigma}_y^2 = 3,3$. Далее было образовано по 50 значений переменных y_1, y_2, y_3 по формуле

$$y_j = y + e_j; \quad j = 1, 2, 3,$$

где e_j — значение случайной нормально распределенной ошибки с нулевым математическим ожиданием и соответствующей дисперсией

$$\sigma_1^2 = 1, \quad \sigma_2^2 = 4, \quad \sigma_3^2 = \frac{1}{4}.$$

По обычной формуле

$$\hat{B}_j = S^{-1} X^T Y_j; \quad j = 1, 2, 3$$

были получены следующие эмпирические уравнения регрессии и их остаточные дисперсии:

$$\hat{y}_1 = 1,13 + 1,43x_1 + 2,48x_2 - 0,97x_3; \hat{\sigma}_1^2 = 1,09;$$

$$\hat{y}_2 = 5,62 + 0,82x_1 + 2,02x_2 - 1,92x_3; \hat{\sigma}_2^2 = 3,96;$$

$$\hat{y}_3 = -0,12 + 2,11x_1 + 3,20x_2 - 0,96x_3; \hat{\sigma}_3^2 = 0,28.$$

1.2. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ МОДЕЛИ

Точечная оценка по существу представляет собой некоторую функцию от выборочных данных, которая в определенном смысле наилучшим образом оценивает неизвестное значение параметра. Практически точечная оценка всегда отличается от истинного значения параметра, поэтому важно иметь надежное представление о возможных границах истинного значения оцениваемого параметра.

Будем называть интервальной оценкой параметра θ^* такой интервал $l = \theta_2 - \theta_1$ на оси θ , который с заранее выбранной вероятностью $P = 1 - \alpha$ покрывает истинное значение оцениваемого параметра, т. е.

$$P\{\theta_1 \leq \theta^* \leq \theta_2\} = 1 - \alpha. \quad (1.48)$$

Построить интервальную оценку для параметра θ^* — значит определить по данным выборки (или по точечной оценке $\hat{\theta}$) конечные значения θ_1 и θ_2 интервала, удовлетворяющего условию (1.48).

Далее будет рассматриваться построение доверительных интервалов для β_j , σ^2 и предсказываемого значения $\hat{y}$ с использованием статистических распределений, табулированных для различного уровня значимости α .

Доверительный интервал для неизвестного значения σ^2 можно построить, используя тот факт, что величина (1.26) имеет χ^2 -распределение с $(n-k)$ степенями свободы. Если учитывать, что χ^2 -распределение несимметрично, для некоторых значений α_2 и α_1 справедливо соотношение

$$P\left[\chi_{\alpha_1}^2, (n-k) \leq \frac{\sigma^2 (n-k)}{\sigma^2} \leq \chi_{\alpha_2}^2, (n-k)\right] = \int_{\chi_{\alpha_1}^2}^{\chi_{\alpha_2}^2} \chi^2(u) du = 1 - \alpha, \quad (1.49)$$

где $\alpha = \chi_{\alpha_2}^2, (n-k) - \chi_{\alpha_1}^2, (n-k)$; здесь; $\chi_{\alpha_2}^2, (n-k)$ и $\chi_{\alpha_1}^2, (n-k)$ — табличные значения χ^2 -распределения.

Преобразовав выражение (1.49), можно получить:

$$P\left[\frac{\hat{\sigma}^2 (n-k)}{\chi_{\alpha_2}^2, (n-k)} \leq \sigma^2 \leq \frac{\hat{\sigma}^2 (n-k)}{\chi_{\alpha_1}^2, (n-k)}\right] = 1 - \alpha, \quad (1.50)$$

где в квадратных скобках $(1-\alpha)$ -доверительный интервал для ис-

тинного значения σ^2 . Ширина доверительного интервала для σ^2

$$l_{\sigma^2} = \frac{\widehat{\sigma}^2 (n-k)}{\chi_{\alpha_1}^2 (n-k)} - \frac{\widehat{\sigma}^2 (n-k)}{\chi_{\alpha_2}^2 (n-k)} = \\ = \widehat{\sigma}^2 (n-k) \left(\frac{1}{\chi_{\alpha_1}^2 (n-k)} - \frac{1}{\chi_{\alpha_2}^2 (n-k)} \right). \quad (1-51)$$

Обычно $\alpha_1 = \alpha_2 = \alpha/2$.

Для построения доверительного интервала для одного отдельно взятого регрессионного коэффициента β_j воспользуемся результатом (1-24), из которого непосредственно следует, что величина $w =$

$$= \frac{\widehat{\beta}_j - \beta_j}{\sigma \sqrt{c_{jj}}} \text{ имеет одномерный нормальный закон распределения}$$

с нулевым математическим ожиданием и единичной дисперсией. По определению (см. приложение) величина

$$\frac{\frac{\widehat{\beta}_j - \beta_j}{\sigma \sqrt{c_{jj}}}}{\sqrt{\frac{\widehat{\sigma}^2 (n-k)}{\sigma^2 (n-k)}}} = \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma} \sqrt{c_{jj}}}$$

подчинена t -распределению с $(n-k)$ степенями свободы. Тогда для уровня значимости α справедливо соотношение

$$P \left[-t_{\alpha/2, (n-k)} \leq \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma} \sqrt{c_{jj}}} \leq t_{\alpha/2, (n-k)} \right] = \int_{-t_{\alpha/2}}^{+t_{\alpha/2}} t(w) dw = 1 - \alpha, \quad (1-52)$$

где $t_{\alpha/2, (n-k)}$ — табличное значение t -распределения. После преобразования (1-52) получим:

$$P [\widehat{\beta}_j - t_{\alpha/2, (n-k)} \widehat{\sigma} \sqrt{c_{jj}} \leq \beta_j \leq \widehat{\beta}_j + t_{\alpha/2, (n-k)} \widehat{\sigma} \sqrt{c_{jj}}] = 1 - \alpha, \quad (1-53)$$

где в квадратных скобках заключен $(1-\alpha)$ -доверительный интервал для неизвестного значения β_j . Ширина доверительного интервала для β_j

$$l_{\beta_j} = 2t_{\alpha/2, (n-k)} \widehat{\sigma} \sqrt{c_{jj}}. \quad (1-54)$$

Выражение (1-53) можно интерпретировать следующим образом. Если у нас имеется множество выборок, то в среднем $100(1-\alpha)\%$ интервалов, построенных по этим выборкам, покроют истинное неизвестное значение β_j .

Следует подчеркнуть, что (1-53) справедливо для одного отдельно взятого коэффициента β_j , и нельзя утверждать, что $P=1-\alpha$

есть вероятность того, что доверительные интервалы, построенные по данным одной выборки для всех k коэффициентов β_j , $j = 1, \dots, k$, одновременно покрывают истинные значения этих коэффициентов. Если бы интервалы были независимы, то вероятность такого утверждения была бы равна $(1-\alpha)^k$. Однако доверительные интервалы нельзя считать независимыми, поскольку коррелированы оценки $\hat{\beta}_j$, что следует из результата (1-23).

Теперь перейдем к определению интервальной оценки для предсказываемого значения $\hat{y}_{(0)}$, соответствующего подстановке в эмпирическое уравнение регрессии

$$\hat{y} = \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k \quad (1-55)$$

некоторого набора фиксированных значений $X_{(0)}^T = (x_{1(0)}, \dots, x_{k(0)})$, или, что то же самое, определим интервал, в котором с вероятностью $P=1-\alpha$ окажется значение величины y , имеющей плотность распределения

$$\varphi[y, x_1 = x_{1(0)}, \dots, x_k = x_{k(0)}, \beta_1 = \hat{\beta}_1, \dots, \beta_k = \hat{\beta}_k].$$

Если бы были известны истинные значения $\beta_1, \dots, \beta_k$ и σ^2 , то величина $y_{(0)}$ соответствовала бы уравнению

$$y_{(0)} = \sum_{j=1}^k \beta_j x_{j(0)} + e_{(0)} \quad (1-56)$$

и имела согласно (1-4) нормальное распределение с параметрами

$$y_{(0)} \sim N_1(X_{(0)}^T B, \sigma^2). \quad (1-57)$$

Тогда доверительный интервал для значения y , соответствующего опыту $X_{(0)}^T$, можно было бы построить сравнительно просто, например, используя тот факт, что примерно 95% нормально распределенной случайной величины заключены в интервале $\pm 2\sigma$:

$$P[X_{(0)}^T B - 2\sigma \leq y_{(0)} \leq X_{(0)}^T B + 2\sigma] \approx 0,95,$$

что для практических целей вполне надежно.

В действительности нам неизвестны истинные значения σ^2 и $\beta_1, \dots, \beta_k$; поэтому при построении доверительного интервала для y , соответствующего опыту $X_{(0)}$, мы вынуждены воспользоваться точечными оценками $\hat{\beta}_1, \dots, \hat{\beta}_k$ и σ^2 .

Рассмотрим разность величин $y_{(0)}$ и $\hat{y}_{(0)}$, получающихся при подстановке вектора $X_{(0)}$ в уравнения (1-55) и (1-56):

$$z = \hat{y}_{(0)} - y_{(0)} = X_{(0)}^T \hat{B} - X_{(0)}^T B - e_{(0)}.$$

Случайная величина z представляет собой линейную комбинацию случайных величин $\hat{B}$ и e , распределенных нормально. Следовательно,

z также имеет нормальное распределение. Определим параметры нормального распределения величины z . Математическое ожидание

$$M[z] = X_{(0)}^T M[\hat{B}] - X_{(0)}^T B - M[e_{(0)}] = 0. \quad (1-58)$$

Поскольку B и e независимы, для дисперсии z можно записать:

$$D[z] = D[X_{(0)}^T \hat{B}] + D[e_{(0)}] = M[(X_{(0)}^T \hat{B} - X_{(0)}^T B) \times \\ \times (X_{(0)}^T \hat{B} - X_{(0)}^T B)^T] + \sigma^2. \quad (1-59)$$

Первое слагаемое выражения (1-58) можно преобразовать следующим образом:

$$M[(X_{(0)}^T \hat{B} - X_{(0)}^T B) (X_{(0)}^T \hat{B} - X_{(0)}^T B)^T] = \\ = M[X_{(0)}^T (\hat{B} - B) (\hat{B} - B)^T X_{(0)}] = X_{(0)}^T M[(\hat{B} - B) (\hat{B} - B)^T] X_{(0)} = \\ = \sigma^2 X_{(0)}^T S^{-1} X_{(0)}.$$

Окончательно имеем:

$$D[z] = \sigma^2 X_{(0)}^T S^{-1} X_{(0)} + \sigma^2. \quad (1-60)$$

В (1-60) первое слагаемое дисперсии ошибки предсказания обусловлено выборочной неточностью определения оценок коэффициентов $\hat{\beta}_1, \dots, \hat{\beta}_k$; второе слагаемое учитывает действие случайной ошибки e .

На основании результатов (1-58) и (1-60) можно записать для z :

$$z \sim N_1[0, \sigma^2(1 + X_{(0)}^T S^{-1} X_{(0)})],$$

откуда непосредственно следует:

$$v = \frac{z}{\sigma \sqrt{1 + X_{(0)}^T S^{-1} X_{(0)}}} \sim N_1(0, 1).$$

Согласно определению (см. приложение) величина

$$\frac{\frac{\hat{y}_{(0)} - y_{(0)}}{\sigma \sqrt{1 + X_{(0)}^T S^{-1} X_{(0)}}}}{\sqrt{\frac{\hat{\sigma}^2(n-k)}{\sigma^2(n-k)}}} = \frac{X_{(0)}^T \hat{B} - y_{(0)}}{\hat{\sigma} \sqrt{1 + X_{(0)}^T S^{-1} X_{(0)}}}$$

имеет t -распределение с $(n-k)$ степенями свободы. Тогда для уровня значимости α можно записать вероятностное уравнение

$$P \left[-t_{\alpha/2, (n-k)} \leq \frac{y_{(0)} - X_{(0)}^T \hat{B}}{\hat{\sigma} \sqrt{1 + X_{(0)}^T S^{-1} X_{(0)}}} \leq t_{\alpha/2, (n-k)} \right] = 1 - \alpha, \quad (1-61)$$

откуда легко перейти к доверительному интервалу

$$P [X_{(0)}^T \hat{B} - t_{\alpha/2}, (n-k) \hat{\sigma} \sqrt{1 + X_{(0)}^T S^{-1} X_{(0)}} \leq y_{(0)} \leq X_{(0)}^T \hat{B} + t_{\alpha/2}, (n-k) \hat{\sigma} \sqrt{1 + X_{(0)}^T S^{-1} X_{(0)}}] = 1 - \alpha. \quad (1-62)$$

Доверительный интервал, заключенный в квадратные скобки выражения (1-62), будем называть интервалом единичного предсказания. Ширина этого интервала

$$l_{y(0)} = 2t_{\alpha/2}, (n-k) \hat{\sigma} \sqrt{1 + X_{(0)}^T S^{-1} X_{(0)}}. \quad (1-63)$$

Определим, чему равна дисперсия ошибки единичного предсказания (1-60) для случая, когда регрессионное уравнение используется в форме (1-33). Величины $\hat{y}_{(0)}$ и $y_{(0)}$ согласно уравнениям (1-36) и (1-56) можно представить в виде

$$\begin{aligned} \hat{y}_{(0)} &= \bar{y} + (X_{(0)} - \bar{X})^T \hat{\mathfrak{B}}; \\ y_{(0)} &= b_0 + X_{(0)}^T \mathfrak{B} + e_{(0)}, \end{aligned}$$

причем

$$\bar{y} = b_0 + \bar{X}^T \mathfrak{B} + \bar{e},$$

где

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i.$$

В этих выражениях $\bar{X}$ есть $(k \times 1)$ вектор средних значений $X^T = (\bar{x}_1, \dots, \bar{x}_k)$. Для разности z можно написать:

$$\begin{aligned} z &= \hat{y}_{(0)} - y_{(0)} = b_0 + \bar{X}^T \mathfrak{B} + \bar{e} + (X_{(0)} - \bar{X})^T \hat{\mathfrak{B}} - b_0 - \\ &- X_{(0)}^T \mathfrak{B} - e_{(0)} = (X_{(0)} - \bar{X})^T (\hat{\mathfrak{B}} - \mathfrak{B}) + \bar{e} - e_{(0)}. \end{aligned}$$

Тогда согласно (1-58) дисперсия z

$$\begin{aligned} D[z] &= M \{ [(X_{(0)} - \bar{X})^T \hat{\mathfrak{B}} - (X_{(0)} - \bar{X})^T \mathfrak{B}] [(X_{(0)} - \bar{X})^T \hat{\mathfrak{B}} - \\ &- (X_{(0)} - \bar{X})^T \mathfrak{B}]^T \} + M [\bar{e} - e_{(0)}]^2. \end{aligned} \quad (1-64)$$

Нетрудно показать, что первое слагаемое в (1-64) равно:

$$\sigma^2 (X_{(0)} - \bar{X})^T S_0^{-1} (X_{(0)} - \bar{X}),$$

где S_0^{-1} определяется выражением (1-34). Второе слагаемое в (1-64) равно $\sigma^2 \left(1 + \frac{1}{n}\right)$. Следовательно, дисперсия ошибки единичного

предсказания для уравнения (1-33) имеет вид:

$$D[z] = \sigma^2 \left[(\mathbf{X}_{(0)} - \bar{\mathbf{X}})^T \mathbf{S}_0^{-1} (\mathbf{X}_{(0)} - \bar{\mathbf{X}}) + 1 + \frac{1}{n} \right]. \quad (1-65)$$

Из (1-65) видно, что $D[z]$ будет минимальна тогда, когда $\mathbf{X}_{(0)} = \bar{\mathbf{X}}$. Ширина доверительного интервала ошибки предсказания в этом случае равна:

$$\Delta y_{(0)} | \mathbf{X}_{(0)} = \bar{\mathbf{X}} = 2t_{\alpha/2}, (n-k) \hat{\sigma} \sqrt{1 + \frac{1}{n}}. \quad (1-66)$$

Из (1-66) следует, что ошибка предсказания по уравнению регрессии неодинакова для различных точек, взятых внутри диапазонов изменения переменных $x_1, \dots, x_k$. Она минимальна для середины диапазона и постепенно увеличивается по мере смещения точки $\mathbf{X}_{(0)}$ к краям диапазона. Полученный результат поясним примером.

Пример 1-4. Рассмотрим эмпирическое уравнение регрессии, соответствующее модели примера 1-2:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_2.$$

Согласно (1-62) и (1-65) значение $y_{(0)}$, соответствующее опыту $x_1=1$ и $x_2=x_{2(0)}$, с вероятностью $P=1-\alpha$ заключено в интервале

$$\mathbf{X}_{(0)}^T \hat{\mathbf{B}} \pm t_{\alpha/2}, (n-2) \hat{\sigma} (1 + \mathbf{X}_{(0)}^T \mathbf{S}^{-1} \mathbf{X}_{(0)})^{1/2}.$$

Воспользуемся выражением для $\mathbf{S}^{-1}$, полученным в примере 1-2. Тогда доверительный интервал для $y_{(0)}$ можно представить в виде

$$\begin{aligned} & [\hat{\beta}_1 + \hat{\beta}_2 x_{2(0)}] \pm t_{\alpha/2}, (n-2) \hat{\sigma} \left[\frac{1}{\sum (x_2 - \bar{x}_2)^2} \times \right. \\ & \times \left\| \begin{array}{c} 1 \\ x_{2(0)} \end{array} \right\|^T \left\| \begin{array}{cc} \frac{1}{n} \sum x_2^2 & -\bar{x}_2 \\ -\bar{x}_2 & 1 \end{array} \right\| \left\| \begin{array}{c} 1 \\ x_{2(0)} \end{array} \right\| + \\ & \left. + 1 \right]^{1/2} = [\hat{\beta}_1 + \hat{\beta}_2 x_{2(0)}] \pm t_{\alpha/2}, (n-2) \hat{\sigma} \left[1 + \right. \\ & \left. + \frac{1}{n} + \frac{(x_{2(0)} - \bar{x}_2)^2}{\sum (x_2 - \bar{x}_2)^2} \right]^{1/2}. \quad (1-67) \end{aligned}$$

Выражение (1-67) на плоскости с координатными осями y и x_2 определяет две ветви гиперболы, осью которой является эмпирическое

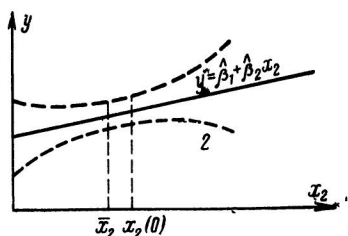


Рис. 1-2. Доверительный интервал для единичного предсказания.

1 — верхняя доверительная граница; 2 — нижняя доверительная граница.

уравнение регрессии $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_2$ (рис. 1-2). Ширина доверительного интервала минимальна, когда $x_{2(0)} = \bar{x}_2$, и увеличивается с возрастанием $|x_{2(0)} - \bar{x}_2|$.

1-3. ПРОВЕРКА ОСНОВНЫХ ГИПОТЕЗ

Пусть G — область всех возможных значений параметра θ , а g — подобласть некоторых возможных значений θ , принадлежащая области G . Проверить гипотезу относительно истинного значения параметра θ — значит на основании ограниченной выборки решить, принадлежит ли истинное значение θ подобласти g или оно лежит в подобласти $(G - g)$. Если g — единственная точка, гипотеза называется простой; в противном случае гипотеза называется сложной. Простую гипотезу можно записать $H_0: \theta = \theta^*$, против альтернативного утверждения $H_1: \theta \neq \theta^*$.

Принимая или отвергая гипотезу H_0 на основании анализа выборочных данных, исследователь рискует сделать ошибку двоякого рода:

I. Принять $H_1: \theta \neq \theta^*$, хотя в действительности $\theta = \theta^*$. Это ошибка первого рода — отрицание справедливой гипотезы. Вероятность появления такой ошибки обозначим $P(I)$.

II. Принять $H_0: \theta = \theta^*$, хотя в действительности $\theta \neq \theta^*$. Это ошибка второго рода — принятие неверной гипотезы. Вероятность ее появления обозначим $P(II)$.

Обычно вместо $P(II)$ используют величину

$$\eta = 1 - P(II), \quad (1-68)$$

которую называют мощностью критерия. Мощность критерия есть вероятность отвергнуть неверную гипотезу. Очевидно, что $0 \leq \eta \leq 1$. Можно показать, что мощность критерия есть функция от возможных значений параметра θ . Это значит, что критерий, достаточно мощный для одних значений θ , окажется слабым в области других значений θ . Существует понятие равномерно наиболее мощного критерия. Критерий называется равномерно наиболее мощным, если $\eta(\theta) > \eta'(\theta)$ для всех значений θ , где $\eta'(\theta)$ — мощность любого другого критерия. К сожалению, равномерно наиболее мощный критерий не всегда удается построить.

Для проверки основных гипотез относительно истинных значений регрессионных коэффициентов модели (1-1) мы будем пользоваться критерием отношения правдоподобия, относительно которого показано, что он является равномерно наиболее мощным и обладает рядом других оптимальных свойств. Поясним смысл этого критерия. Пусть необходимо проверить справедливость гипотезы $H_0: \theta = \theta^*$

против гипотезы $H_1: \theta \neq \theta^*$. Для этого из совокупности с плотностью распределения $\varphi(x, \theta)$ делается выборка $x_1, \dots, x_n$ и строится функция правдоподобия $L(x_1, \dots, x_n, \theta)$. Далее формируется отношение $Q = L(g)/L(G)$, где $L(g)$ — максимальное значение функции правдоподобия при условии, что истинное значение параметра θ лежит в подобласти g , т. е. в нашем случае при условии $\theta = \theta^*$, а $L(G)$ — максимальное значение функции правдоподобия при условии, что θ лежит в области G . По существу $L(G) = L(x_1, \dots, x_n, \hat{\theta})$, где $\hat{\theta}$ — выборочное значение параметра θ , полученное по методу максимального правдоподобия. Если значение $L(g)$ близко к значению $L(G)$, т. е. значения $\hat{\theta}$ и θ^* близки, то есть основания принять гипотезу H_0 и отвергнуть H_1 ; если же различие между $L(g)$ и $L(G)$ значительно, то больше оснований принять гипотезу H_1 . Очевидно, что $0 \leq Q \leq 1$, и мы отвергаем гипотезу H_0 , когда Q мало, и принимаем H_0 , когда Q велико. Теперь нужно выбрать постоянную A такой, чтобы можно было отвергать H_0 , когда $Q < A$, и принимать H_0 , когда $Q \geq A$.

Было бы желательно выбрать A так, чтобы и $P(I)$, и $P(II)$ были минимальны. Однако можно показать, что для фиксированного объема выборки уменьшение $P(I)$ влечет увеличение $P(II)$, и наоборот. Поэтому обычно задаются вероятностью ошибки первого рода $P(I) = \alpha$, что для данной выборки однозначно определяет мощность используемого критерия $\eta(\theta^*) = 1 - P(II)$. Чтобы определить величину A для выбранного значения $P(I) = \alpha$, нужно знать плотность распределения Q для случая, когда гипотеза H_0 верна. Обозначим эту плотность $h(Q, H_0)$. Тогда искомое значение A соответствует равенству

$$P(I) = \int_0^A h(Q, H_0) dQ = \alpha.$$

Для мощности критерия Q можно записать:

$$\eta(\theta^*) = \int_0^A h(Q, H_1) dQ.$$

Отметим очень важное свойство критерия отношения правдоподобия: если Q есть монотонная функция от v , то величину v можно использовать как критерий для проверки гипотезы H_0 , эквивалентный критерию Q , что облегчает пользование критерием.

Таким образом, процедура проверки гипотезы по критерию отношения правдоподобия включает следующие этапы:

1) выдвигается гипотеза, например, $H_0: \theta = \theta^*$ против альтернативного утверждения $H_1: \theta \neq \theta^*$;

2) производится случайная выборка $x_1, \dots, x_n$ из совокупности с распределением $\varphi(x, \theta)$;

3) строится функция правдоподобия $L(x_1, \dots, x_n, \theta)$ и определяются ее значения $L(g) = \max[L(x_1, \dots, x_n, \theta = \theta^*)]$ и $L(G) = L(x_1, \dots, x_n, \hat{\theta})$;

4) формируется отношение $Q = L(g)/L(G)$ и выделяется величина v , от которой Q зависит монотонно;

5) находится распределение $f(v, H_0)$ величины v при условии $\theta = \theta^*$;

6) задается вероятность появления ошибки первого рода $\alpha = P(I)$;

7) из условия $\int_{f(0)}^{f(A)} f(v, H_0) dv = \alpha$ определяется постоянная A

так, что H_0 принимается, если полученное значение $v \geq A$, и отвергается, если $v < A$;

8) вычисляется мощность критерия v для данных значений α и θ^* .

Теперь рассмотрим применение критерия отношения правдоподобия для проверки двух важнейших гипотез относительно коэффициентов регрессионной модели (1-1).

Первая гипотеза формулируется так: $H'_0: B=B^*$ против $H'_1: B \neq B^*$, где $B^* = (k \times 1)$ вектор известных величин. Это равносильно утверждению: истинное значение каждого коэффициента β_j равно соответствующей величине β_j^* . Если положить $B^* = 0$, то с помощью гипотезы $H'_0: B=0$ можно установить, существует ли связь вида (1-1) между переменной y и переменными $x_1, \dots, x_k$, включенными в уравнение регрессии. Действительно, если гипотеза $H'_0: B=0$ принимается, то между переменными y и $x_1, \dots, x_k$ с вероятностью $P=1-P(I)$ нет регрессионной связи вида (1-1); если же гипотеза $H'_0: B=0$ отвергается, это значит, что между переменной y и хотя бы одной переменной x_j из числа $x_1, \dots, x_k$ существует регрессионная зависимость, соответствующая уравнению (1-1).

Для проверки гипотезы $H'_0: B=B^*$ будем опираться на ту же выборку в виде матриц наблюдений X и Y , которая использовалась для определения оценок $\hat{B}$ и $\hat{\sigma}^2$. Поэтому выражение для функции правдоподобия будет аналогично выражению (1-6)

$$L(E, B, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{(Y - XB)^T (Y - XB)}{2\sigma^2} \right]. \quad (1-69)$$

Теперь нужно определить $L(G)$ и $L(g)$. Область G в данном случае есть $(k+1)$ -мерное пространство параметров $\beta_1, \dots, \beta_k, \sigma^2$, ограниченное условиями $-\infty < \beta_j < +\infty$; $j=1, \dots, k$; $0 < \sigma^2 < \infty$. Подобласть g есть пространство тех же параметров, ограниченное условиями гипотезы $H'_0: \beta_1 = \beta_1^*, \dots, \beta_k = \beta_k^*$ и неравенством $0 < \sigma^2 < \infty$, т. е. подобласть g представляет собой одномерное пространство, и, следовательно, проверяемая гипотеза является сложной.

Чтобы найти $L(g)$, подставим в (1-69) условие $B=B^*$ и прологарифмируем обе части выражения (1-69):

$$\begin{aligned} \ln L(E, B, \sigma^2) &= -\frac{n}{2} \ln 2\pi - \\ &- \frac{n}{2} \ln \sigma^2 - \frac{(Y - XB^*)^T (Y - XB^*)}{2\sigma^2}. \end{aligned} \quad (1-70)$$

Величина $L(g)$ представляет собой максимум выражения (1-70) по переменной σ^2 , что соответствует условию

$$\frac{\partial}{\partial \sigma^2} \ln L(E, B, \sigma^2) = \frac{(Y - XB^*)^T (Y - XB^*)}{2\sigma^4} - \frac{n}{2\sigma^2} = 0,$$

откуда получаем:

$$\sigma_g^2 = \frac{(\mathbf{Y} - \mathbf{XB}^*)^T (\mathbf{Y} - \mathbf{XB}^*)}{n}.$$

Теперь для $L(g)$ можно записать:

$$L(g) = L(\mathbf{E}, \mathbf{B} = \mathbf{B}^*, \sigma^2 = \sigma_g^2) = \frac{n^{n/2} e^{-n/2}}{(2\pi)^{n/2} [(\mathbf{Y} - \mathbf{XB}^*)^T (\mathbf{Y} - \mathbf{XB}^*)]^{n/2}}.$$

Согласно определению, сформулированному выше, величину $L(G)$ можно представить в виде

$$L(G) = L(\mathbf{E}, \mathbf{B} = \hat{\mathbf{B}}, \sigma^2 = s^2), \quad (1-71)$$

где $\hat{\mathbf{B}}$ и s^2 — соответственно оценки $\mathbf{B}$ и σ^2 , максимизирующие функцию (1-69), т. е. определяемые формулами (1-11) и (1-13). Подставив

$\hat{\mathbf{B}} = \mathbf{S}^{-1} \mathbf{X}^T \mathbf{Y}$ и $s^2 = \frac{1}{n} (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$ в (1-71), получим:

$$L(G) = \frac{n^{n/2} e^{-\frac{n}{2}}}{(2\pi)^{n/2} [(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})]}.$$

Отношение $Q = L(g)/L(G)$ имеет вид:

$$Q = \left[\frac{(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})}{(\mathbf{Y} - \mathbf{XB}^*)^T (\mathbf{Y} - \mathbf{XB}^*)} \right]^{n/2}. \quad (1-72)$$

Если $h(Q, \mathbf{B}^*)$ есть распределение Q при условии, что $\mathbf{B} = \mathbf{B}^*$, то гипотеза H'_0 : $\mathbf{B} = \mathbf{B}^*$ отвергается, если рассчитанное по выборке $\mathbf{X}$ и $\mathbf{Y}$ значение (1-72) попадает в критическую область $0 < Q \leq A$, где A таково, что

$$\int_0^A h(Q, \mathbf{B}^*) dQ = \alpha. \quad (1-73)$$

В формуле (1-73) α есть вероятность ошибки первого рода. Чтобы определить A , мы должны задаться значением α и найти распределение Q при условии $\mathbf{B} = \mathbf{B}^*$.

Определение $h(Q, \mathbf{B}^*)$ начнем с рассмотрения вспомогательного тождества

$$\begin{aligned} (\mathbf{Y} - \mathbf{XB}^*)^T (\mathbf{Y} - \mathbf{XB}^*) &= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + \\ &+ (\hat{\mathbf{B}} - \mathbf{B}^*)^T \mathbf{X}^T \mathbf{X} (\hat{\mathbf{B}} - \mathbf{B}^*). \end{aligned} \quad (1-74)$$

Если в выражении (1-74) сделать подстановку $\hat{\mathbf{B}} = \mathbf{S}^{-1} \mathbf{X}^T \mathbf{Y}$, то после несложных преобразований получим:

$$\begin{aligned} (\mathbf{Y} - \mathbf{XB}^*)^T (\mathbf{Y} - \mathbf{XB}^*) &= (\mathbf{Y} - \mathbf{XB}^*)^T (\mathbf{I}_n - \mathbf{XS}^{-1} \mathbf{X}^T) (\mathbf{Y} - \mathbf{XB}^*) + \\ &+ (\mathbf{Y} - \mathbf{XB}^*)^T \mathbf{XS}^{-1} \mathbf{X}^T (\mathbf{Y} - \mathbf{XB}^*). \end{aligned}$$

Обозначим:

$$\begin{aligned} \mathbf{Z} &= \mathbf{Y} - \mathbf{X}\mathbf{B}^*; \\ \mathbf{T}_1 &= \mathbf{I}_n - \mathbf{X}\mathbf{S}^{-1}\mathbf{X}^T; \\ \mathbf{T}_2 &= \mathbf{X}\mathbf{S}^{-1}\mathbf{X}^T. \end{aligned}$$

Учитывая результат (1-4), несложно показать, что

$$\mathbf{Z} \sim N_n(\mathbf{X}\mathbf{B} - \mathbf{X}\mathbf{B}^*, \sigma^2 \mathbf{I}_n).$$

Согласно результату (1-16) матрицы $\mathbf{T}_1$ и $\mathbf{T}_2$ идемпотентные и $\text{rg}\|\mathbf{T}_1\| = (n-k)$, а $\text{rg}\|\mathbf{T}_2\| = k$. Тогда на основании теоремы приложения и с учетом результата (1-27) можно утверждать, что

$$\frac{\mathbf{Z}^T \mathbf{T}_1 \mathbf{Z}}{\sigma^2} \sim \chi^2(n-k); \quad (1-75)$$

$$\frac{\mathbf{Z}^T \mathbf{T}_2 \mathbf{Z}}{\sigma^2} \sim \chi^2(n-k, \lambda_1), \quad (1-76)$$

где

$$\lambda_1 = \frac{(\mathbf{B} - \mathbf{B}^*)^T \mathbf{X}^T (\mathbf{X}\mathbf{S}^{-1}\mathbf{X}^T) \mathbf{X} (\mathbf{B} - \mathbf{B}^*)}{2\sigma^2} = \frac{(\mathbf{B} - \mathbf{B}^*)^T \mathbf{S} (\mathbf{B} - \mathbf{B}^*)}{2\sigma^2}. \quad (1-77)$$

Кроме того, случайные величины (1-75) и (1-76) независимы, поскольку

$$\mathbf{T}_1 \mathbf{T}_2 = (\mathbf{I}_n - \mathbf{X}\mathbf{S}^{-1}\mathbf{X}^T) (\mathbf{X}\mathbf{S}^{-1}\mathbf{X}^T) = \mathbf{0}.$$

Тогда согласно определению нецентрального F -распределения получаем:

$$v_1 = \frac{\mathbf{Z}^T \mathbf{T}_2 \mathbf{Z}}{\mathbf{Z}^T \mathbf{T}_1 \mathbf{Z}} \frac{n-k}{k} \sim F'(k, n-k, \lambda_1). \quad (1-78)$$

Путем простых матричных преобразований отношение Q можно выразить в виде

$$Q = \left(\frac{1}{1 + \frac{v_1 - k}{n-k}} \right)^{n/2},$$

т. е. как монотонную функцию от v_1 . Следовательно, v_1 можно использовать для проверки гипотезы $H'_0: \mathbf{B} = \mathbf{B}^*$ как критерий, эквивалентный критерию Q . Из результатов (1-77) и (1-78) видно, что при условии $\mathbf{B} = \mathbf{B}^*$ величина v_1 имеет центральное F -распределение с $(k, n-k)$ степенями свободы. Поэтому критическая область для v_1 , соответствующая области $0 < Q \leq A$, есть $F_\alpha(k, n-k) \leq v_1 < \infty$, поскольку $F_\alpha(k, n-k)$ удовлетворяет условию

$$\int_{F_\alpha(k, n-k)}^{\infty} F(v_1, k, n-k) dv_1 = \alpha.$$

Мощность v_1 -критерия, т. е. вероятность того, что расчетное значение v_1 попадает в критическую область $F_\alpha(k, n-k) \leq v_1 < \infty$

$< \infty$, когда $\lambda_1 \neq 0$ и, следовательно, $B \neq B^*$, определяется выражением

$$\eta(v_1) = \int_{F_\alpha(k, n-k)}^{\infty} F'(v_1, k, n-k, \lambda_2) dv_1.$$

Величина $\eta(v_1)$ есть возрастающая функция параметра нецентральности λ_1 , определяемого выражением (1-77). Таким образом, процедура проверки гипотезы $H'_0: B=B^*$ сводится к вычислению величины

$$v_1 = \frac{(Y - XB^*)^T XS^{-1}X^T (Y - XB^*)}{(Y - XB^*)^T (I_n - XS^{-1}X^T)(Y - XB^*)} \frac{n-k}{k} \quad (1-79)$$

и сравнению ее с табличным значением F -распределения для уровня значимости α и степеней свободы k и $(n-k)$. Если вычисленное значение $v_1 \geq F_\alpha(k, n-k)$, то гипотеза $H'_0: B=B^*$ отвергается; если $v_1 < F_\alpha(k, n-k)$, то гипотеза считается справедливой.

Для практически важного случая, когда $H'_0: B=0$, выражение (1-79) можно упростить, используя соотношения (1-13), (1-15) и (1-21):

$$v_1 = \frac{Y^T XS^{-1}X^T Y}{Y^T (I_n - XS^{-1}X^T) Y} \frac{n-k}{k} = \frac{\widehat{B}^T X^T X \widehat{B}}{\widehat{\sigma}^2 k} = \frac{\sum_{i=1}^n \widehat{y}_i^2}{\widehat{\sigma}^2 k}, \quad (1-80)$$

где $\widehat{y}_i$ — значение y , рассчитанное по эмпирическому уравнению регрессии для $X_{(i)}$; $i = 1, \dots, n$.

Если расчетное значение (1-80) больше взятого из таблиц значения $F_\alpha(k, n-k)$, то гипотеза $H'_0: B=0$ отбрасывается, и, следовательно, с вероятностью $P=1-P(I)=1-\alpha$ можно утверждать, что в исследуемом уравнении (1-1) переменная y регрессионно связана хотя бы с одной переменной из числа $x_1, \dots, x_k$. Если же (1-80) меньше $F_\alpha(k, n-k)$, то принимается гипотеза $H'_0: B=0$, и с вероятностью $P=1-P(II)=\eta(v_1)$ можно утверждать, что между переменной y и переменными $x_1, \dots, x_k$ не существует регрессионной зависимости.

Вторая важная гипотеза, на которой необходимо остановиться, формулируется следующим образом: только часть коэффициентов $\beta_1, \dots, \beta_r$, где $r < k$, равна предполагаемым значениям $\beta_1^*, \dots, \beta_r^*$, а относительно остальных $(k-r)$ коэффициентов не делается никаких предположений. Наиболее частое употребление этой гипотезы, встречающееся в практических задачах, относится к случаю, когда проверяется один коэффициент $H''_0: \beta_j=0$. Поочередная проверка гипотезы $H''_0: \beta_j=0$ для всех коэффициентов $\beta_1, \dots, \beta_k$ позволяет определить совокупность так называемых влиятельных или существенных независимых переменных из числа тех, которые введены в уравнение (1-1). Другими словами, гипотеза $H'_0: B=0$ позволяет проверить наличие регрессионной связи вида

(1-1) между y и хотя бы одной переменной из числа $x_1, \dots, x_k$. Если такая связь установлена, то проверяя гипотезу $H''_0: \beta_j=0$ (или $H''_1: \beta_j \neq 0$) для $j=1, \dots, k$, можно отобрать те переменные, которые участвуют в образовании этой связи.

За недостатком места не будем останавливаться на подробном выводе проверочной функции для гипотезы H''_0 , а перейдем сразу к рассмотрению критерия отношения правдоподобия для проверки гипотезы $H''_0: \beta_j=0$, который имеет вид:

$$v_2 = \frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{c_{jj}}}, \quad (1-81)$$

где $\hat{\beta}_j$ — выборочная оценка коэффициента β_j ; $\hat{\sigma}^2$ — остаточная дисперсия уравнения регрессии; c_{jj} — диагональный элемент обратной матрицы, соответствующий переменной x_j .

Величина v_2 имеет нецентральное t -распределение с $(n-k)$ степенями свободы и параметром нецентральности

$$\lambda_2 = \frac{\beta_j}{\sigma \sqrt{c_{jj}}}. \quad (1-82)$$

Критическая область для критерия v_2 будет $t_\alpha (n-k) \leq v_2 < \infty$

Если расчетное значение v_2 больше табличного значения t -распределения для уровня значимости α и $(n-k)$ степеней свободы, то гипотеза $H''_0: \beta_j=0$ отвергается и переменную x_j можно считать влияющей с надежностью $P=1-P(I)$. Если же $v_2 < t_\alpha (n-k)$, то гипотеза $H''_0: \beta_j=0$ принимается, и с вероятностью

$$\eta(v_2) = \int_{t_\alpha(n-k)}^{\infty} t'(v_2, n-k, \lambda_2) dv_2 = 1-P(II) \quad (1-83)$$

можно утверждать, что переменная x_j не участвует в регрессионной зависимости (1-1).

Используя критерий v_2 для проверки гипотезы $H''_0: \beta_j=0$, исследователь может допустить одну из двух возможных ошибок:

1) ошибка первого рода: принять $\beta_j \neq 0$, хотя в действительности $\beta_j=0$;

2) ошибка второго рода: принять $\beta_j=0$, хотя в действительности $\beta_j \neq 0$.

С точки зрения задачи отбора совокупности влияющих переменных появление второй ошибки более нежелательно, так как в этом случае из уравнения регрессии будет исключена влияющая переменная. Вероятность отвергнуть гипотезу $H''_0: \beta_j=0$, когда в действительности $\beta_j \neq 0$, равна мощности критерия (1-83), которая есть возрастающая функция параметра нецентральности λ_2 . Последнее означает, что для данного уравнения регрессии с конечной дисперсией ошибки σ^2 критерий v_2 будет относить проверяемый коэффициент β_j к ненулевому, когда и в действительности $\beta_j \neq 0$, с вероятностью, пропорциональной истинному значению β_j и обратно пропорциональной величине диагонального элемента обратной мат-

рицы c_{jj} . Рассмотрим подробнее диагональный элемент c_{jj} матрицы $\mathbf{C} = \mathbf{S}^{-1}$. Разобьем матрицу $\mathbf{S}$ на подматрицы:

$$\mathbf{S} = \begin{vmatrix} \mathbf{X}_j^T \mathbf{X}_j & \mathbf{X}_j^T \mathbf{X}_{(k-1)} \\ \mathbf{X}_{(k-1)}^T \mathbf{X}_j & \mathbf{X}_{(k-1)}^T \mathbf{X}_{(k-1)} \end{vmatrix}, \quad (1-84)$$

где $\mathbf{X}_j$ — $(n \times 1)$ вектор-столбец матрицы $\mathbf{X}$, соответствующий n значениям переменной x_j ; $\mathbf{X}_{(k-1)}$ — матрица $\mathbf{X}$ без столбца $\mathbf{X}_j$.

Аналогично разобьем матрицу $\mathbf{C}$:

$$\mathbf{C} = \begin{vmatrix} c_{jj} & \mathbf{C}_{j(k-1)} \\ \mathbf{C}_{(k-1)j} & \mathbf{C}_{(k-1)(k-1)} \end{vmatrix}, \quad (1-85)$$

С учетом равенства $\mathbf{C} = \mathbf{S}^{-1}$ алгебраическим преобразованием выражений (1-84) и (1-85) можно получить следующее соотношение для c_{jj} :

$$c_{jj} = \frac{1}{\mathbf{X}_j^T \mathbf{X}_j - \mathbf{X}_j^T \mathbf{X}_{(k-1)} (\mathbf{X}_{(k-1)}^T \mathbf{X}_{(k-1)})^{-1} \mathbf{X}_{(k-1)}^T \mathbf{X}_j} = \frac{1}{V}. \quad (1-86)$$

В свою очередь знаменатель выражения (1-86) можно привести к виду

$$V = \hat{\sigma}_j^2 (n-1) (1 - \hat{R}_{j \cdot 1, 2, \dots, (k-1)}^2), \quad (1-87)$$

где $\hat{\sigma}_j^2$ — выборочная дисперсия переменной x_j , подсчитанная по данным $\mathbf{X}_j$; n — число опытов в матрице $\mathbf{X}$; $\hat{R}_{j \cdot 1, 2, \dots, (k-1)}$ — выборочный коэффициент множественной корреляции между переменной x_j и остальными независимыми переменными $x_1, \dots, x_{(k-1)}$.

Подставив (1-87) в выражение (1-82), получим:

$$\lambda_2 = \frac{\beta_j}{\sigma} \hat{\sigma}_j \sqrt{(n-1) (1 - \hat{R}_{j \cdot 1, 2, \dots, (k-1)}^2)}. \quad (1-88)$$

Результат (1-88) позволяет сделать следующие выводы:

1. Для фиксированных конечных значений β_j , σ^2 и выборки n мощность критерия v_2 будет тем выше, чем больше диапазон изменения значений переменной x_j в матрице наблюдений $\mathbf{X}$. Это значит, что при организации эксперимента по сбору статистических данных желательно «раскачивать» независимые переменные, добиваясь максимального размаха.

2. Мощность критерия v_2 возрастает с уменьшением коррелированности столбцов матрицы $\mathbf{X}$. Когда значения переменной x_j представляют собой линейные комбинации значений других переменных, т. е.

$\hat{R}_{j \cdot 1, 2, \dots, (k-1)} = 1$, мощность критерия v_2 минимальна ($\lambda_2 = 0$), и он не почувствует сколь угодно большого значения коэффициента β_j из интервала $-\infty < \beta_j < \infty$. Если же значения переменной x_j не коррелированы со значениями других переменных, т. е. $\hat{R}_{j \cdot 1, 2, \dots, (k-1)} = 0$,

что соответствует случаю ортогональной матрицы X , мощность критерия v_2 будет максимальной. При $0 < R_{j,1,2,\dots,(k-1)} < 1$ вероятность принять $\beta_j = 0$, когда в действительности $\beta_j \neq 0$, зависит от степени коррелированности столбцов матрицы X .

Второй вывод является объяснением парадоксальной ситуации, нередко встречающейся на практике, когда в результате проверки гипотеза $H'_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ отвергается, но в то же время гипотеза $H''_0: \beta_j = 0$ принимается как справедливая для всех коэффициентов $\beta_1, \dots, \beta_k$. Пример такой ситуации приводится в [Л 9]. В этом случае коррелированность столбцов матрицы X настолько велика, что соотношение между β_j и c_{jj} для всех $j = 1, \dots, k$ не обеспечивает достаточной мощности критерия v_2 для надежного обнаружения отличных от нуля коэффициентов, а критерий v_1 не столь чувствителен к особенностям матрицы X , и, как следует из (1-77), распределение величины v_1 превращается в центральное F распределение, если $\beta_1 = \beta_2 = \dots = \beta_k = 0$, т. е. когда справедлива гипотеза $H'_0: \beta = 0$.

Пример 1-5. Действительная модель имеет вид:

$$y = 0,7 x_1 + 0,6 x_2 - 0,4 x_3 + e, \quad (1-89)$$

где ошибка e представляет собой нормально распределенную случайную величину с нулевым математическим ожиданием и единичной дисперсией. Не зная структуры действительной модели, исследователь сформулировал предполагаемую модель в виде

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + e. \quad (1-90)$$

По $n = 50$ значениям переменных y, x_1, x_2, x_3, x_4 были получены оценки коэффициентов модели (1-90)

$$\hat{\beta}_1 = 0,73, \hat{\beta}_2 = 0,68, \hat{\beta}_3 = -0,44, \hat{\beta}_4 = -0,008.$$

Остаточная дисперсия равна $\hat{\sigma}^2 = 1,12$. Для отбора существенных переменных из числа x_1, x_2, x_3, x_4 последовательно проверялась гипотеза $H''_0: \beta_j = 0$ для $j = 1, 2, 3, 4$. Для проверки использовался критерий v_2 , расчетное значение которого сравнивалось с табличным значением t -распределения для $\alpha = 0,05$ и $(n - k) = 50 - 4 = 46$, $t_{0,05; 46} = 1,96$. Последовательно для $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ и $\hat{\beta}_4$ были получены следующие значения v_2 : $v_2 \hat{\beta}_1 = 4,6$; $v_2 \hat{\beta}_2 = 5,4$; $v_2 \hat{\beta}_3 = 2,2$; $v_2 \hat{\beta}_4 = 0,034 < 1,96$.

Как видно, в данном случае критерий v_2 оказался достаточно мощным, чтобы правильно отобрать совокупность влиятельных переменных, соответствующих действительной модели (1-89). Однако такой благополучный результат получается далеко не всегда. Стремясь охватить больший класс возможных моделей, исследователь мог бы сформировать предполагаемую модель не в виде (1-90), а в виде полного квадратичного полинома от четырех переменных:

$$y = \sum_{j=1}^4 \beta_j x_j + \sum_{j=1}^4 \beta_{jj} x_j^2 + \sum_{\substack{j,q=1 \\ (j < q)}}^4 \beta_{jq} x_j x_q + e, \quad (1-91)$$

используя для получения оценок коэффициентов (1-91) те же $n = 50$ значений переменных y, x_1, x_2, x_3, x_4 .

Если бы критерий v_2 работал безошибочно, то, последовательно применяя его для проверки гипотезы $H''_0: \beta_j=0$ для $j=1, \dots, 14$, исследователь пришел бы от модели (1-91) к модели (1-89). Посмотрим, что получается на самом деле.

Найденные оценки коэффициентов (1-91):

$$\begin{aligned}\hat{\beta}_1 &= 0,184; & \hat{\beta}_2 &= 0,16; & \hat{\beta}_3 &= -0,58; \\ \hat{\beta}_4 &= 0,011; & \hat{\beta}_{11} &= 0,5; & \hat{\beta}_{22} &= 0,25; \\ \hat{\beta}_{33} &= 0,11; & \hat{\beta}_{44} &= 0,82; & \hat{\beta}_{12} &= 0,33; \\ \hat{\beta}_{13} &= -0,31; & \hat{\beta}_{14} &= 0,67; & \hat{\beta}_{23} &= -0,27; \\ \hat{\beta}_{24} &= -0,09; & \hat{\beta}_{34} &= -0,73.\end{aligned}$$

Остаточная дисперсия для (1-91) равна $\hat{\sigma}^2=1,03$. Последовательное применение v_2 -критерия выявило такой порядок отбрасывания невлиятельных переменных:

- 1) $x_4 - v_2 \hat{\beta}_4 = 0,25$; 2) $x_2 - v_2 \hat{\beta}_2 = 0,21$; 3) $x_3^2 - v_2 \hat{\beta}_{33} = 0,26$;
4) $x_2 - v_2 \hat{\beta}_2 = 0,5$; 5) $x_1 - v_2 \hat{\beta}_1 = 0,55$; 6) $x_1 - v_2 \hat{\beta}_1 = 0,97$;
7) $x_1^2 - v_2 \hat{\beta}_{11} = 1,08$; 8) $x_2^2 - v_2 \hat{\beta}_{22} = 1,8$.

После отбрасывания каждой переменной все оценки коэффициентов и значения v_2 -критерия для оставшихся переменных пересчитывались заново. Окончательный вид эмпирического уравнения регрессии для модели (1-91) имел вид:

$$\begin{aligned}y &= 0,46 - 0,75 x_3 + 1,0 x_4^2 - 0,9 x_1 x_3 + \\ &+ 0,64 x_1 x_4 - 0,6 x_2 x_3 - 0,9 x_3 x_4.\end{aligned}\quad (1-92)$$

Как видно, по структуре уравнение (1-92) не имеет ничего общего с действительной зависимостью (1-89). Полученный результат хорошо иллюстрирует отрицательное влияние коррелированности переменных на работоспособность v_2 -критерия.

Вспомним, что мощность критерия есть вероятность отвергнуть неверную гипотезу. В модели (1-91) для коэффициентов типа β_{jq} , $j, q=1, 2, 3, 4$ неверной будет гипотеза $H: \beta_{jq} \neq 0$, которая отвергается в пяти случаях из десяти; а для коэффициентов типа β_j , $j=1, 2, 3$ неверной будет гипотеза $H: \beta_j = 0$, которая отвергается в единственном случае из трех. Из приведенной корреляционной матрицы независимых переменных модели (1-91) видно, что при сравнительно слабой корреляции между переменными x_1, x_2, x_3, x_4 переменные типа x_j^2 и $x_j x_q$ образуют очень сильно коррелированные пары, что резко снижает мощность v_2 -критерия и не позволяет получить удовлетворительный результат.

Картина существенно улучшается, если не вводить в модель квадраты переменных. В этом случае из корреляционной матрицы выпадает блок, обведенный пунктиром. Для модели вида

$$y = \sum_{j=1}^n \beta_j x_j + \sum_{\substack{j,q=1 \\ (j < q)}}^4 \beta_{jq} x_j x_q + c \quad (1-93)$$

[illegible]

после отбрасывания невлиятельных переменных было получено эмпирическое уравнение регрессии

$$\hat{y} = 0,34 + 0,65x_1 + 0,57x_2 - 0,51x_3 + 0,23x_1x_2, \quad (1-94)$$

которое весьма удовлетворительно отражает действительную зависимость (1-89). Остаточная дисперсия уравнения (1-94) равна $\hat{\sigma}^2 = 1,09$.

Корреляционная матрица переменных полного квадратичного полинома приведена на стр. 38.

ГЛАВА ВТОРАЯ

ПРИМЕНЕНИЕ МНОЖЕСТВЕННОЙ РЕГРЕССИИ ДЛЯ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКОГО ОПИСАНИЯ ПРОМЫШЛЕННОГО ОБЪЕКТА

2-1. ВВОДНЫЕ ЗАМЕЧАНИЯ

В этой главе рассматриваются особенности применения аппарата множественной регрессии в задачах получения статистического математического описания промышленного объекта.

Будем полагать, что в объекте реализуется непрерывный технологический процесс, который условно можно представить в виде структурной схемы, изображенной на рис. 1-1. Объединим все контролируемые параметры группы *A* и *B* в общую группу $X = (x_1, \dots, x_l)$, где $l = m + r$, и будем называть их входными или технологическими переменными; соответственно параметр *y* будем называть выходной переменной.

Для переменных, определяющих текущее состояние объекта, введем следующие допущения:

1) изменения каждого контролируемого параметра $x_1, \dots, x_l$ представляют собой нормально распределенный случайный стационарный процесс, обладающий свойством эргодичности;

2) каждый параметр группы *D* характеризует случайно и независимо действующий фактор, среди которых нет доминирующих;

3) все параметры группы *D* не коррелированы с контролируемыми параметрами, и их изменения представляют случайный стационарный процесс с нулевым математическим ожиданием.

Введенные допущения достаточно реальны и выполняются для большинства промышленных объектов.

Статическое математическое описание определяет связь между соответствующими мгновенными значениями переменной *y* и переменных $x_1, \dots, x_l$. В общем виде статическое математическое описание сложного объекта представляет собой совокупность уравнений типа

$$y_j(t) = \psi_j[x_1(t), \dots, x_l(t), d_1(t), \dots, d_s(t)], \quad (2-1)$$

где $j = 1, \dots, q$; *q* — число выходных переменных объекта.

Поскольку согласно допущениям каждый параметр $d_1, \dots, d_s$ характеризует случайно и независимо действующий фактор, не коррелированный с входными переменными, то каждое уравнение совокупности (2-1) можно привести к виду

$$y(t) = \psi[x_1(t), \dots, x_l(t)] + \xi[d_1(t), \dots, d_s(t)] = \psi[x_1(t), \dots, x_l(t)] + e(t), \quad (2-2)$$

где $e(t)$ — эквивалентный шум, приведенный к выходу объекта и заменяющий действие всех неконтролируемых параметров.

В соответствии с принятыми допущениями структурную схему объекта можно привести к виду на рис. 2-1. Шум e нарушает однозначность связи между переменными y и $x_1, \dots, x_l$, и потому ее можно рассматривать лишь в вероятностном смысле.



Рис. 2-1. Приведенная структурная схема сложного объекта.

Поскольку эквивалентный шум порождается множеством случайно и независимо действующих факторов $d_1, \dots, d_s$, среди которых нет доминирующих, то на основании центральной предельной теоремы [Л. 14] можно считать, что

$e(t)$ является случайным процессом с нулевым математическим ожиданием и нормальным законом распределения. Тогда для каждого фиксированного момента времени математическое ожидание выходной переменной y

$$M[y] = \psi(x_1, \dots, x_l).$$

Для практических целей зависимость $\psi(x_1, \dots, x_l)$ удобно выразить в виде полинома

$$\psi(x_1, \dots, x_l) = \sum_j \beta_j x_j + \sum_{j,q} \beta_{jq} x_j x_q + \dots \quad (2-3)$$

Используя формально линеаризованные переменные, статическое математическое описание объекта для каждого фиксированного момента времени можно представить в виде математической модели

$$y(t_i) = \beta_1 x_1(t_i) + \dots + \beta_k x_k(t_i) + e(t_i), \quad (2-4)$$

где $e(t_i)$ — нормально распределенная случайная ошибка с нулевым математическим ожиданием; k — число членов полинома (2-3).

В дальнейшем будем опускать индекс t_i , кроме специально оговоренных случаев, имея в виду, что модель (2-4) рассматривается для фиксированного момента времени.

Задача построения статического математического описания сложного объекта сводится к экспериментальному определению оценок модели (2-4), для чего может быть использован аппарат множественной регрессии.

Статистическим материалом для получения выборочных оценок $\hat{\beta}_1, \dots, \hat{\beta}_k$ могут служить экспериментальные данные, полученные на объекте в режиме его нормальной эксплуатации. Эти данные представляют собой наборы значений контролируемых переменных

y и $x_1, \dots, x_l$, зафиксированные в дискретные моменты времени. Как и прежде, набор значений $\{y_i, x_{1i}, \dots, x_{li}\}$, зафиксированных в момент времени t_i и соответствующих i -му текущему состоянию объекта, будем называть опытом.

Все n опытов, проведенных в процессе эксперимента, будем представлять в виде матриц наблюдений Y и X ; первая — размером $(n \times 1)$, вторая — размером $(n \times k)$, где k — число членов в модели (2-4) с учетом линеаризованных переменных полинома (2-3). В общем случае полином (2-3) может быть сколь угодно высокой степени, однако, как показывает практика, для построения математического описания промышленного объекта достаточно ограничиться вторым — третьим порядком.

Выбор порядка и формирование вида полинома (2-3) являются весьма ответственными неформальными этапами в задаче построения математического описания промышленного объекта.

Поскольку обычно вид зависимости $\psi(x_1, \dots, x_l)$ неизвестен, можно пойти по формальному пути: представить $\psi(x_1, \dots, x_l)$ в виде полинома достаточно высокого порядка, найти оценки всех его коэффициентов, а затем, отбрасывая с помощью критерия v_2 невлиятельные члены, остановиться на полиноме, который с приемлемой для практических целей точностью аппроксимирует неизвестную зависимость $\psi(x_1, \dots, x_l)$ (вопрос оценки работоспособности математического описания рассматривается в § 2-5). Однако такой путь имеет ряд серьезных недостатков. Во-первых, чем больше членов в модели (2-4), тем больше опытов необходимо иметь в матри-

цах наблюдений, чтобы получить оценки $\hat{\beta}_1, \dots, \hat{\beta}_k$ с удовлетворительной точностью, и, следовательно, тем продолжительнее и дороже будет эксперимент по сбору статистических данных. Кроме того, при большом k резко возрастают объем и трудоемкость вычислений, и не всякой ЭВМ они могут оказаться под силу. Во-вторых, как показано в примере 1-5, с увеличением порядка полинома увеличивается корреляция между переменными, участвующими в уравнении (2-4), что в свою очередь снижает мощность v_2 -критерия, и полученный путем отбрасывания невлиятельных членов полином по составу переменных может существенно отличаться от действительной полиномиальной аппроксимации, соответствующей бесконечно большой выборке, а это приведет к неверным выводам при технологической интерпретации математического описания.

Более правильным представляется путь использования априорной информации о процессе, содержащейся в практическом опыте технологов и специалистов, знакомых с исследуемым процессом. В этом случае в модель (2-4) вводятся не все члены полинома выбранной степени, а только те из них, которые были отобраны как существенные в результате обсуждения со специалистами-технологами. При этом могут быть использованы методы объективной оценки субъективных знаний, например методы ранговой корреляции, парных сравнений и т. д. Найденные для модели (2-4) оценки коэффициентов дополнительно проверяются с помощью критерия v_2 .

Отмеченное выше несовершенство v_2 -критерия заставляет исследователей искать более надежные способы отбора совокупности влиятельных переменных. В частности, представляют интерес появившиеся публикации, в которых предлагаются различные методы упорядоченного перебора переменных с использованием критерия информативности. Наиболее законченные результаты содержатся в [Л. 10—13].

Как уже отмечалось во введении, специфические особенности производственных условий требуют разработки и соблюдения специальных мер, обеспечивающих корректность применения аппарата множественной регрессии для построения математического описания промышленного объекта.

В этой главе будут рассмотрены следующие вопросы:

- 1) выбор интервала между соседними опытами и оценка продолжительности эксперимента по сбору статистических данных;
- 2) учет погрешностей регистрации при сборе статистических данных;
- 3) оценка работоспособности математического описания;
- 4) некоторые вычислительные рекомендации.

2-2. ПРОДОЛЖИТЕЛЬНОСТЬ ЭКСПЕРИМЕНТА И ИНТЕРВАЛ СЪЕМА ДАННЫХ

При рассмотрении вопроса о требуемом объеме выборки в некоторых руководствах [Л. 15] приводятся рекомендации, опирающиеся на следующий критерий: объем выборки n будет достаточным, если с заданной вероятностью P

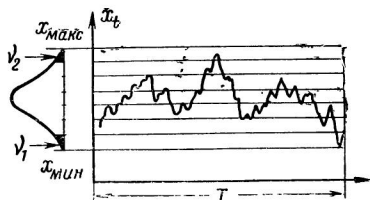


Рис. 2-2. К расчету продолжительности эксперимента T .

будет отличаться от истинного значения не более чем на заданную величину ошибки δ . Однако такой подход дает сильно завышенную оценку для n , что невыгодно, так как проведение эксперимента на промышленном объекте связано с большими материальными затратами. Кроме того, одного критерия статистической надежности недостаточно; если речь идет о сборе данных на промышленном объекте

с непрерывным изменением технологических переменных, так как остается неясным, за какой промежуток времени следует провести требуемые n опытов. Известно, что регрессионное уравнение может эффективно использоваться лишь в том интервале, в котором наблюдались изменения переменных при сборе статистических данных. Всякая экстраполяция уравнения за пределы обследованного интервала является несостоятельной и может привести к большим погрешностям [Л. 9, 16]. Поэтому, если проводить n опытов за короткий отрезок времени, в течение которого технологические переменные успеют «пробежать» лишь часть своего диапазона, математическое описание объекта может оказаться непригодным для рабочего диапазона в целом. Отсюда появляется необходимость учитывать характер изменения во времени всех технологических переменных объекта.

Сформулируем задачу следующим образом. Задан рабочий диапазон конечной величины для технологической переменной x , изменения которой представляют собой случайный стационарный процесс (рис. 2-2):

$$\Delta x = x_{\max} - x_{\min}. \quad (2-5)$$

Весь диапазон Δx разбит на ряд одинаковых квантов в соответствии с разрешающей способностью измерительного прибора.

Предположим, что известны дискретность проведения опытов Δt и вероятности ν_1 и ν_2 попадания случайной величины x в нижний и верхний кванты диапазона Δx .

Требуется определить время наблюдения T над переменной x , в течение которого в серии измерений $n=T/\Delta t$ переменная x с вероятностью P будет зафиксирована в верхнем и нижнем квантах хотя бы по одному разу. Естественно ожидать, что при выполнении этого требования будет охвачен весь диапазон изменения переменной, так как вероятности пребывания вблизи середины диапазона больше, чем вблизи его границ.

Обычно технологические переменные хорошо налаженных производств имеют небольшую колебательность, и, следовательно, вероятности попадания в крайние кванты диапазона ν_1 и ν_2 будут малы. В этом случае можно воспользоваться формулой Пуассона, согласно которой вероятность того, что переменная x появится в крайнем кванте ровно m раз, определяется выражением

$$P_m = \frac{\lambda^m}{m!} e^{-\lambda}. \quad (2-6)$$

Параметр λ можно представить в виде

$$\lambda = aT, \quad (2-7)$$

где a — среднее число попаданий в крайний квант в единицу времени; T — полное время наблюдений.

Следовательно, λ характеризует среднее число попаданий переменной в крайний квант диапазона за время эксперимента.

Если переменная x имеет симметричное распределение внутри диапазона, и, следовательно, $\nu_1 = \nu_2 = \nu$, то вероятность того, что за n наблюдений значения переменной x будут обнаружены хотя бы по одному разу и в верхнем, и в нижнем квантах, равна [Л. 2]:

$$P = (1 - e^{-\lambda})^2. \quad (2-8)$$

Задавшись подходящим значением P , из выражения (2-8) можно найти λ , а затем по формуле (2-7) определить необходимое время наблюдений T . В табл. 2-1 приведены значения λ для распространенных в инженерной практике значений P .

Т а б л и ц а 2-1

P	0,94	0,95	0,96	0,97	0,98	0,99
λ	3,52	3,68	3,90	4,19	4,6	5,3

Среднее число попаданий переменной в крайний квант диапазона в единицу времени

$$a = \nu / \Delta t. \quad (2-9)$$

Следовательно,

$$\lambda = \frac{T}{\Delta t} \nu = n\nu, \quad (2-10)$$

откуда

$$T = \frac{\Delta t \lambda}{\nu}. \quad (2-11)$$

Вычислив отношение $\Delta t/\nu$ и взяв из таблицы значение λ , соответствующее выбранной вероятности P , по формуле (2-11) можно найти требуемое время наблюдений T . Оно измеряется в тех же

единицах, что и Δt . На рис. 2-3 представлены зависимости T от $\Delta t/v$ при двух наиболее употребительных значениях P .

Для практической оценки T вероятности v_1 и v_2 можно заменить соответствующими частотами $\hat{v}_1$ и $\hat{v}_2$. Материалом для определения $\hat{v}_1$ и $\hat{v}_2$ могут служить многочисленные архивные данные, имеющиеся на каждой производственной установке.

При окончательном выборе продолжительности эксперимента $T_{\text{экс}}$ необходимо принимать во внимание все значения T_j , рассчитанные для технологических переменных x_j .

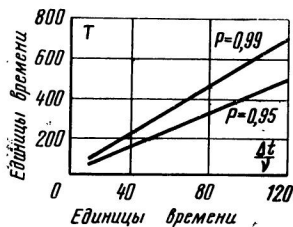


Рис. 2-3. Графики для расчета T .

носивлю требованию $R_y(0) \approx 0$, где $R_y(0)$ — автокорреляционная функция выходной переменной y .

Практически интервал Δt выбирается по условию

$$\Delta t \geq \theta_y^* \quad (2-12)$$

Здесь θ_y^* — время корреляции переменной y , определяемое как $R_y(\theta^*) = 0,05R_y(0)$.

Время корреляции θ_y^* легко определить по эмпирической кривой автокорреляционной функции $\hat{R}_y(\theta)$, для построения которой должен быть организован специальный эксперимент с достаточно частой (порядка 10—20 мин) фиксацией значений y [Л. 17]. Приблизительно значение θ_y^* можно оценить по диаграммной записи y . Для этого на диаграмме отмечают приближенное среднее значение и подсчитывают число пересечений F за время t . Практически установлено [Л. 17], что среднее число пересечений $f_0 = F/t$ удовлетворительно стабилизируется за время, в течение которого $F = 40 \div 70$. Тогда $\theta_y^* \approx 2/f_0$. Для непрерывных технологических производств величина θ_y^* колеблется в пределах от единиц до десятков часов.

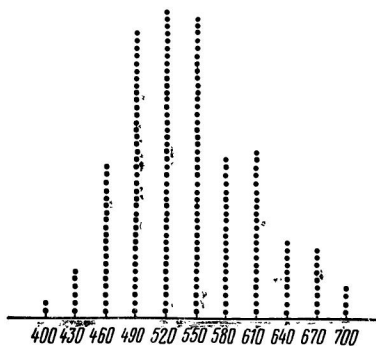


Рис. 2-4. Гистограмма распределения значений технологической переменной.

При выборе Δt необходимо также учитывать ограничения, вытекающие из разрешающей способности контрольно-измерительной аппаратуры. Например, в большинстве случаев состав продуктов контролируется с помощью лабораторных химических анализов, продолжительность которых нередко составляет несколько часов. Экспериментатор не может приступить к отбору новых проб, пока не будут готовы результаты предыдущих замеров.

При выборе верхнего предела Δt следует принимать во внимание такие факторы, как возможное изменение параметров объекта во времени, нарушение условий, в которых проводится эксперимент, и максимальные затраты, связанные с его проведением. Если интервал Δt взять очень большим, то результаты эксперимента окажутся подверженными влиянию изменений технологических параметров объекта. Поэтому интервал Δt целесообразно выбирать по возможности ближе к нижнему пределу.

Пример 2-1. При построении математического описания процесса коагуляции латекса после анализа диаграммных записей выходной переменной было решено выбрать $\Delta t = 3$ ч. Гистограмма распределения выходной переменной, построенная для 200 данных, взятых из операторских записей, изображена на рис. 2-4.

Из гистограммы следует, что частоты $\hat{v}_1$ и $\hat{v}_2$ можно взять равными $\hat{v}_1 = \hat{v}_2 = \hat{v} \sim 3/200$. Тогда отношение $\Delta t/\nu$ равно $\Delta t/\nu = \frac{3}{3/200} = 200$.

Согласно табл. 2-1 для вероятности $P = 0,95$ имеем $\lambda = 3,68$. Подставив найденные значения в формулу (2-11), получим оценку продолжительности эксперимента $T = 200 \cdot 3,68 = 736$ ч, т. е. при круглосуточной работе эксперимент по сбору статистических данных должен продолжаться месяц. За это время будет собрано $n = T/\Delta t \approx 245$ опытов.

2-3. ВЛИЯНИЕ ПОГРЕШНОСТЕЙ РЕГИСТРАЦИИ СТАТИСТИЧЕСКИХ ДАННЫХ НА ТОЧНОСТЬ МАТЕМАТИЧЕСКОГО ОПИСАНИЯ

Аппарат множественной регрессии построен в предположении, что все данные, содержащиеся в матрице X , соответствуют истинным значениям переменных в момент наблюдения. Однако в реальных условиях промышленного объекта технологические переменные регистрируются с некоторой погрешностью, обусловленной либо ошибками измерительных приборов, либо неточностью химических анализов. Погрешность измерения переменных увеличивает неопределенность уравнения регрессии и в этом смысле эквивалентна действию добавочного внешнего шума. Практически это проявляется в возрастании остаточной дисперсии уравнения по сравнению с остаточной дисперсией, вычисленной при точно измеренных данных. В общем случае приращение остаточной дисперсии за счет погрешностей измерения может быть столь значительным, что полученное уравнение регрессии теряет практический смысл. Поэтому важно оценить влияние ошибок регистрации статистических данных на точность получаемого уравнения регрессии.

Представим i -е измерение независимой переменной x_j в виде суммы

$$x_{ji} = x_{ji}^0 + \varepsilon_{ji},$$

где x_{ji}^0 — истинное значение переменной x_j ; ε_{ji} — значение ошибки измерения.

Аналогично можно представить ошибку измерения зависимой переменной y .

Будем предполагать, что ошибки измерения по каждой переменной представляют собой случайные величины с нулевым математическим ожиданием и дисперсией ζ_j^2 , т. е.

$$M[\varepsilon_j] = 0; M[\varepsilon_j \varepsilon_j] = \zeta_j^2. \quad (2-13)$$

Далее предположим, что ошибки измерения по каждой переменной не коррелированы с ошибками по другим переменным и со значениями самих переменных, т. е.

$$\begin{aligned} M[\varepsilon_j \varepsilon_q] &= 0; M[\varepsilon_j (x_j - M[x_j])] = 0; \\ M[\varepsilon_j (x_q - M[x_q])] &= 0. \end{aligned} \quad (2-14)$$

Введенные предположения достаточно реальны и выполняются в большинстве практических случаев. Задача состоит в том, чтобы, располагая матрицей наблюдений X , содержащей значения переменных $x_1, \dots, x_k$, измеренных с ошибками, найти оценки коэффициентов $\hat{\beta}_1, \dots, \hat{\beta}_k$, освобожденные от искажающего влияния ошибок измерения, и определить величину приращения остаточной дисперсии, обусловленную ошибками регистрации отдельных переменных. Для облегчения выводов и анализа будем пользоваться соотношениями для генеральных значений параметров. В конце параграфа основные результаты будут выражены через выборочные оценки.

Согласно формуле (1-45) дисперсия ошибки σ^2 связана с генеральным значением $R_{y \cdot x_1, \dots, x_k}$ соотношением

$$\sigma^2 = \sigma_y^2 (1 - R_{y \cdot x_1, \dots, x_k}^2). \quad (2-15)$$

Дисперсия выходной переменной σ_y^2 не зависит от ошибок измерения $x_1, \dots, x_k$, и поэтому влияние ошибок сказывается лишь на величине $R_{y \cdot x_1, \dots, x_k}$. Согласно формуле (1-46) для генерального значения $R_{y \cdot x_1, \dots, x_k}$ можно записать:

$$R_{y \cdot x_1, \dots, x_k}^2 = 1 - \frac{D}{\sigma_y^2 D_{yy}}. \quad (2-16)$$

Элементы определителя D представляют собой ковариации переменных

$$d_{jq} = M[(x_j - M[x_j])(x_q - M[x_q])]. \quad (2-17)$$

В (2-17) заменим x на $(x^0 + \varepsilon)$. Тогда с учетом предположений (2-13) и (2-14) после несложных преобразований (2-17) получим:

$$\left. \begin{aligned} d_{jq} &= d_{jq}^0 = M[(x_j^0 - M[x_j^0])(x_q^0 - M[x_q^0])]; \\ d_{jj} &= d_{jj}^0 + \zeta_j^2 = \sigma_j^{02} + \zeta_j^2, \end{aligned} \right\} \quad (2-18)$$

где d_{jq}^0 — ковариация переменных x_j и x_q , измеренных без ошибок; $d_{jj}^0 = (\sigma_j^0)^2$ — дисперсия переменной x_j , измеряемой без ошибок.

Чтобы выяснить, как зависит $R_{y, x_1, \dots, x_k}$ от погрешностей измерения переменных $x_1, \dots, x_k$, рассмотрим частные производные $\partial R^2 / \partial \zeta_j^2$ ($j = 1, \dots, k$) применительно к формуле (2-16). Используя теорему о дифференцировании определителей, получаем [Л.18]:

$$\frac{\partial R^2}{\partial \zeta_j^2} = - \frac{(b_j)^2}{\sigma_y^2} = - \frac{(b^*_j)^2}{\sigma_j^2}. \quad (2-19)$$

Все величины, стоящие в правой части равенства (2-19), положительны; следовательно, погрешность измерения каждой независимой переменной x_j всегда вызывает уменьшение значения множественного коэффициента корреляции $R_{y, x_1, \dots, x_k}$. Это уменьшение пропорционально квадрату величины регрессионного коэффициента в стандартизованном виде b^*_j и обратно пропорционально дисперсии переменной x_j .

Для оценки количественного изменения величины $R_{y, x_1, \dots, x_k}$ за счет ошибок измерения заменим в (2-19) дифференциалы на соответствующие приращения

$$\left. \begin{aligned} dR^2 &\approx \Delta R^2 = R^2 - (R^0)^2; \\ d\zeta_j^2 &\approx \Delta \zeta_j^2 = \zeta_j^2, \end{aligned} \right\} \quad (2-20)$$

где R^0 — значение множественного коэффициента корреляции в условиях отсутствия ошибок измерения, т. е. когда все $\zeta_j^2 = 0$. Тогда согласно (2-19) ошибки измерения каждой переменной вызывают уменьшение $(R^0)^2$ на величину

$$\Delta R_{(j)}^2 \approx \frac{\zeta_j^2}{\sigma_j^2} (b^*_j)^2. \quad (2-21)$$

Выражение (2-21) показывает, что вклад погрешности измерения каждой переменной x_j в приращение остаточной дисперсии определяется не только величиной собственно ошибки измерения ζ_j^2 , но также диапазоном изменения этой переменной σ_j^2 и степенью влияния ее на зависимую переменную b^*_j . Отсюда вытекает одно практически важное замечание: чем больше размах колебаний независимых переменных при сборе статистических данных, тем меньше скажется погрешность регистрации на точности математического опи-

сания. Соотношение (2-21) позволяет проранжировать все технологические переменные соответственно их роли в снижении точности математического описания объекта из-за погрешности измерения. Такая ранжировка позволит выявить особо «неприятные» переменные, которые в первую очередь требуют повышения точности регистрации при сборе статистических данных. В формуле (2-21), как и во всех последующих результатах, значения регрессионных коэффициентов b^*_j и дисперсий переменных σ^2_j вычисляются по фактически наблюдавшимся данным, содержащим ошибки измерения.

Используя описанный метод, нетрудно показать, что приращение дисперсии за счет ошибки измерения зависимой переменной y равно:

$$\Delta R^2_{(y)} \approx \frac{\zeta_y^2}{\sigma_y^2}.$$

Воспользовавшись теоремой математического анализа о полном приращении функции многих переменных, запишем:

$$\Delta R^2 = \sum_{j=1}^k \Delta R^2_{(j)} \approx \sum_{j=1}^k \frac{\zeta_j^2}{\sigma_j^2} (b^*_j)^2, \quad (2-22)$$

где ΔR^2 — полное отрицательное приращение $(R^0_{y \cdot x_1, \dots, x_k})^2$ за счет ошибок измерения по всем независимым переменным $x_1, \dots, x_k$.

С учетом знаков в равенствах (2-19) и (2-20) выражение (2-15) можно представить в виде

$$\begin{aligned} \sigma^2 &= \sigma_y^2 \{1 - [(R^0_{y \cdot x_1, \dots, x_k})^2 - \Delta R^2]\} = \\ &= \sigma_y^2 [1 - (R^0)^2] + \sigma_y^2 \Delta R^2 = (\sigma^0)^2 + \Delta \sigma^2, \end{aligned} \quad (2-23)$$

где $\Delta \sigma^2$ характеризует количественное увеличение дисперсии уравнения регрессии. После подстановки (2-22) в (2-23) получим формулу, позволяющую рассчитать полное приращение дисперсии, вызванное погрешностью регистрации независимых переменных:

$$\Delta \sigma^2 \approx \sigma_y^2 \sum_{j=1}^k \frac{\zeta_j^2}{\sigma_j^2} (b^*_j)^2. \quad (2-24)$$

Влияние погрешности регистрации x_j на величину коэффициента b_j определяется частной производной [Л. 18]:

$$\frac{\partial b_j}{\partial \zeta_j^2} = - \frac{b_j}{\sigma_j^2} \frac{1}{1 - R^2_{j \cdot 1, 2, \dots, (k-1)}}, \quad (2-25)$$

где $R_{j \cdot 1, 2, \dots, (k-1)}$ — множественный коэффициент корреляции независимой переменной x_j на остальные независимые переменные. Выражение (2-25) показывает, что ошибки в измерении переменной x_j всегда вызывают уменьшение алгебраической величины соответствующего коэффициента b_j ; это уменьшение пропорционально зна-

чению b_j и коэффициенту корреляции между переменной x_j и остальными переменными $x_1, x_2, \dots, x_{(k-1)}$ и обратно пропорционально дисперсии переменной x_j . Чем теснее коррелированы независимые переменные, тем сильнее искажаются регрессионные коэффициенты ошибками регистрации.

Чтобы выяснить влияние на b_j ошибок измерения зависимой переменной y , рассмотрим частную производную

$$\frac{\partial b_j}{\partial \zeta_y^2} = \frac{\partial}{\partial \zeta_y^2} \left(-\frac{D_{yy}}{D_{yy}} \right) = 0.$$

Из последнего соотношения следует, что погрешности измерения зависимой переменной ε_y не влияют на величины коэффициентов уравнения регрессии.

Чувствительность коэффициента b_j к ошибкам измерения остальных независимых переменных определяется частными производными

$$\frac{\partial b_j}{\partial \zeta_q^2} = -b_q \frac{D_{yy \cdot jq}}{D_{yy}}. \quad (2-26)$$

В общем случае величина $b_q \frac{D_{yy \cdot jq}}{D_{yy}}$ может быть как положительной, так и отрицательной; поэтому погрешности в измерении независимых переменных x_q , где $q=1, \dots, (k-1)$ и $q \neq j$, могут вызывать как увеличение, так и уменьшение величины коэффициента b_j . Однако изменение b_j будет всегда пропорционально значению соответствующего коэффициента b_q . Для приближенной оценки количественного изменения величины b_j заменим в (2-25) и (2-26) дифференциалы на соответствующие приращения:

$$\begin{aligned} \partial b_j &\approx \Delta b_j = b_j - b_j^0; \\ \partial \zeta_j^2 &\approx \Delta \zeta_j^2 = \zeta_j^2. \end{aligned}$$

Тогда согласно (2-25) ошибки измерения переменной x_j вызывают уменьшение коэффициента b_j на величину

$$\Delta b_{j(j)} \approx \frac{\zeta_j^2}{\sigma_j^2} \frac{b_j}{1 - R_{j,1,2,\dots,(k-1)}^2}, \quad (2-27)$$

а изменение коэффициента b_j за счет погрешностей измерения переменной x_q определяется величиной

$$\Delta b_{j(q)} \approx -b_q \zeta_q^2 \frac{D_{yy \cdot jq}}{D_{yy}}. \quad (2-28)$$

В общем случае полное изменение Δb_j коэффициента b_j за счет ошибок измерения всех независимых переменных можно рассчитать по формуле

$$\Delta b_j = \sum_{q=1}^k \Delta b_{j(q)} \approx - \sum_{q=1}^k b_q \zeta_q^2 \frac{D_{yy \cdot jq}}{D_{yy}}. \quad (2-29)$$

В частном случае, когда независимые переменные не коррелированы между собой, все определители $D_{yy \cdot qq} = 0$ для $q \neq j$ и формула (2-29) принимает простой и удобный для расчетов вид:

$$\Delta b_j = \Delta b_{j(j)} \approx \frac{\hat{\xi}_j^2}{\hat{\sigma}_j^2} b_j, \quad (2-30)$$

поскольку в этом случае определители $D_{yy \cdot jj}$ и D_{yy} будут равны просто произведению своих диагональных элементов.

Высокий уровень погрешностей при регистрации статистических данных может существенно исказить действительное соотношение между величинами коэффициентов уравнения регрессии, что приведет к неверным выводам при технологической интерпретации полученных результатов. Формулы (2-29) и (2-30) дают возможность провести коррекцию коэффициентов и получить уравнение регрессии, свободное от искажающего влияния ошибок измерения.

Переходя к выборочным оценкам, основные результаты этого параграфа можно представить в следующем виде:

$$\Delta \hat{R}_{(j)}^2 \approx \frac{\hat{\xi}_j^2}{\hat{\sigma}_j^2} (\hat{b}^*_{jj})^2; \quad (2-31)$$

$$\Delta \hat{\sigma}^2 \approx \frac{n-1}{n-k} \hat{\sigma}_y^2 \sum_{j=1}^k \frac{\hat{\xi}_j^2}{\hat{\sigma}_j^2} \hat{b}^{*2}_{jj}; \quad (2-32)$$

$$\Delta \hat{b}_{j(q)} \approx -\hat{b}_q \hat{\xi}_q^2 \frac{\hat{D}_{yy \cdot jq}}{\hat{D}_{yy}}; \quad q = 1, \dots, k. \quad (2-33)$$

Выборочную оценку дисперсии ошибки $\hat{\xi}^2$ можно получить либо из паспортных характеристик используемого измерительного прибора, либо экспериментальным путем. Остальные данные для формул (2-31), (2-32) и (2-33) получаются в результате обработки матриц наблюдений X и Y , содержащих статистические данные, измеренные с ошибками. Выведенные соотношения позволяют решить следующие практически важные задачи:

1. Выделение переменных, ошибки измерения которых составляют наиболее существенную часть приращения остаточной дисперсии. Это позволяет сформулировать требования к точности регистрации отдельных технологических переменных.

2. Вычисление общего приращения остаточной дисперсии за счет погрешностей измерения независимых переменных.

3. Получение регрессионных коэффициентов, освобожденных от искажающего влияния ошибок измерения.

Приведенные выше результаты получены для линейного уравнения регрессии, т. е. когда $k=1$. Если в качестве модели (2-3) используется полином выше первого порядка, все формулы усложняются. Например, для квадратичного полинома формулы (2-31) и (2-32) соответственно имеют вид:

$$\Delta \hat{R}_{(j)кв}^2 \approx \frac{\hat{\xi}_j^2}{\hat{\sigma}_y^2} \left(\hat{b}_j^2 + \sum_{q,r=1}^l \hat{a}_{qr} \hat{b}_{jq} \hat{b}_{jr} \right); \quad (2-34)$$

$$\Delta \hat{\sigma}_{\text{кв}}^2 \approx \frac{n-1}{n-k} \sum_{j=1}^l \hat{\zeta}_j^2 \left(\hat{b}_j^2 + \sum_{q,r=1}^l \hat{a}_{qr} \hat{b}_{jq} \hat{b}_{jr} \right). \quad (2-35)$$

Пример 2-2. Математическое описание процесса коагуляции было получено в виде регрессионного уравнения, связывающего производительность лентоотливочной машины с рядом технологических параметров. После отбрасывания малозначимых регрессионных коэффициентов уравнение в стандартизованном масштабе измерения переменных имеет вид:

$$\hat{y}^* = -0,12 x^*_4 - 0,17 x^*_5 - 0,3 x^*_{11} + 0,18 x^*_{12} + 0,7 x^*_{14} + 0,14 x^*_{15} + 0,38 x^*_{17}. \quad (2-36)$$

Полученное уравнение предполагалось использовать для ориентировочного прогнозирования эффективности работы объекта по текущим значениям технологических переменных $x_4, x_5, x_{11}, x_{12}, x_{14}, x_{15}, x_{17}$.

Если учесть, что объект обладает большим транспортным запаздыванием вследствие медленного движения сита лентоотливочной машины, то текущее прогнозирование может быть применено при управлении объектом как способ своевременного обнаружения отклонений в режиме.

Для выяснения путей дальнейшего совершенствования полученной модели анализировалось влияние ошибок регистрации отдельных технологических переменных. Для этого вычислялись вклад ошибок измерения отдельных переменных в приращение остаточной дисперсии $\hat{\sigma}^2$ и величина общего приращения остаточной дисперсии за счет ошибок измерения технологических переменных.

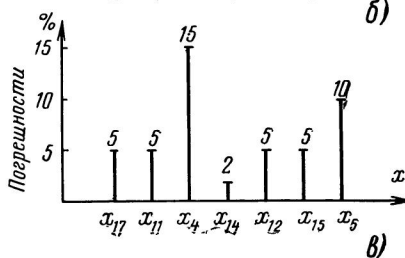
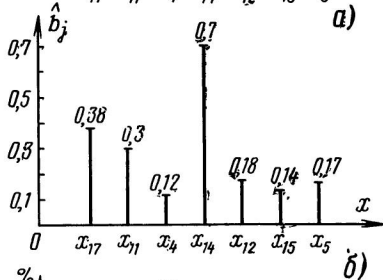
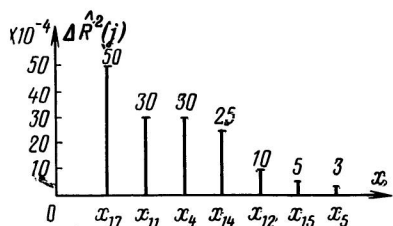
Вклад погрешности регистрации переменной x_j , равный по величине (2-31), в дальнейшем для краткости будем называть просто вкладом переменной x_j . Общее приращение остаточной дисперсии вычисляется по формуле (2-32). Для иллюстрации результатов расчетов в табл. 2-2 приведены выборочная дисперсия погрешности измерения $\hat{\zeta}_j^2$, выборочная дисперсия переменной $\hat{\sigma}_j^2$, абсолютная величина регрессионного коэффициента $\hat{b}_j^*$, приборная точность измерения переменной в процентах, величина вклада $\Delta \hat{R}_{(j)}^2$ по каждой технологической переменной, входящей в уравнение (2-36).

Таблица 2-2

Переменные	$\hat{\zeta}_j^2$	$\hat{\sigma}_j^2$	$\hat{b}_j^*$	Погрешность	$\Delta \hat{R}_{(j)}^2$
x_{17}	4,0	113,0	0,38	5	0,005
x_{11}	3,0	90,0	0,3	5	0,003
x_4	351	1 660	0,12	15	0,003
x_{14}	0,16	31,2	0,70	2	0,0025
x_{12}	0,3	96,9	0,18	5	0,001
x_{15}	1 100	47 700	0,14	5	0,0005
x_5	0,64	68,5	0,17	10	0,0003

Для наглядности полученные результаты представлены в виде диаграмм на рис. 2-5. На диаграмме рис. 2-5,а изображены ранжированные величины вкладов $\Delta R^2_{(j)}$; диаграмма рис. 2-5,б дает представление о сравнительном влиянии отдельных переменных на y , на диаграмме рис. 2-5,в приведены погрешности измерения переменных. Анализ диаграмм позволяет сделать интересные выводы:

1. Основными составляющими приращения $\Delta \hat{\sigma}^2$ являются вклады переменных x_{17} , x_{11} , x_4 и x_{14} . При этом особенно выделяется вклад



переменной x_{17} , хотя погрешность ее измерения сравнительно невелика. В то же время вклад переменной x_4 почти вдвое меньше, хотя погрешность ее измерения в 3 раза больше, чем погрешность измерения переменной x_{17} . На этом примере хорошо видно, что в роли переменных, особенно «неприятных» в смысле увеличения остаточной дисперсии, могут выступать переменные, которые регистрируются сравнительно точно, т. е. абсолютное значение погрешности измерения не является решающим для выбора приборов и методов регистрации переменных при подготовке эксперимента по сбору статистических данных. В нашем случае наиболее точно регистрируемая переменная x_{14} вносит весьма ощутимый вклад в $\Delta \hat{\sigma}^2$. Величина общего приращения остаточной дисперсии

$$\Delta \hat{\sigma}^2 = \frac{n-1}{n-7} \sum_{j=1}^7 \Delta R^2_{(j)} = 0,65.$$

Рис. 2-5. Диаграммы анализа погрешностей измерения.

общее приращение остаточной

измерения весьма невелико ($\sim 8\%$ от $\hat{\sigma}^2$). Поэтому в рассмотренном случае нецелесообразно тратить усилия на повышение точности регистрации переменных и следует искать иные пути совершенствования полученной модели (включение новых переменных, учет взаимодействий и т. д.).

Точность регрессионного уравнения (2-36) можно повысить, если уменьшить погрешность регистрации переменных x_{17} , x_{11} , x_4 , x_{14} . Однако дисперсия за счет ошибок

2-4. ОЦЕНКА РАБОТОСПОСОБНОСТИ МАТЕМАТИЧЕСКОГО ОПИСАНИЯ

Математическое описание в виде эмпирического уравнения регрессии может быть использовано при управлении технологическим объектом для прогнозирования значения выходной переменной y . Однако не всякое уравнение, полученное по экспериментальным данным, будет соответствовать требованиям, предъявляемым к точности управления объектом. Даже если гипотеза $H'_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ бракуется и тем самым подтверждается связь между y и переменными $x_1, \dots, x_k$, все же остается неясным, пригодно ли полученное математическое описание для целей управления или обеспечиваемая им точность прогнозирования недостаточна.

Для процессов с непрерывным производством (например, в химической, стекольной, бумажной промышленности) требования к точности управления можно сформулировать в виде ограничения на разность между действительным и предсказанным значениями выходной переменной: $\delta_{\max} \geq |y - \hat{y}|$. Математическое описание можно считать работоспособным, если $\delta \leq \delta_{\max}$. В качестве оценки работоспособности полученного математического описания можно использовать критерий, основанный на сопоставлении ошибок предсказания выходной переменной по среднему значению и по эмпирическому уравнению регрессии.

Уравнение

$$\hat{y}^* = \bar{y}, \quad (2-37)$$

где $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, обеспечивает наилучшее предсказание выходной переменной y в случае, когда неизвестны количественные соотношения между исследуемыми переменными объекта. Ошибка предсказания по уравнению (2-37)

$$\delta^*_i = y_i - \hat{y}^*_i = y_i - \bar{y}.$$

Используя основные допущения и понятие t -распределения случайной величины, нетрудно показать [Л. 20], что в каждом i -м наблюдении величина

$$t^*_i = \delta^*_i / \hat{\sigma}_y \sqrt{\frac{n+1}{n}}$$

является случайной, подчиняющейся центральному t -распределению с $(n-1)$ степенями свободы:

$$t^* \sim t(n-1). \quad (2-38)$$

На основании (2-38) можно записать равенство

$$P\{|\delta^*_i| = |y_i - \bar{y}| \leq \Delta^*_i\} = 1 - \alpha, \quad (2-39)$$

где $\Delta^* = t_{\alpha/2}(n-1) \hat{\sigma}_y \sqrt{\frac{n+1}{n}}$; $\hat{\sigma}_y^2$ — выборочная дисперсия выходной переменной; $t_{\alpha/2}(n-1)$ — табличное значение t -распределения для уровня значимости $\alpha/2$ и $(n-1)$ степеней свободы.

Из (2-39) следует, что ширина интервала единичного предсказания для уравнения (2-37) с вероятностью $P=1-\alpha$ не превышает по абсолютной величине значения Δ^* .

Теперь рассмотрим ошибку предсказания среднего значения выходной переменной по эмпирическому уравнению

$$\hat{y} = \hat{b}_1 (x_1 - \bar{x}_1) + \dots + \hat{b}_k (x_k - \bar{x}_k). \quad (2-40)$$

Значение $\hat{y}_t = \bar{y}$ соответствует технологической ситуации, когда $x_1 = \bar{x}_1, \dots, x_k = \bar{x}_k$. Согласно результату (1-66) ширина $(1-\alpha)$ -доверительного интервала единичного предсказания для уравнения (2-40) равна:

$$\Delta_t = t_{\alpha/2} (n-k) \hat{\sigma} \sqrt{1 + \frac{1}{n}}.$$

Заметим, что $t(n-1) = t(n-k)$, если $(n-k) > 30$, что имеет место практически всегда.

В качестве критерия, оценивающего эффективность применения эмпирического регрессионного уравнения, рассмотрим отношение

$$\hat{\omega} = \frac{\Delta_t}{\Delta_i^*} = \frac{t_{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n}}}{t_{\alpha/2} \hat{\sigma}_y \sqrt{1 + \frac{1}{n}}} = \frac{\hat{\sigma}}{\hat{\sigma}_y}. \quad (2-41)$$

Чтобы лучше уяснить физический смысл полученного критерия, преобразуем выражение (2-41). Используя соотношения

$$\left. \begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-k} (Y^T Y - \hat{B}^T X^T Y); \\ \hat{\sigma}_y^2 &= \frac{1}{n-1} Y^T Y, \end{aligned} \right\} \quad (2-42)$$

выражение (2-41) можно записать в виде

$$\hat{\omega} = \sqrt{\frac{n-1}{n-k} \left(1 - \frac{\hat{B}^T X^T Y}{Y^T Y} \right)}.$$

В разложении (2-42) величина $\hat{B}^T X^T Y$ соответствует уменьшению полной суммы квадратов $Y^T Y$ за счет переменных $x_1, \dots, x_k$. Переходя к пределу при $n \rightarrow \infty$, получаем:

$$\omega = \lim_{n \rightarrow \infty} \hat{\omega} = \sqrt{1 - \frac{B^T X^T Y}{Y^T Y}}. \quad (2-43)$$

Из (2-43) следует, что значения ω ограничены пределами $0 \leq \omega \leq 1$. В случае, когда $\omega = 0$, имеет место равенство

$$B^T X^T Y = Y^T Y,$$

т. е. в объекте отсутствует случайное возмущение e , и весь разброс

выходной переменной y обусловлен изменением технологических переменных $x_1, \dots, x_k$; предсказание по уравнению регрессии будет безошибочным. Величина $\omega=1$ соответствует случаю, когда

$$B^*X^*Y=0,$$

т. е. коэффициенты $\beta_1, \dots, \beta_k$ равны нулю, и нет никакой связи между выходной переменной объекта и теми технологическими переменными, которые введены в уравнение регрессии; весь разброс y обусловлен исключительно действием шумового поля e .

Если значение ω лежит в интервале $0 < \omega < 1$, то между y и переменными $x_1, \dots, x_k$ существует определенная зависимость, которая действует на фоне некоторого шумового поля, вносящего погрешность в предсказание по уравнению регрессии.

Следует отметить, что интервалы Δ^*_i и Δ_i расширяются при приближении к границе диапазона. Однако на основных свойствах ω -критерия это не сказывается.

Для практических целей удобнее вместо $\hat{\omega}$ использовать обратную величину, которую назовем коэффициентом $\hat{\gamma}$:

$$\hat{\gamma} = \frac{\Delta^*_e}{\Delta_e} = \frac{\hat{\sigma}_y}{\hat{\sigma}}.$$

Коэффициент γ показывает, во сколько раз уменьшается интервал ошибки предсказания при переходе от предсказания выходной переменной по среднему значению к предсказанию по эмпирическому уравнению регрессии.

Коэффициент $\hat{\gamma}$ дает возможность сравнить различные по форме и составу переменных уравнения регрессии и выбрать то из них, которое обеспечивает наилучшее предсказание. На практике всегда важно знать, насколько модель должна соответствовать описываемому объекту, требуется ли дальнейшее совершенствование полученного математического описания, нужно ли дополнительно повысить точность предсказания. Ответы на эти вопросы может дать пороговое значение $\hat{\gamma}_{\text{мин}}$, которое рассчитывается на основании общих требований, предъявляемых к точности управления данным объектом $\delta_{\text{макс}}$.

Интересно сопоставить коэффициент γ с коэффициентом множественной корреляции $R_{y \cdot x_1, \dots, x_k}$, который хотя и не имеет для регрессионного анализа такой четкой статистической интерпретации, как в корреляционном анализе, но все же часто используется исследователями в роли косвенного показателя работоспособности уравнения регрессии.

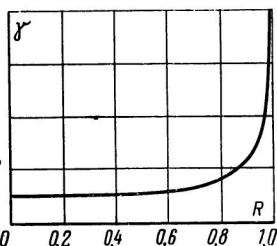


Рис. 2-6. Зависимость коэффициента $\hat{\gamma}$ от R .

На рис. 2-6 приведена графическая зависимость γ от R , из которой следует, что γ начинает резко возрастать в области больших значений R . Вероятно, эмпирическое уравнение регрессии имеет практический смысл, если $\hat{\gamma} \geq 2$, т. е. когда ошибка предсказания по уравнению регрессии хотя бы в 2 раза меньше, чем ошибка предсказания по среднему значению $\bar{y}$. Это соответствует значению $R \geq 0,86$. Из рис. 2-6 видно, что любая мера, обеспечивающая даже незначительное возрастание коэффициента множественной корреляции за пределами $R=0,86$, будет практически оправдана, так как при этом работоспособность уравнения может повыситься весьма существенно.

2-5. ОСОБЕННОСТИ ВЫЧИСЛИТЕЛЬНОЙ ПРОЦЕДУРЫ

Вычислительную процедуру множественной регрессии можно разбить на несколько этапов:

- 1) формирование системы нормальных уравнений;
- 2) получение обратной матрицы;
- 3) вычисление оценок $\hat{\beta}_1, \dots, \hat{\beta}_k$;
- 4) проверка гипотез и вычисление некоторых дополнительных величин;
- 5) выделение невлиятельных переменных.

Остановимся на каждом этапе подробнее. Согласно формулам (1-12) и (1-32) процедура формирования системы нормальных уравнений сводится к вычислению сумм квадратов и произведений по всем n наблюдениям переменных, входящих в уравнение регрессии. Форма (1-32) предпочтительнее, поскольку в этом случае возникает возможность вычисления $\hat{b}_0$ и улучшаются в вычислительном смысле условия последующего обращения матрицы. Иногда, чтобы избежать непосредственного центрирования каждого наблюдения в матрицах наблюдений X и Y , пользуются тождеством

$$\sum_{i=1}^n x_{ji}x_{qi} - \frac{1}{n} \sum_i x_{ji} \sum_i x_{qi} = \sum_i (x_{ji} - \bar{x}_j)(x_{qi} - \bar{x}_q).$$

Величины типа $\sum_i x_{ji}$ обычно представляют собой большие числа,

что при перемножении ведет к значительной потере точности. Поэтому в программе для ЦВМ лучше предусмотреть непосредственное центрирование каждого наблюдения.

С вычислительной точки зрения очень важно, чтобы элементы обращаемой матрицы незначительно различались по величине. Этого можно достичь, если использовать в качестве коэффициентов нормальных уравнений выборочные коэффициенты корреляции, т. е., помимо центрирования, провести стандартизацию наблюдений по формулам (1-42). Корреляционную матрицу целесообразно выводить на печать, поскольку она содержит информацию о парном взаимодействии переменных, которая может быть полезна при технологическом анализе математического описания.

Особо следует остановиться на образовании линеаризованных переменных, когда модель выбрана в виде полинома выше первого порядка. Возможны два пути:

1) образование линеаризованных переменных из нестандартизованных данных с последующей стандартизацией обобщенной матрицы наблюдений;

2) стандартизация первоначальной матрицы наблюдений и образование линеаризованных переменных по стандартизованным наблюдениям.

С точки зрения объема вычислений второй путь более громоздок, зато, как будет показано ниже, он более эффективен в смысле точности получаемых оценок.

Сильная взаимосвязь переменных, входящих в уравнение регрессии, не только снижает мощность χ^2 -критерия, но и является причиной плохой обусловленности системы нормальных уравнений, что порождает серьезные вычислительные трудности. Даже если входные переменные объекта слабо коррелированы, при образовании линеаризованных переменных все равно появляется сильная корреляция между такими переменными, например, квадратичного полинома, как x_j и x_j^2 ; x_j и $x_j x_q$; x_j^2 и $x_j x_q$. Это хорошо видно по корреляционной матрице примера 1-5. В [Л. 21] показано, что коэффициенты корреляции между этими переменными могут быть представлены в виде

$$r[x_j, x_j^2] = \frac{1}{\sqrt{1 + \frac{\delta_j^2}{2}}};$$

$$r[x_j, x_j x_q] = \frac{1}{\sqrt{1 + \delta_q^2 + \frac{\delta_q^2}{\delta_j^2}}};$$

$$r[x_j^2, x_j x_q] = \frac{1}{\sqrt{\left(1 + \frac{1}{2} \delta_j^2\right) (1 + \delta_q^2 + \delta_q^2 / \delta_j^2)}}.$$

Величины $\delta_j = \frac{\sigma_{x_j}}{M[x_j]}$, $\delta_q = \frac{\sigma_{x_q}}{M[x_q]}$ представляют собой коэффици-

циенты вариации входных переменных объекта x_j и x_q . Из приведенных выражений следует, что коэффициенты корреляции будут малы, когда коэффициенты вариации велики, а это достигается центрированием технологических переменных $x_1, \dots, x_l$. Полученные таким образом линеаризованные переменные целесообразно в свою очередь стандартизовать по формулам (1-42), чтобы привести к единому масштабу измерения все переменные $x_1, \dots, x_k$, входящие в уравнение регрессии, что позволит сопоставлять их влияние на выходную переменную при технологической интерпретации математического описания.

К первому этапу следует отнести и обсуждение проблемы выпавших наблюдений.

При сборе статистических данных на промышленном объекте нередко случается, что в i -м опыте значение x_j -й переменной по тем или иным причинам оказывается незарегистрированным. Этот опыт, естественно, не может полноценно использоваться для вычисления оценок $\hat{\beta}_1, \dots, \hat{\beta}_k$. Таких опытов в матрице наблюдений за время эксперимента может накопиться немало, и исследователь, решая, как поступить, может либо исключить все опыты, содержащие хотя бы одно выпавшее наблюдение, либо обратиться к специальным методам, позволяющим полезно использовать информацию, содержащуюся в неполноценных опытах.

Упоминаемые в литературе методы учета выпавших наблюдений можно разделить на две группы:

1) методы восстановления наиболее вероятного значения выпавшего наблюдения [Л. 22, 23];

2) методы, использующие для вычисления сумм квадратов $\sum x_j^2$ и парных произведений $\sum x_j x_q$ и $\sum x_j x_u$ все имеющиеся наблюдения, т. е. пределы суммирования для различных сочетаний переменных берутся различными в зависимости от количества наблюдений, общих для данной пары x_j и x_q .

Методы первой группы, как правило, основаны на весьма громоздких вычислениях, и вряд ли усилия в этом направлении можно считать оправданными. Интересные результаты по эффективности методов второй группы приведены в [Л. 24], содержащей также обширную библиографию. Авторами с помощью ЦВМ было исследовано восемь регрессионных моделей, включающих различное количество переменных с различной степенью коррелированности. Массивы данных, по которым определялись оценки коэффициентов моделей, содержали от 10 до 80% опытов с выпавшими наблюдениями. При получении оценок коэффициентов каждой модели выпавшие данные учитывались двумя способами:

1) все опыты с выпавшими данными вычеркивались, и оценки определялись по оставшимся данным обычным путем;

2) для формирования системы нормальных уравнений использовались суммы квадратов и парных произведений, вычисленные по всем наблюдениям, общим для данной пары.

Оценки коэффициентов $\hat{\beta}_1, \dots, \hat{\beta}_k$ для первого и второго способов сопоставлялись по величине дисперсии и близости к истинным значениям коэффициентов модели $\beta_1, \dots, \beta_k$. Из приведенных в работе результатов следует, что в подавляющем большинстве случаев первый способ обеспечивал существенно лучшие оценки коэффициентов по сравнению со вторым способом. Лишь в единственном случае, когда после вычеркивания опытов с выпавшими наблюдениями в массиве осталось всего 10% первоначальных данных, по второму способу был получен более точный результат.

Резюмируя сказанное, можно сделать следующий вывод. Принимая во внимание стремление упростить вычислительный алгоритм и сократить и без того громоздкую вычислительную процедуру множественной регрессии, целесообразно исключить из матриц наблюдений опыты, содержащие выпавшие наблюдения, и вести обработку оставшихся полноценных опытов обычным путем.

Прежде чем обсуждать этап получения обратной матрицы $[(X^*X)^{-1}]$, заметим, что элементы обратной матрицы необходимы только для вычисления дисперсии и ковариации оценок $\hat{\beta}_1, \dots, \hat{\beta}_k$. Если

дисперсии и ковариации не нужны, то оценки коэффициентов могут быть получены просто путем решения системы нормальных уравнений без обращения матрицы. Специфической особенностью обрабатываемой матрицы S является ее симметричность. Для получения инверсии $C=S^{-1}$ удобен метод Дулиттла, основанный на следующем рассуждении.

Пусть $C=\|C_1, \dots, C_k\|$, где C_j — j -й столбец матрицы C . Тогда $S\|C_1, \dots, C_k\|=\|I_{(1)}, \dots, I_{(k)}\|$, где $I_{(j)}$ — j -й столбец единичной матрицы I размером $(k \times k)$. Для получения элементов матрицы C нужно решить k систем уравнений типа $SC_j=I_{(j)}$; $j=1, \dots, k$.

Пример 2-3. Требуется определить элементы матрицы C , обратной матрице

$$X^T X = \begin{vmatrix} 2 & 4 & 2 \\ 4 & 10 & 2 \\ 2 & 2 & 12 \end{vmatrix}.$$

Найдем C_1 , C_2 и C_3 , решая следующие системы уравнений:

$$\begin{aligned} 2c_{11}+4c_{12}+2c_{13} &= 1; & 2c_{21}+4c_{22}+2c_{23} &= 0; & 2c_{31}+4c_{32}+2c_{33} &= 0; \\ 4c_{11}+10c_{12}+2c_{13} &= 0; & 4c_{21}+10c_{22}+2c_{23} &= 1; & 4c_{31}+10c_{32}+2c_{33} &= 0; \\ 2c_{11}+2c_{12}+12c_{13} &= 0; & 2c_{21}+2c_{22}+12c_{23} &= 0; & 2c_{31}+2c_{32}+12c_{33} &= 1. \end{aligned}$$

Обратная матрица C имеет вид:

$$C = \|C_1 C_2 C_3\| = \frac{1}{8} \begin{vmatrix} 29 & -11 & -3 \\ -11 & 5 & 1 \\ -3 & 1 & 1 \end{vmatrix}.$$

Если из-за сильной коррелированности переменных система нормальных уравнений плохо обусловлена, то даже незначительные ошибки в вычислении элементов матрицы $X^T X$ вызывают большие ошибки при вычислении оценок $\hat{\beta}_1, \dots, \hat{\beta}_k$, причем ошибки возрастают с увеличением порядка матрицы $X^T X$. При получении математического описания сложного промышленного объекта приходится иметь дело с матрицами до 20-го порядка, и точность вычислений в таких случаях играет весьма важную роль. Практика показывает, что при использовании стандартизации для образования линейаризованных переменных и последующей стандартизации линейаризованных переменных вычислительная машина типа «Минск-22» успешно справляется с процедурой обращения матрицы даже при высоких значениях коэффициентов парной корреляции. В особо трудных случаях можно воспользоваться методом коррекции обратной матрицы, предложенным М. Г. Шульцем [Л. 25] и развитым Х. Хотеллингом [Л. 26]. Метод основан на использовании рекуррентного соотношения

$$C^{(f+1)} = C^{(f)} [2I_k - SC^{(f)}],$$

для которого доказано, что разность $C^{(f+1)} - C^{(f)}$ стремится к нулевой матрице при $f \rightarrow \infty$.

Пример 2-4. Пусть S и $C=S^{-1}$ — матрицы вида

$$S = \begin{vmatrix} 2 & -3 \\ -3 & 5 \end{vmatrix}; \quad C = \begin{vmatrix} 5 & 3 \\ 3 & 2 \end{vmatrix}.$$

Предположим, что из-за ошибок вычисления мы получили не точное значение C , а некоторую матрицу $C^* = \begin{vmatrix} 4,989 & 3,011 \\ 3,012 & 1,001 \end{vmatrix}$. Тогда $2I - SC^* = \begin{vmatrix} 1,058 & -0,049 \\ -0,093 & 1,078 \end{vmatrix}$, и, следовательно, первое скорректированное приближение есть

$$C^{(1)} = C^* (2I - SC^*) = \begin{vmatrix} 4,908 & 3,001 \\ 3,002 & 1,999 \end{vmatrix},$$

что в целом существенно ближе к действительной матрице C . Продолжая эту процедуру, можно получить дальнейшее уточнение элементов обратной матрицы.

Если вычислена обратная матрица C , то оценки коэффициентов в стандартизованном масштабе легко определяются по формуле

$$\hat{b}^*_j = \sum_{q=1}^b c_{jq} r_{yq}; \quad j = 1, \dots, k.$$

Не приводя коэффициенты $\hat{b}^*_j$ к натуральному масштабу, можно вычислить коэффициент множественной корреляции $\hat{R}_{y \cdot x_1, \dots, x_k}$ по формуле (1-47) и остаточную дисперсию в стандартизованной форме $\hat{\sigma}_0^2$ по формуле (1-45), причем $\hat{\sigma}_y^2 = 1$. Более точно остаточная дисперсия вычисляется по формуле

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^n (y^*_i - \hat{y}^*_i)^2}{n - k}, \quad (2-44)$$

где $\hat{y}^*_i$ — значение выходной переменной в i -м опыте, полученное подстановкой значений $x^*_{1i}, \dots, x^*_{ki}$ в найденное уравнение регрессии. Алгебраически выражения (1-45) и (2-44) тождественны, однако с вычислительной точки зрения формула (2-44) дает лучший результат.

Выведенные на печать разности $(y^*_i - \hat{y}^*_i)$; $i = 1, \dots, n$ трудно обозримы, однако их графическое представление бывает весьма полезно. Значения $\hat{y}^*_i$ могут быть использованы для проверки гипотезы H'_0 : $\beta_1 = \dots = \beta_k = 0$ по формуле (1-80), где $\hat{\sigma}^2 = \hat{\sigma}_0^2$. Если все переменные стандартизовались и, следовательно, матрица S представляет собой корреляционную матрицу, то порядок выделения невлиятельных переменных можно определить следующим образом:

1. Вычислить для всех коэффициентов величины $v_{2j} = \hat{b}^*_j / \hat{\sigma}_0 \sqrt{c_{jj}/(n-1)}$; $j = 1, \dots, k$.

2. Выделить минимальную v_{2q} и сравнить ее с константой $t_{\text{табл}} = 1,96$, если вероятность ошибки первого рода выбрана равной $P(I) = \alpha = 0,05$. При выборе другого значения α константа $t_{\text{табл}}$

равна соответствующему значению t -распределения для $(n-1)$ степени свободы.

3. Если $u_{2\min} = u_{2q} \geq t_{\text{табл}}$, то с вероятностью $P = 1 - P(I)$ можно считать, что все переменные полученного уравнения регрессии являются существенными.

Если $u_{2\min} < t_{\text{табл}}$, то гипотеза $H''_0: \beta_q = 0$ принимается, соответствующая переменная x_q считается невлиятельной и исключается из уравнения, после чего все оценки коэффициентов и других величин должны быть пересчитаны, и программа выделения невлиятельных переменных возвращается к первому этапу.

Цикл повторяется до тех пор, пока не будут выделены и исключены все невлиятельные переменные.

Чтобы не повторять процедуру обращения матрицы после исключения каждой невлиятельной переменной, можно воспользоваться формулами В. Кочрена [Л. 27], позволяющими корректировать оценки коэффициентов и элементы обратной матрицы с учетом исключенной переменной x_q :

$$\hat{b}^*_{j \text{ кор}} = \hat{b}^*_j - \frac{c_{jq}}{c_{qq}} \hat{b}^*_q; \quad j = 1, 2, \dots, (q-1), (q+1), \dots, (k-1), k;$$

$$c_{j \text{ кор}} = c_{jq} - \frac{c_{jq}c_{gq}}{c_{qq}}; \quad j, g = 1, 2, \dots, (q-1), (q+1), \dots, (k-1), k;$$

$$\hat{b}^*_{q \text{ кор}} = 0; \quad c_{qj \text{ кор}} = c_{jq \text{ кор}} = 0; \quad j = 1, \dots, k.$$

После выделения всех невлиятельных переменных оставшиеся коэффициенты пересчитываются в натуральном масштабе измерения переменных, и вычисляется натуральное значение остаточной дисперсии $\hat{\sigma}^2$ с учетом изменившегося значения $k-m$, где m — количество оставшихся членов эмпирического уравнения регрессии.

ГЛАВА ТРЕТЬЯ

РЕКУРРЕНТНЫЕ АЛГОРИТМЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКОГО ОПИСАНИЯ ДРЕЙФУЮЩИХ ОБЪЕКТОВ

При построении математического описания промышленных объектов часто приходится сталкиваться с фактом неконтролируемого изменения модели объекта во времени. Объекты, обладающие таким свойством, называют нестационарными или дрейфующими, а процесс изменения характеристик во времени — дрейфом характеристик объекта. Причина появления дрейфа — наличие неконтролируемых воздействий, значения которых не могут быть учтены при построении математического описания. Дрейф характеристик процесса возникает, например, из-за изменения активности катализатора, старения агрегата, сезонных изменений условий работы и т. п. Как правило, процесс изменения дрейфующих параметров значительно медленнее процесса изменения контролируемых переменных объекта. Если изменение управляющих воздействий производится при управлении некоторым объектом в среднем, например, 1 раз в течение каждого

часа, то неконтролируемая активность катализатора может практически не изменяться в течение суток, старение агрегата будет сказываться на выходном продукте лишь после нескольких месяцев эксплуатации и т. д.

Экспериментальное построение математического описания объекта обычно предшествует решению задачи выбора управляющего воздействия. Ясно, что при этом найденная модель должна соответствовать моменту осуществления управления, т. е. должна отражать текущее состояние управляемого объекта.

Часто промышленные объекты обладают существенным временным запаздыванием. Это приводит к тому, что во всякий момент времени решать вопрос о значениях параметров модели объекта приходится на основании старой информации, собранной в предшествующие моменты. При этом необходимо некоторым образом прогнозировать значения параметров модели.

Существенной особенностью задачи, рассматриваемой в данной главе, является то, что неизвестные параметры модели объекта представляют собой процесс, т. е. их действительные значения изменяются во времени. В связи с этим задача оценки неизвестных параметров является задачей оценки значений процесса. Подобные задачи являются предметом изучения теории оптимальной фильтрации [Л. 5, 28].

Применение методов теории фильтрации для построения аналитическим путем оптимального в некотором смысле алгоритма оценки дрейфующих параметров модели объекта предполагает априорное знание (т. е. знание, предшествующее построению и использованию алгоритма) определенных свойств дрейфа [Л. 29, 30]. Например, если процесс изменения параметров модели случайный и стационарный, то для того, чтобы найти наилучший (в смысле минимума среднего квадрата ошибки) линейный алгоритм решения задачи, т. е. алгоритм, при котором искомые оценки представляют собой линейные комбинации результатов измерения выхода объекта, необходимо знать корреляционные функции дрейфующих параметров [Л. 28].

Ограниченность аналитического подхода состоит прежде всего в том, что далеко не всегда имеется вся информация о характере изменения параметров модели, необходимая для аналитического построения оптимального алгоритма.

В этом случае, по-видимому, разумно попытаться восполнить недостающую информацию о дрейфе экспериментальным путем, например, определяя оценки неизвестных параметров корреляционных функций дрейфующих коэффициентов, если вид этих функций известен, или параметров алгоритма решения задачи, если структура алгоритма решения задана, используя информацию, поступающую от объекта при его эксплуатации. Такой путь решения задачи называют адаптивным [Л. 31, 32].

В предыдущих главах при построении математического описания объекта по выборке оценивались только математические ожидания коэффициентов, так как в предположении отсутствия дрейфа математические ожидания и действительные значения коэффициентов в любой момент времени совпадают. В этом разделе мы отказываемся от требования неизменности коэффициентов. При этом знания только математического ожидания коэффициентов уже недостаточно для построения математического описания. Принятый нами адаптивный путь решения задачи приводит к необходимости получения по выборке, помимо математических ожиданий коэффициентов,

оценки еще ряда параметров, значения которых определяются неизвестными априори свойствами дрейфа.

При практическом решении задач построения математического описания дрейфующего объекта возможны ситуации двух типов:

1. Получена выборка наблюдений за входом и выходом объекта на некотором интервале времени от t_0 до $t > t_0$; необходимо найти оценки вектора коэффициентов модели, близкие к действительным значениям коэффициентов в момент $t + \alpha$, где $\alpha \geq 0$.

2. Осуществляется непрерывное слежение за изменением коэффициентов модели. Во всякий текущий момент t требуется найти наилучшее в некотором смысле приближение к действительным значениям коэффициентов в момент $t + \alpha$.

Вторая ситуация по существу аналогична первой при стремлении длины интервала наблюдений к бесконечности. В этой главе будет рассмотрен один из возможных подходов к решению задачи оценки параметров модели дрейфующих объектов.

3-1. ПОСТАНОВКА ЗАДАЧИ И ОПИСАНИЕ МЕТОДА РЕШЕНИЯ

Рассмотрим постановку задачи построения математического описания дрейфующего объекта. Пусть имеется выборка n значений входных переменных объекта $x(i)$ и выходной переменной $y(i)$. Здесь

$$i = t_0 + \Delta t; t_0 + 2\Delta t; \dots; t_0 + n\Delta t,$$

где i — дискретные моменты измерения входа и выхода объекта; Δt — интервалы времени, через которые осуществляется эксперимент.

В дальнейшем будем полагать, что i принимает значения 1, 2, ..., n . Значения входных переменных в моменты съема информации представим в виде матрицы

$$X_n = \begin{pmatrix} x_1(1) & x_2(1) & \dots & x_k(1) \\ x_1(n) & x_2(n) & \dots & x_k(n) \end{pmatrix}, \quad (3-1)$$

а наблюдаемые значения выхода — в виде вектора-столбца

$$Y_n = \begin{pmatrix} y(1) \\ y(2) \\ \dots \\ y(n) \end{pmatrix}. \quad (3-2)$$

Положим, что

$$y(i) = u(i) + e(i), \quad (3-3)$$

где $u(i)$ — действительное значение выхода объекта в момент i ; $e(i)$ — ошибка измерения выхода.

Пусть модель объекта имеет вид:

$$u(i) = B^T(i) x(i), \quad (3-4)$$

где $B(i)$ — значение вектора действительных значений коэффициентов в момент i ; $x(i)$ — значение вектора входных переменных в момент i .

Будем полагать, что последовательности значений $B(i)$ и $x(i)$ являются случайными, стационарными и эргодическими. Относитель-

Но ошибки $e(i)$ сохраним предположения, введенные в § 1-1. Таким образом, исходные предпосылки отличаются от рассмотренных в гл. 1 лишь тем, что действительные значения коэффициентов модели изменяются во времени. Задача заключается в нахождении оценки $\widehat{B}(n+\alpha)$ вектора коэффициентов в момент $(n+\alpha)$, где $\alpha = 0, 1, 2, \dots$, $\alpha \ll n$, наилучшей в некотором определенном смысле. При этом, если $\alpha > 0$, наша задача является задачей прогнозирования значений вектора коэффициентов, а при $\alpha = 0$ — задачей фильтрации. При объединении обоих случаев можно говорить о слежении за процессом изменения коэффициентов модели. Наш подход к решению задачи прежде всего характеризуется тем, что мы будем полагать, что структура алгоритма вычисления оценок коэффициентов модели объекта всюду задается с точностью до значений некоторого небольшого числа параметров γ . Оптимальные свойства заданного алгоритма достигаются путем выбора значений параметров γ в рамках заданной структуры [Л. 33]. Реализуя такой подход, естественно, можно потерять что-то в смысле точности нахождения оценок $\widehat{B}(n+\alpha)$, если принятая структура алгоритма будет отличаться от оптимальной для действительного характера процесса $B(i)$. С другой стороны, такой подход позволяет не привязываться к информации о свойствах процесса $B(i)$, которая в практических задачах часто отсутствует. Один из возможных способов реализации такого подхода широко известен как метод «замороженных» коэффициентов [Л. 29] или метод «скользящего интервала» [Л. 34]. Считается, что в течение некоторого интервала наблюдений $T+\alpha$ можно принять, что $B(i) \sim \text{const}$, где

$$i = n+\alpha; n-1+\alpha; \dots; n-T.$$

Тогда, пользуясь информацией, получаемой в интервале наблюдений $T(i=n-T, \dots, n)$, можно найти наилучшие в смысле минимума среднего квадрата ошибки значения оценок $\widehat{B}(n+\alpha)$. Метод решения при этом допущении полностью аналогичен методу регрессионного анализа, рассмотренному ранее. Однако при реализации такого подхода возникает необходимость наилучшего выбора длины интервала T , так как увеличение этого интервала приводит, с одной стороны, к росту ошибок определения коэффициентов за счет изменения их действительных значений за время выборки, а с другой стороны, уменьшает влияние ошибки эксперимента на полученные оценки. Поэтому фактически такой метод решения относится к рассматриваемой нами группе методов, оптимальные свойства которых достигаются путем наилучшего в некотором смысле выбора значений параметров алгоритма в рамках заданной структуры. Здесь будут рассматриваться алгоритмы вычисления оценок $\widehat{B}(n+\alpha)$, представленные в рекуррентной форме, связывающей оценку для момента $(n+\alpha)$ с ранее найденной оценкой для момента $(n+\alpha-1)$ и вновь поступившей информацией в момент n . В общем виде рекуррентный алгоритм записывается следующим образом:

$$\widehat{B}(n+\alpha) = F[\widehat{B}(n+\alpha-1), x(n), y(n), \gamma], \quad (3-5)$$

где F — некоторая функция, определяющая структуру алгоритма; γ — свободные параметры алгоритма, значения которых могут выбираться произвольно в рамках заданной структуры.

Подобная форма лучше, чем какая-либо иная, отвечает сущности задачи слежения за дрейфующими параметрами объекта. Положим, что оценка $\widehat{B}(\alpha)$ вектора коэффициентов для момента α известна (или задан способ ее вычисления по выборке наблюдений в моменты $i=1, 2, 3, \dots, n$), а структура алгоритма выбрана в виде рекуррентного соотношения (3-5). Тогда для имеющейся выборки оценка $\widehat{B}(n+\alpha)$ будет зависеть лишь от величины γ . При отсутствии какой-либо информации о процессе изменения $B(i)$ выбор наилучшего в некотором смысле значения γ должен быть произведен по имеющейся выборке.

Заметим, что при всяком значении γ и выборке X_n и Y_n , используя алгоритм (3-5), можно получить последовательность оценок $\widehat{B}(\alpha+1), \widehat{B}(\alpha+2), \dots, \widehat{B}(\alpha+n)$. Имея эту последовательность, вычислим величину $S_\alpha(\gamma)$, представляющую собой сумму квадратов ошибок слежения за измеряемым выходом объекта по выборке:

$$S_\alpha(\gamma) = \sum_{j=\alpha}^n [y(j) - \widehat{B}^T(j) x(j)]^2. \quad (3-6)$$

Величину $S_\alpha(\gamma)$ мы будем использовать в качестве критерия для нахождения оценки $\widehat{\gamma}_{\text{опт}}$ по выборке. Тогда оценка $\widehat{\gamma}_{\text{опт}}$ определится из следующего условия:

$$\min_{\gamma} \sum_{j=\alpha}^n [y(j) - \widehat{B}^T(j) x(j)]^2$$

при

$$\widehat{B}(j) = F[\widehat{B}(j-1), x(j-\alpha), y(j-\alpha), \gamma]; \quad (3-7)$$

а значения оценок коэффициентов для момента $(n+\alpha)$

$$\widehat{B}(n+\alpha) = F[\widehat{B}(n+\alpha-1), x(n), y(n), \widehat{\gamma}_{\text{опт}}], \quad (3-8)$$

где в свою очередь

$$\widehat{B}(n+\alpha-1) = F[\widehat{B}(n+\alpha-2), x(n-1), y(n-1), \widehat{\gamma}_{\text{опт}}]$$

и т. д.

Заметим, что при использовании алгоритма вида (3-5) оценка $\widehat{B}(j)$ всегда является функцией параметра γ , так что вместо $\widehat{B}(j)$ везде следовало бы писать $\widehat{B}(\gamma, j)$. Здесь, так же как в дальнейшем изложении, аргумент γ опускается.

При использовании критерия (3-7) возникает вопрос о свойствах получаемых оценок параметров алгоритма. Рассмотрим величину $Q(\gamma)$, представляющую собой средний квадрат ошибки слежения за действительным выходом объекта в момент n , когда n стремится

к бесконечности, т. е. в установившемся режиме слежения

$$Q(\gamma) = M \left\{ \lim_{n \rightarrow \infty} [u(n) - \hat{B}^T(n) \mathbf{x}(n)]^2 \right\}. \quad (3-9)$$

Здесь M — символ математического ожидания, взятого по множеству реализаций.

Значение γ , минимизирующее (3-9), обозначим $\gamma_{\text{опт}}$. Тогда можно доказать следующее утверждение относительно оценки $\hat{\gamma}_{\text{опт}}$.

Пусть алгоритм (3-5) обеспечивает для заданного $\hat{B}(\alpha)$ и значений γ , выбираемых из некоторого определенного диапазона, слежение за дрейфом коэффициентов модели, так что разность $\xi(i) = y(i) - \hat{B}^T(i) \mathbf{x}(i)$ является стационарным случайным процессом, для которого справедлива гипотеза эргодичности, и пусть $\alpha > 0$, т. е. решается задача прогнозирования, и ошибка $e(i)$ во всякий момент i является независимой случайной величиной с нулевым математическим ожиданием и дисперсией $\hat{\sigma}^2$. Тогда оценка $\hat{\gamma}_{\text{опт}}$ является состоятельной в том смысле, что с вероятностью единица

$$\lim_{n \rightarrow \infty} \hat{\gamma}_{\text{опт}} = \gamma_{\text{опт}}. \quad (3-10)$$

Приведем доказательство этого утверждения. В силу эргодичности процесса $\xi(j)$ имеем с вероятностью единица:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n - \alpha} S_{\alpha}(\gamma) &= \lim_{n \rightarrow \infty} \frac{1}{n - \alpha} \sum_{j=\alpha}^n [y(j) - \hat{B}^T(j) \mathbf{x}(j)]^2 = \\ &= M [y(j) - \hat{B}^T(j) \mathbf{x}(j)]^2. \end{aligned} \quad (3-11)$$

Раскроем выражение под знаком математического ожидания в (3-11), подставив вместо $y(j)$ выражение (3-3):

$$\begin{aligned} M [y(j) - \hat{B}^T(j) \mathbf{x}(j)]^2 &= M [u(j) - \hat{B}^T(j) \mathbf{x}(j)]^2 + \sigma^2 - \\ &- 2M \{e(j) [u(j) - \hat{B}^T(j) \mathbf{x}(j)]\}. \end{aligned} \quad (3-12)$$

Последнее слагаемое в (3-12) при $\alpha > 0$ в силу постулированных свойств ошибки $e(j)$ равно нулю, так как при $\alpha > 0$ оценка $\hat{B}(j)$ не зависит от $e(j)$. В силу стационарности процесса $\xi(j)$ имеем:

$$\begin{aligned} M \{[y(j) - \hat{B}^T(j) \mathbf{x}(j)]^2\} &= \lim_{n \rightarrow \infty} M \{[y(n) - \hat{B}^T(n) \mathbf{x}(n)]^2\} = \\ &= M \{ \lim_{n \rightarrow \infty} [u(n) - \hat{B}^T(n) \mathbf{x}(n)]^2 \} + \sigma^2 = Q(\gamma) + \sigma^2. \end{aligned}$$

Таким образом, при n , стремящемся к бесконечности, величина $\frac{1}{n - \alpha} S_{\alpha}(\gamma)$ с вероятностью единица совпадает с величиной $Q(\gamma) + \sigma^2$.

Это означает, что значение $\gamma_{\text{опт}}$, минимизирующее средний квадрат ошибки прогнозирования действительного выхода объекта $Q(\gamma)$,

будет одновременно минимизировать и $\hat{S}_\alpha(\gamma)$ при n , стремящемся к бесконечности, т. е. соотношение (3-10) имеет место.

Доказанное утверждение свидетельствует о том, что используемый для нахождения $\hat{\gamma}_{\text{опт}}$ критерий (3-6) обеспечивает с вероятностью единица при $\alpha > 0$ и достаточно больших n минимальную величину среднего квадрата ошибки прогнозирования действительного выхода объекта. Заметим, что все сказанное относится лишь к случаю решения задачи прогнозирования и не имеет места при фильтрации выхода объекта.

При доказательстве состоятельности оценки $\hat{\gamma}_{\text{опт}}$ предполагалось, что $\xi(j)$ — стационарный случайный процесс. Это означает, что с самого первого эксперимента режим слежения является установившимся, при котором математическое ожидание и дисперсия случайной величины $\xi(j)$ остаются неизменными.

Однако оценки будут состоятельными и в том случае, когда переходный процесс заканчивается за конечное число шагов m , т. е. когда стационарность $\xi(j)$ имеет место лишь при $j > m$. Для того чтобы убедиться в этом, перепишем (3-11) следующим образом:

$$\lim_{n \rightarrow \infty} \frac{1}{n - \alpha} S_\alpha(\gamma) = \lim_{n \rightarrow \infty} \frac{1}{n - \alpha} \sum_{j=\alpha}^m \xi^2(j) + \\ + \lim_{n \rightarrow \infty} \frac{n - m}{n - \alpha} \frac{1}{n - m} \sum_{j=m+1}^n \xi^2(j).$$

Первое слагаемое в правой части равно нулю, а второе в силу эргодичности $\xi(j)$ с вероятностью единица стремится к $M\{\xi^2(j)\}$. Таким образом, мы вновь имеем (3-11), и все последующие выводы оказываются справедливыми при $j > m$.

Длительность переходного процесса (m) при некотором значении γ зависит от выбора начального приближения $\hat{B}(\alpha)$. Переходный процесс будет практически отсутствовать, если $\hat{B}(\alpha)$ близко к $B(\alpha)$.

Такой выбор можно обеспечить, принимая в качестве $\hat{B}(\alpha)$ среднеквадратическую оценку вектора коэффициентов по выборке. Основанием для этого являются допущения о стационарности процесса $B(i)$ и сравнительно медленном характере дрейфа.

В случаях, когда используется алгоритм типа (3-5), а коэффициенты модели описываются стационарным случайным процессом с известными статистическими свойствами, выбор $\gamma_{\text{опт}}$ может быть осуществлен аналитически путем минимизации $Q(\gamma)$ (3-9). Пример решения подобной задачи рассмотрен в § 3-5. При этом в случае, когда алгоритм (3-5) является линейным, т. е. значение $\hat{B}(n)$ представляется для любого n как некоторая линейная комбинация наблюдений $y(1), y(2), \dots, y(n)$, для нахождения $\gamma_{\text{опт}}$ достаточно знать корреляционную функцию процесса $B(n)$.

Итак, построение математического описания дрейфующего объекта по выборке наблюдений в моменты $i=1, 2, \dots, n$ рассматривается здесь как задача прогнозирования значений коэффициентов модели объекта для момента времени $(n+\alpha)$. Решение ищется с по-

мощью рекуррентных алгоритмов вида (3-5), структура которых считается заданной, а значения неопределенных параметров γ оцениваются по выборке.

В § 3-2 и 3-3 рассмотрен алгоритм типа (3-5), вытекающий из метода текущего регрессионного анализа [Л. 35], а также некоторые иные алгоритмы, использующие структуру методов стохастической аппроксимации [Л. 32].

Оптимальные с точки зрения минимума среднего квадрата ошибки прогнозирования свойства используемых алгоритмов достигаются путем выбора значений параметров γ алгоритмов в соответствии с критерием (3-6) по информации, содержащейся в выборке. Начальное значение $\hat{B}(\alpha)$, принимаемое за значение коэффициента в момент α , находится путем решения обычной задачи регрессионного анализа по имеющейся выборке наблюдений. В § 3-4 описан алгоритм поиска $\hat{\gamma}_{\text{опт}}$ по конечной выборке наблюдений на основе критерия (3-6). В § 3-5 рассмотрен вопрос о выборе $\gamma_{\text{опт}}$ при известной корреляционной функции процесса $B(i)$.

В гл. 4 рассмотрен метод кусочно-линейной аппроксимации характеристик дрейфующих объектов, построенный на базе алгоритмов гл. 3. Здесь также приводятся алгоритм решения задачи по конечной выборке и необходимые выражения для нахождения $\gamma_{\text{опт}}$ в случае известных характеристик дрейфа.

Реализация рассматриваемых здесь методов построения математических описаний дрейфующих объектов ориентирована на использование вычислительных средств в виде управляющих вычислительных машин или специализированных устройств.

3-2. МЕТОД ТЕКУЩЕГО РЕГРЕССИОННОГО АНАЛИЗА

Пусть по некоторому числу наблюдений n над входом и выходом объекта в моменты $i=1, 2, \dots, n$ ищется оценка коэффициентов математического описания объекта в момент времени $(n+\alpha)$, где $\alpha=1, 2, \dots, v$, причем $v \ll n$. Влияние информации, полученной в моменты времени $n, n-1$ и т. д., на точность оценки $\hat{B}(n+\alpha)$, вообще говоря, неодинаково. Для того чтобы учесть изменение ценности информации, введем функцию веса $\varphi(n, j, \gamma)$. Здесь n — число наблюдений в выборке; j — номер текущего наблюдения; γ — некоторые параметры функции веса, значения которых будут выбираться в процессе настройки алгоритма.

Для определения структуры алгоритма вычисления $\hat{B}(n+\alpha)$ воспользуемся критерием наименьших квадратов, произведя взвешивание квадратов ошибок при различных экспериментах с помощью функции $\varphi(n, j, \gamma)$:

$$L = \sum_{j=1}^n [y(j) - \hat{B}^T(n+\alpha) x(j)]^2 \varphi(n, j, \gamma). \quad (3-13)$$

Вычисление $\hat{B}(n+\alpha)$ из условия минимума L по $\hat{B}(n+\alpha)$ приводит к методу текущего регрессивного анализа [Л. 35, 36].

В зависимости от выбора функции веса $\varphi(n, j, \gamma)$ алгоритм, соответствующий критерию (3-13), будет давать различные оценки $\hat{B}(n+\alpha)$. Таким образом, критерий (3-13) определяет лишь структуру

ру алгоритма вычисления $\widehat{B}(n+\alpha)$, не гарантируя в общем случае получение оценок $\widehat{B}(n+\alpha)$, близких в каком бы то ни было смысле к $B(n+\alpha)$. Для достижения минимальной величины среднего квадрата ошибки прогнозирования выхода объекта в момент $(n+\alpha)$ необходимо наилучшим образом выбрать функцию веса $\varphi(n, j, \gamma)$. Рассмотрим функции веса ограниченного класса. Будем полагать, что с увеличением «возраста» наблюдений ценность их для определения оценки $\widehat{B}(n+\alpha)$ убывает, т. е. функция веса $\varphi(n, j, \gamma)$ уменьшается с ростом $|n-j|$. Это предположение, хотя и кажется весьма естественным, накладывает определенные ограничения на тип процессов $B(i)$, для которого рассматриваемый ниже метод может оказаться полезным. В частности, к такому типу относятся процессы с монотонно убывающими корреляционными функциями, например экспоненциально коррелированные процессы. Рассмотрим некоторые виды функций $\varphi(n, j, \gamma)$. Пусть

$$\varphi(n, j, \gamma) = \begin{cases} 1 & \text{при } n-j \geq \gamma; \\ 0 & \text{при } n-j < \gamma. \end{cases} \quad (3-14)$$

Такой выбор функции веса соответствует методу «скользящего интервала» [Л. 34]. С вычислительной точки зрения более удобно выбрать

$$\varphi(n, j, \gamma) = \gamma^{n-j}, \quad \text{где } 0 < \gamma \leq 1. \quad (3-15)$$

В случае (3-14) и (3-15) наилучший выбор функции веса сводится к определению оптимального значения параметра γ , обеспечивающего минимальную величину суммы квадратов ошибок прогнозирования (3-6).

Как отмечалось выше, естественной формой записи алгоритмов построения математического описания дрейфующих объектов являются рекуррентные выражения, связывающие старые оценки параметров модели с вновь поступившей информацией. Общая запись рекуррентного алгоритма определяется выражением (3-5).

Определим такой алгоритм для метода текущего регрессионного анализа [Л. 37], для чего введем следующие обозначения:

$$X_i = \begin{vmatrix} x_1(1) & x_2(1) & \dots & x_h(1) \\ \dots & \dots & \dots & \dots \\ x_1(i) & x_2(i) & \dots & x_h(i) \end{vmatrix}; \quad (3-16)$$

$$Y_i = \begin{vmatrix} y(1) \\ y(2) \\ \dots \\ y(i) \end{vmatrix}; \quad (3-17)$$

$$\Psi_i = \begin{vmatrix} \psi(i, 1, \gamma) & & 0 \\ & \psi(i, 2, \gamma) & \\ \dots & \dots & \dots \\ 0 & & \psi(i, i, \gamma) \end{vmatrix}; \quad (3-18)$$

$$\widehat{B}(i+\alpha) = \begin{vmatrix} \widehat{\beta}_1(i+\alpha) \\ \widehat{\beta}_2(i+\alpha) \\ \dots \\ \widehat{\beta}_n(i+\alpha) \end{vmatrix}. \quad (3-19)$$

Тогда величина L в (3-13) для выборки объема i может быть записана следующим образом:

$$L = (Y_i - X_i \widehat{B}(i+\alpha))^T \Psi_i (Y_i - X_i \widehat{B}(i+\alpha)). \quad (3-20)$$

Минимизация L по $\widehat{B}(i+\alpha)$ приводит к системе нормальных уравнений вида (см. § 1-1)

$$X_i^T \Psi_i X_i \widehat{B}(i+\alpha) = X_i^T \Psi_i Y_i, \quad (3-21)$$

откуда

$$\widehat{B}(i+\alpha) = (X_i^T \Psi_i X_i)^{-1} X_i^T \Psi_i Y_i. \quad (3-22)$$

Заметим, что если Ψ_i — единичная матрица, то (3-22) совпадает с (1-13).

Допустим, что после определения оценки $\widehat{B}(i+\alpha)$ появилась дополнительная информация $x(i+1)$, $y(i+1)$, используя которую необходимо подсчитать $\widehat{B}(i+\alpha+1)$. Для этого можно вновь воспользоваться формулой (3-22), изменив входящие в нее матрицы. Наша цель состоит в построении рекуррентного алгоритма вычисления $\widehat{B}(i+\alpha+1)$, использующего уже подсчитанные величины $\widehat{B}(i+\alpha)$. Пусть функция веса $\varphi(i, j, \gamma)$ выбрана в виде (3-15). Тогда

$$\begin{aligned} \widehat{B}(i+\alpha+1) &= \|X_{i+1}^T \Psi_{i+1} X_{i+1}\|^{-1} X_{i+1}^T \Psi_{i+1} Y_{i+1} = \\ &= [\gamma X_i^T \Psi_i X_i + x(i+1)x^T(i+1)]^{-1} [\gamma X_i^T \Psi_i Y_i + x(i+1)y(i+1)]. \end{aligned} \quad (3-23)$$

Запись (3-23) совершенно очевидна. Воспользуемся известной формулой обращения матриц [Л. 6, 37]

$$(P + xx^T)^{-1} = P^{-1} - \frac{P^{-1}xx^TP^{-1}}{x^TP^{-1}x + 1}. \quad (3-24)$$

Здесь P — квадратная матрица ($k \times k$); x — вектор-столбец размерности k . Применяя формулу (3-24) к первому сомножителю (3-23) и обозначая $(X_i^T \Psi_i X_i)^{-1} = C_i$, находим:

$$\begin{aligned} \widehat{B}(i+\alpha+1) &= \left[\frac{1}{\gamma} C_i - \frac{1}{\gamma^2} \frac{C_i x(i+1) x^T(i+1) C_i}{\frac{1}{\gamma} x^T(i+1) C_i x(i+1) + 1} \right] \times \\ &\times [\gamma X_i^T \Psi_i X_i + x(i+1)y(i+1)] = \widehat{B}(i+\alpha) + \frac{1}{\gamma} C_i x(i+1) y(i+1) - \\ &- \frac{1}{\gamma} \frac{C_i x(i+1) x^T(i+1) C_i}{x^T(i+1) C_i x(i+1) + 1} [\gamma X_i^T \Psi_i X_i + x(i+1)y(i+1)]. \end{aligned} \quad (3-25)$$

Проведем несложные преобразования выражения (3-25):

$$\begin{aligned} \widehat{B}(i + \alpha + 1) &= \widehat{B}(i + \alpha) + \frac{1}{\gamma} C_i x(i + 1) \times \\ &\times \left[1 - \frac{x^T(i + 1) C_i x(i + 1)}{x^T(i + 1) C_i x(i + 1) + \gamma} \right] y(i + 1) - \\ &- C_i x(i + 1) \frac{x^T(i + 1) \widehat{B}(i + \alpha)}{x^T(i + 1) C_i x(i + 1) + \gamma}. \end{aligned} \quad (3-26)$$

Произведя вычитание в (3-26) в квадратной скобке, окончательно имеем:

$$\begin{aligned} \widehat{B}(i + \alpha + 1) &= \widehat{B}(i + \alpha) + \frac{C_i x(i + 1)}{x^T(i + 1) C_i x(i + 1) + \gamma} [y(i + 1) - \\ &- x^T(i + 1) \widehat{B}(i + \alpha)], \end{aligned} \quad (3-27)$$

где

$$C_i = (X_i^T \Psi_i X_i)^{-1} = \frac{1}{\gamma} \left[C_{i-1} - \frac{C_{i-1} x(i) x^T(i) C_{i-1}}{x^T(i) C_{i-1} x(i) + \gamma} \right]. \quad (3-28)$$

Выражение (3-27) дает нам в окончательной форме искомый рекуррентный алгоритм. Выражение (3-28) представляет собой рекуррентный алгоритм вычисления обратной матрицы при добавлении новых наблюдений.

Величина параметра γ в (3-27) и (3-28) определяет скорость изменения функции веса. Если исходить из того, что вес наблюдений убывает с удалением от текущего момента i , то величину γ следует выбирать от δ до 1, где δ — малая положительная величина. Значению $\gamma=1$ соответствует случай равных весов всех наблюдений выборки. Выражения (3-27) и (3-28) определяют при этом рекуррентный алгоритм метода наименьших квадратов, который может быть использован для решения задач гл. 1. При использовании алгоритма (3-27) для вычисления прогнозируемых значений коэффициентов $\widehat{B}(n+\alpha)$ по выборке из n наблюдений необходимо, помимо начальных значений оценок $\widehat{B}(\alpha)$ и величины γ , выбрать еще и начальное значение матрицы C_0 (или обратной ей C_0^{-1}). В соответствии

с соображениями, изложенными в гл. 2, принимаем величину $\widehat{B}(\alpha)$ равной оценке наименьших квадратов по имеющейся выборке, т. е.

$$\widehat{B}(\alpha) = (X_n^T X_n)^{-1} X_n^T Y_n. \quad (3-29)$$

Выбор начального значения матрицы C_0 (или C_0^{-1}) несуществен с точки зрения состоятельности оценки $\widehat{\gamma}_{\text{опт}}$, однако оказывает влияние на время установления режима слежения. Элементы этой матрицы поэтому желательно выбрать близкими к значениям в установившемся режиме слежения. Однако при $n \rightarrow \infty$ значение $M(C_n^{-1})$ будет зависеть от γ . Для того чтобы исключить эту зависимость,

выберем функцию веса в такой форме:

$$\varphi(i, j, \gamma) = (1-\gamma) \gamma^{|i-j|}. \quad (3-30)$$

Тогда в установившемся режиме слежения имеем в случае стационарного процесса $\mathbf{x}(i)$:

$$\lim_{i \rightarrow \infty} M[C_i^{-1}] = M[\mathbf{x}(i) \mathbf{x}^T(i)]. \quad (3-31)$$

Оценкой последнего выражения является матрица $\frac{1}{n} \mathbf{X}_n^T \mathbf{X}_n$, которую можно подсчитать по имеющейся выборке. Поэтому целесообразно следующим образом выбрать матрицу $\mathbf{C}_0$:

$$\mathbf{C}_0 = n (\mathbf{X}_n^T \mathbf{X}_n)^{-1}. \quad (3-32)$$

При использовании функции веса вида (3-30) рекуррентные выражения (3-27) и (3-28) примут следующий вид:

$$\begin{aligned} \mathbf{B}(i + \alpha + 1) = & \widehat{\mathbf{B}}(i + \alpha) + \frac{\mathbf{C}_i \mathbf{x}(i + 1)}{\mathbf{x}^T(i + 1) \mathbf{C}_i \mathbf{x}(i + 1) + \frac{\gamma}{1-\gamma}} [y(i + 1) - \\ & - \mathbf{x}^T(i + 1) \widehat{\mathbf{B}}(i + \alpha)]; \end{aligned} \quad (3-33)$$

$$\mathbf{C}_i = \frac{1}{\gamma} \left[\mathbf{C}_{i-1} - \frac{\mathbf{C}_{i-1} \mathbf{x}(i) \mathbf{x}^T(i) \mathbf{C}_{i-1}}{\mathbf{x}^T(i) \mathbf{C}_{i-1} \mathbf{x}(i) + \frac{\gamma}{1-\gamma}} \right]. \quad (3-34)$$

3-3. АЛГОРИТМЫ СТОХАСТИЧЕСКОЙ АППРОКСИМАЦИИ

Алгоритм (3-27) можно записать в несколько более общей форме:

$$\begin{aligned} \widehat{\mathbf{B}}(i + \alpha + 1) = & \widehat{\mathbf{B}}(i + \alpha) + \Gamma(i + 1) \mathbf{x}(i + 1) [y(i + 1) - \\ & - \mathbf{x}^T(i + 1) \widehat{\mathbf{B}}(i + \alpha)]. \end{aligned} \quad (3-35)$$

Здесь $\Gamma(i+1)$ — некоторая матрица. В рассмотренном выше случае

$$\Gamma(i + 1) = \frac{\mathbf{C}_i}{\mathbf{x}^T(i + 1) \mathbf{C}_i \mathbf{x}(i + 1) + \gamma}. \quad (3-36)$$

Структура рекуррентного алгоритма вычисления оценок коэффициентов модели (3-35) совпадает со структурой алгоритмов метода стохастической аппроксимации [Л. 32]. По существу выражение (3-35) определяет перемещение точки $\widehat{\mathbf{B}}(i+\alpha)$ в момент $(i+1)$ в пространстве коэффициентов. Это перемещение осуществляется в таком направлении, чтобы величина квадрата ошибки предсказания выхода в момент $(i+1)$, а именно: $[y(i+1) - \mathbf{B}^T(i+\alpha) \mathbf{x}(i+1)]^2$, уменьшалась. Шаг перемещения определяется матрицей $\Gamma(i+1)$. Выражения (3-28)

и (3-36) дают, например, величину шага для метода текущего регрессионного анализа, а при $\gamma=1$ для метода множественной регрессии.

Здесь стоит обратить внимание на следующее. Движение в пространстве коэффициентов осуществляется в некоторый момент $(i+1)$ при использовании алгоритма (3-35) в таком направлении, чтобы уменьшить ошибку предсказания выхода в тот же момент $(i+1)$. Однако можно использовать алгоритм (3-35) для случаев, когда $\alpha > 0$. При этом оценка коэффициентов, подсчитанная в момент $(i+1)$, предназначена для прогнозирования коэффициентов в момент времени $(i+\alpha+1)$. Совершенно очевидно, что получить положительный эффект в этом случае возможно только тогда, когда действительные значения коэффициентов в моменты $(i+1)$ и $(i+\alpha+1)$ оказываются близкими, т. е. когда дрейф коэффициентов медленный.

В [Л. 32, 38] изучались закономерности, которым должны удовлетворять величины шагов алгоритма стохастической аппроксимации вида (3-35) в задачах последовательной оценки неизменяющихся параметров моделей объектов, а также рассматривалась модификация процедуры стохастической аппроксимации в условиях линейного дрейфа оцениваемых параметров [Л. 49]. Попытаемся выявить некоторые особенности выбора матрицы $\Gamma(i+1)$ в алгоритме (3-35), когда коэффициенты модели изменяются случайным образом.

Прежде всего, если оцениваемые параметры не изменяются во времени, а используемый алгоритм позволяет получить их состоятельные оценки, то после некоторого, возможно достаточно большого числа шагов n оценка $\hat{B}(n)$ будет близка к действительному значению, и движение в пространстве коэффициентов должно прекратиться. При наличии ошибок эксперимента ($\sigma^2 \neq 0$) оценки коэффициентов перестанут изменяться только тогда, когда величина шага станет равной нулю. Это значит, что в случае последовательной оценки неизменяющихся параметров величина шага перемещения точки в пространстве коэффициентов должна стремиться к нулю при увеличении числа сделанных шагов. Действительно, нетрудно убедиться [Л. 37], что при использовании рекуррентных формул (3-27) и (3-28) при $\gamma=1$ $\lim_{n \rightarrow \infty} \Gamma(n) = 0$, где $\Gamma(n)$ определяется формулой (3-36).

В отличие от случая неизменяющихся параметров при оценке дрейфующих коэффициентов движение точки $\hat{B}(n)$ в пространстве коэффициентов не должно прекращаться при n , стремящемся к бесконечности; точка $\hat{B}(n)$ должна постоянно перемещаться вслед за действительным значением коэффициентов. Таким образом, при оценке дрейфующих параметров

$$\lim_{i \rightarrow \infty} \Gamma(n) \neq 0. \quad (3-37)$$

В простейшем случае условию (3-37) удовлетворяет матрица $\Gamma(n)$ вида

$$\Gamma(n) = \gamma I_k, \quad (3-38)$$

где γ — константа, а I_k — единичная матрица. При этом если значение в (3-38) выбрать из диапазона $0 < \gamma < \gamma_0$, то с помощью алгоритма (3-35) можно осуществить слежение за дрейфующими коэффициентами модели [Л. 48]. Предельное значение $\gamma = \gamma_0$ зависит от моментов процесса $x(i)$ и определяет границу устойчивости процесса слежения [Л. 32, 48].

В [Л. 39] для последовательной оценки коэффициентов модели объекта использован алгоритм вида [3-35], где матрица $\Gamma(i+1)$ выбрана следующим образом:

$$\Gamma(i+1) = \frac{1}{\mathbf{x}^T(i+1) \mathbf{x}(i+1) + \gamma} \mathbf{I}_n, \quad (3-39)$$

а γ может принимать только неотрицательные значения.

Заметим, что в случаях, когда при достаточно больших i матрица $\mathbf{C}_i$ в (3-36) оказывается близкой к диагональной с равными элементами на главной диагонали, выражения (3-39) и (3-36) практически совпадают. Подобная ситуация имеет место, например, когда значение γ в (3-36) близко к единице, а компоненты вектора $\mathbf{x}(i)$ независимы, имеют равные дисперсии и нулевые математические ожидания.

Величины параметра γ в (3-38) и (3-39) неопределенны, и для получения их значений в случае решения задачи прогнозирования по конечной выборке наблюдений необходимо воспользоваться критерием (3-6). При известной корреляционной функции процесса $B(i)$ и дисперсии ошибки σ^2 оптимальные значения параметра γ могут быть найдены аналитическим путем.

3-4. ОЦЕНКА ПАРАМЕТРА АЛГОРИТМА γ ПО КОНЕЧНОЙ ВЫБОРКЕ НАБЛЮДЕНИЙ

Рассмотрим алгоритм нахождения оценки параметра γ в соответствии с критерием (3-6) по конечной выборке наблюдений для случая, когда для вычисления оценок коэффициентов используется рекуррентная формула (3-35), а матрица $\Gamma(i+1)$ выбрана в виде (3-39). При этом алгоритм имеет следующий вид:

$$\hat{B}(i+\alpha) = \hat{B}(i+\alpha+1) + \frac{\mathbf{x}(i)}{\mathbf{x}^T(i) \mathbf{x}(i) + \gamma} [y(i) - \hat{B}^T(i+\alpha-1) \mathbf{x}(i)]. \quad (3-40)$$

Пусть количество наблюдений в выборке равно $n \gg \alpha$. В качестве начального значения коэффициентов $\hat{B}(\alpha)$ примем оценку наименьших квадратов по выборке, т. е.

$$\hat{B}(\alpha) = (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T \mathbf{Y}_n. \quad (3-41)$$

По заданной выборке, используя алгоритм (3-40) при $\hat{B}(\alpha)$, подсчитанном по формуле (3-41), и при любом значении γ можно найти последовательность оценок коэффициентов $\hat{B}(\alpha+1)$, $\hat{B}(\alpha+2)$, ..., $\hat{B}(\alpha+n)$. Определим оценку $\gamma_{\text{опт}}$ оптимального значения параметра γ из условия

$$S_\alpha(\hat{\gamma}_{\text{опт}}) = \min_{\gamma \geq 0} S_\alpha(\gamma) = \min_{\gamma \geq 0} \sum_{j=\alpha}^n [y(j) - \hat{B}^T(j) \mathbf{x}(j)]^2, \quad (3-42)$$

где значения $\hat{B}(j)$ находятся путем применения рекуррентного алгоритма (3-40), а $\hat{\gamma}_{\text{опт}}$ принимает только неотрицательные значения.

Заметим, что случаю отсутствия дрейфа соответствует бесконечно большое значение $\hat{\gamma}_{\text{опт}}$. При этом

$$\hat{B}(n + \alpha) = \hat{B}(\alpha) = (X_n^T X_n)^{-1} X_n^T Y_n, \quad (3-43)$$

т. е. предсказанное значение вектора коэффициентов оказывается равным оценке, полученной по выборке методом наименьших квадратов.

Как видно из (3-40), величина коррекции j -го коэффициента в некоторый момент времени зависит от отношения j -й проекции вектора $x(i)$ к квадрату его длины [квадрат длины вектора равен скалярному произведению $x^T(i)x(i)$]. При этом если средние значения переменных неодинаковы, то неодинаково быстро будет происходить движение в направлении различных координат в пространстве коэффициентов. Это в свою очередь вызовет ухудшение процесса слежения за дрейфующими параметрами объекта. Более того, в [Л. 39, 40] показано, что наилучшие результаты с точки зрения скорости сходимости (или скорости уменьшения ошибки определения коэффициентов) алгоритмов рассматриваемого нами типа достигаются в случае, когда все переменные x_j имеют одинаковые дисперсии и нулевые математические ожидания. Поэтому при решении рассматриваемой задачи целесообразно использовать переменные в стандартизованном масштабе.

Остановимся на способе поиска значения $\hat{\gamma}_{\text{опт}}$, обеспечивающего минимум величины $S_\alpha(\gamma)$. Эта задача является одномерной задачей поиска минимума функции при отсутствии помех. Некоторые осложнения возникают из-за того, что для выборки объема n наивысшая степень, с которой γ входит в $S_\alpha(\gamma)$, равна $2n$. Это становится очевидным, если выразить оценку $\hat{B}(n + \alpha)$ через все имеющиеся наблюдения, как это сделано в выражении (3-59). При этом функция $S_\alpha(\gamma)$ может иметь несколько экстремальных точек в диапазоне изменения от 0 до ∞ .

При поиске $\gamma_{\text{опт}}$ нас интересует глобальный минимум функции $S_\alpha(\gamma)$, который может, вообще говоря, достигаться и в граничных точках, т. е. при $\gamma = 0$ или при $\gamma = \infty$.

Учитывая сравнительно малый объем вычислений, требуемый для нахождения значения $S_\alpha(\gamma)$ при любой величине γ , можно рекомендовать следующую процедуру поиска $\hat{\gamma}_{\text{опт}}$.

1. Выбираем значение $\gamma_{\text{макс}}$, такое, что $S_\alpha(\gamma_{\text{макс}}) \approx S_{\text{ост}}$, где $S_{\text{ост}}$ — остаточная сумма квадратов. Если переменные x_j представлены в стандартизованном масштабе, то $\gamma_{\text{макс}} \gg k$, где k — число оцениваемых коэффициентов модели (в общем случае в это число входит и свободный член уравнения, для которого фиктивная переменная тождественно равна единице).

2. Строим функцию $S_\alpha(\gamma)$, производя сканирование по γ с выбранным шагом в интервале значений γ от нуля до $\gamma_{\text{макс}}$.

3. При необходимости уточнить значение $\hat{\gamma}_{\text{опт}}$ производим дополнительные расчеты внутри интервала неопределенности положе-

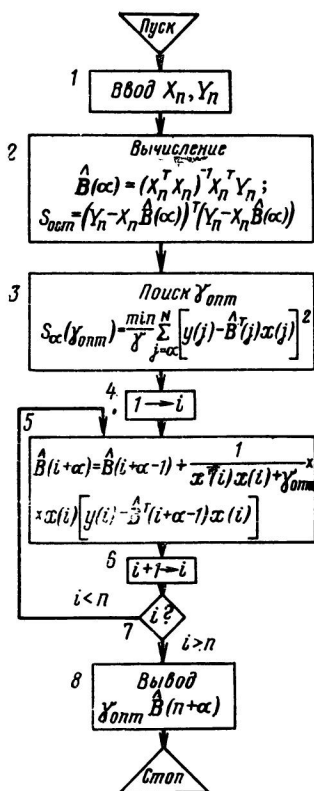


Рис. 3-1. Структурная схема алгоритма прогнозирования параметров дрейфующего объекта по конечной выборке наблюдений.

вания параметров дрейфующего объекта при конечной выборке наблюдений. При изображении структурных схем алгоритмов применена обычно используемая символика. Здесь треугольниками обозначены операторы пуска и останова; ромбами — логические операторы; прямоугольниками — арифметические операторы и крупные блоки, содержащие различные операторы; связи между блоками показаны линиями. Оператор присвоения, который часто встречается при построении подобных структурных схем, обозначен стрелкой. Например, запись $i+1 \rightarrow i$ следует понимать так: переменной i присваивается старое значение, увеличенное на единицу.

На рис. 3-1 приведена общая структурная схема алгоритма решения задачи с помощью рассмотренного метода. В блоке 2 осуществляется вычисление начальных приближений для коэффициен-

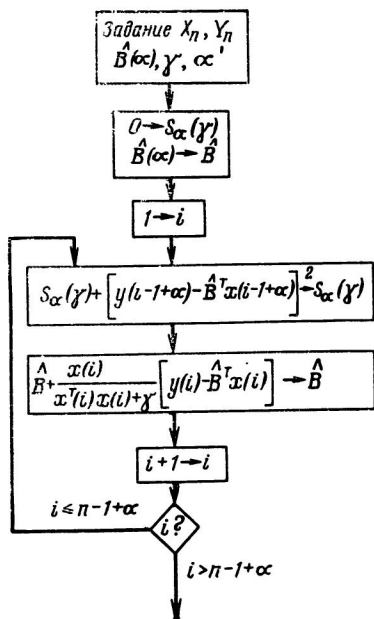


Рис. 3-2. Структурная схема алгоритма вычисления $S_\alpha(\gamma)$.

ния глобального экстремума, полученного в предыдущем пункте. Для этой цели могут быть использованы такие эффективные методы локализации экстремума, как метод Фибоначчи и метод «золотого сечения», подробно рассмотренные в [Л. 41]. На рис. 3-1 и 3-2 приведены структурные схемы алгоритмов решения задачи прогнозирования

тов $\hat{B}(\alpha)$ и остаточной суммы квадратов. С вычислительной точки зрения реализуемые здесь методы не отличаются от рассмотренных в § 1-1. В блоке 3 ищется $\hat{\gamma}_{\text{опт}}$.

Для этого необходимо вычислить значения $S_{\alpha}(\gamma)$ для заданного числа значений γ , последовательный выбор которых определяется используемым методом поиска. В блоках 5—7 с использованием найденного $\hat{\gamma}_{\text{опт}}$ вычисляется по рекуррентным формулам $\hat{B}(n+\alpha)$. Программа заканчивается выбором найденных значений $\hat{\gamma}_{\text{опт}}$ и $\hat{B}(n+\alpha)$ (блок 8).

На рис. 3-2 приведена структурная схема вычисления $S_{\alpha}(\gamma)$, дополняющая схему рис. 3-1. Здесь использован алгоритм (3-40). В качестве оценки вектора коэффициентов для момента α принято значение $\hat{B}(\alpha)$, вычисленное в блоке 2 схемы рис. 3-1.

В заключение рассмотрим примеры применения изложенного алгоритма.

Пример 3-1. Модель исследуемого объекта имеет вид:

$$u(i) = \beta(i) x(i). \quad (3-44)$$

Через равные интервалы времени проделано $n=10$ экспериментов, результаты которых сведены в табл. 3-1.

Таблица 3-1

i	$x(i)$	$y(i)$	i	$x(i)$	$y(i)$
1	-0,74	-0,232	6	-1,667	-0,373
2	1,667	0,28	7	0,623	-0,160
3	-1,83	0,063	8	1,167	-0,338
4	-0,778	-0,214	9	0,662	-0,551
5	1,945	0,058	10	-0,74	0,211

Необходимо оценить значение коэффициента в момент $(n+1)$. Воспользуемся алгоритмом, структурная схема которого изображена на рис. 3-1. Начальное значение $\hat{\beta}(\alpha) = \hat{\beta}(1)$ в данном случае оказывается равным $\hat{\beta}(1) = -0,05$. Остаточная сумма квадратов $S_{\text{ост}} = 0,77$.

Задаемся значением $\gamma_{\text{макс}} = 10$; величина $S_{\alpha}(10) = 0,776$. Вычисляем функцию $S_{\alpha}(\gamma)$ в интервале значений от 0 до 10 с шагом изменения $\gamma = 1$ (рис. 3-3).

Производя дополнительные расчеты при изменении γ от 0,5 до 1,5, получаем $\hat{\gamma}_{\text{опт}} = 1,2$. Пользуясь значением $\hat{\gamma}_{\text{опт}}$, находим $\hat{\beta}(n+1) = -0,36$.

На рис. 3-4 показано, как изменяется оценка $\hat{\beta}(i)$ при $i=1, 2, \dots, 11$, когда $\hat{\gamma}_{\text{опт}} = 1,2$ (кривая 1). Для сравнения там же показан процесс изменения действительного значения коэффициентов $\beta(i)$

(кривая 2). Понятно, что в практических задачах кривая 2 нам неизвестна в отличие от рассмотренной в этом примере модельной задачи.

Пример 3-2. При исследовании процесса формирования и сушки ленты синтетического каучука [Л. 42] было установлено, что для

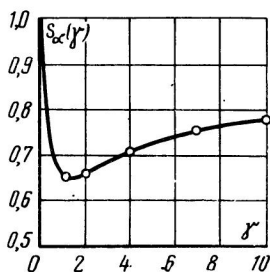


Рис. 3-3. Зависимость $S_\alpha(\gamma)$ для примера 3-1.

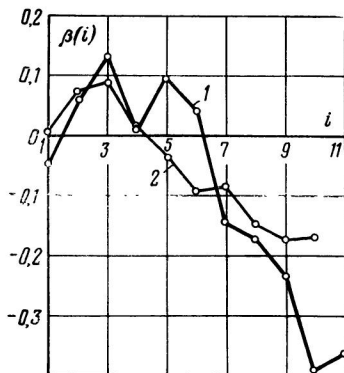


Рис. 3-4. Прогнозирование коэффициента $\beta(i)$ для примера 3-1

при $\gamma = \hat{\gamma}_{\text{опт}} = 1,2$.

1 — оценка $\hat{\beta}(i)$; 2 — действительное значение $\beta(i)$.

выбора режима сушки достаточно найти значение условной толщины ленты y . Кроме того, оптимальные режимы в смысле максимальной скорости сушки при сохранении механических свойств ленты зависят от следующих параметров, характеризующих каучук: дефо латекса (x_2), содержание масла в каучуке (x_3), РН в аппарате (x_4). Предполагалось, что зависимость между условной толщиной ленты в оптимальных режимах сушки и перечисленными параметрами линейна и медленно дрейфует во времени из-за изменения механического состояния сита лентоотливочной машины. Для проверки наличия связи между перечисленными параметрами была извлечена выборка, содержащая $n=38$ экспериментальных точек. Считалось, что при проведении этого эксперимента состояние сита лентоотливочной машины не изменялось, т. е. собранный экспериментальный материал соответствует случаю отсутствия дрейфа. По выборке была построена модель вида Δ :

$$y = 0,362 x_2 + 0,35 x_3 - 0,465 x_4. \quad (3-45)$$

В уравнении (3-45) все переменные представлены в стандартном масштабе. Остаточная сумма квадратов $S_{\text{ост}}=19,85$.

Проверим с помощью алгоритма, приведенного на рис. 3-1, правильность предположения об отсутствии дрейфа. На рис. 3-5 изобра-

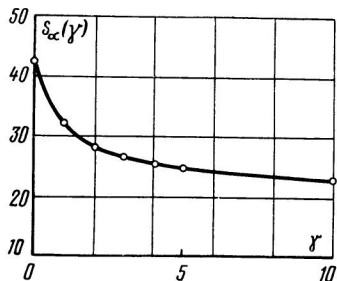


Рис. 3-5. Зависимость $S_\alpha(\gamma)$ для примера 3-2.

жена зависимость $S(\gamma)$, построенная путем использования алгоритма рис. 3-1. Как видно из рис. 3-5 $S_1(\gamma)$ монотонно убывает при увеличении γ . Этот расчет подтверждает правильность предположения о неизменности состояния сита лентоотливочной машины за время сбора экспериментальных данных.

3-5. ОПТИМАЛЬНЫЙ ВЫБОР ПАРАМЕТРА АЛГОРИТМА γ В СЛУЧАЕ ИЗВЕСТНОЙ КОРРЕЛЯЦИОННОЙ ФУНКЦИИ ПРОЦЕССА $B(i)$ И ДИСПЕРСИИ σ^2

Помимо постановки задачи построения модели по конечной выборке наблюдений, которая рассматривалась в предыдущих разделах, возможна еще и иная постановка. Пусть наблюдение за входом и выходом объекта осуществляется в течение бесконечно большого интервала времени, так что к моменту n , когда нужно определить оценку $\hat{B}(n+\alpha)$, можно было бы располагать бесконечно большим числом наблюдений. Для решения этой задачи методом, изложенным выше, потребовалось бы запомнить весь массив наблюдений. В этом случае удобнее поиск $\gamma_{\text{опт}}$ производить последовательно, находя оценку $S_\alpha(\gamma)$ для различных значений γ по разным интервалам наблюдений конечной длины. Такой способ решения задачи настройки алгоритма относится к методам экспериментальной оптимизации или оптимизации при наличии помех, рассматриваемым, например, в [Л. 41, 43].

До сих пор в этой главе рассматривались задачи, где действительный характер процесса $B(i)$ и дисперсия ошибки σ^2 были неизвестны. Если такая информация имела бы, то можно было бы получить строго оптимальный алгоритм решения задачи, структура которого, вероятно, отличалась бы от принятой нами. Задачи такого типа рассматривались ранее, например, в [Л. 30, 37].

Возможна, однако, и такая постановка задачи: пусть нам известны некоторые свойства процесса $B(i)$ и дисперсия σ^2 , выбран критерий оптимальности решения, и в то же время задана с точностью до некоторого параметра γ структура алгоритма решения. В этом случае наилучшее в смысле заданного критерия значение параметра γ может быть найдено аналитически [Л. 44].

Рассмотрим решение задачи определения значения $\gamma_{\text{опт}}$, когда структура алгоритма выбрана в виде (3-35), а $\Gamma(n+1)$ в виде (3-38), т. е.

$$\hat{B}(n+\alpha) = \hat{B}(n+\alpha-1) + \gamma x(n) [y(n) - \hat{B}^T(n+\alpha-1) x(\Delta n)]. \quad (3-46)$$

Пусть модель объекта имеет вид:

$$u(n) = B^T(n) x(n), \quad (3-47)$$

где n — дискретные моменты времени, в которых происходит съем информации со входа и выхода объекта. Будем полагать, как и выше, $n=1, 2, \dots$. Пусть $x(i)$ — (k) -мерный вектор с координатами $x_1(i), x_2(i), \dots, x_k(i)$.

Примем, что всякая переменная $x_j(i)$ представляет собой независимую стационарную случайную последовательность с нулевым математическим ожиданием и дисперсией $\sigma_{x_j}^2 = 1$, т. е.

$$M[x_j(i) x_j(i + \tau)] = \begin{cases} 0 & \text{при } \tau \neq 0; \\ 1 & \text{при } \tau = 0. \end{cases} \quad (3-48)$$

Кроме того, будем считать, что

$$M[x_j(i) x_\mu(i + \tau)] = 0 \text{ при } j \neq \mu, \quad (3-49)$$

т. е. ковариационная матрица вектора $\mathbf{x}(i)$ в каждый момент i имеет вид:

$$M[\mathbf{x}(i) \mathbf{x}^T(i)] = \mathbf{I}_k, \quad (3-50)$$

где $\mathbf{I}_k$ — единичная матрица размерности k . В дальнейшем нам понадобятся еще четвертые моменты распределения переменных x_j :

$$M[x_j^4] = \mu_j. \quad (3-51)$$

Будем считать, что переменные $x_1, \dots, x_k$ имеют одинаковые четвертые моменты, т. е.

$$\mu_1 = \mu_2 = \dots = \mu_k = \mu. \quad (3-52)$$

Выбор ковариационной матрицы вектора $\mathbf{x}(i)$ в виде единичной матрицы $\mathbf{I}_k$ в данной задаче не является ограничением, так как линейным преобразованием исходных переменных всегда можно свести задачу к случаю, когда ковариационная матрица входных переменных имеет вид (3-50) (ортогональное преобразование ковариационной матрицы [Л. 6, §5]). С другой стороны, выбор ковариационной матрицы вектора $\mathbf{x}(i)$ в виде единичной матрицы влечет существенное упрощение приводимых ниже выкладок.

Особенностью выбранной нами модели (3-47) является то, что она не содержит свободного члена. Практически это означает, что величина свободного члена в каждый момент i известна и исключена из наблюдаемых результатов. Рассмотрение случая, когда имеется дрейф всех коэффициентов модели, включая свободный член, принципиальных затруднений не вызывает, однако связано с более громоздкими выкладками и поэтому здесь не приводится.

Положим далее, что $\mathbf{B}(i) = (k)$ -мерный стационарный случайный процесс, причем известны корреляционные функции

$$\begin{aligned} R_{ii}(\tau) &= M\{[\beta_i(j) - m_{\beta_i}][\beta_i(j + \tau) - m_{\beta_i}]\} = \\ &= \sigma_{ii}^2 f_{ii}(\tau); \quad i = 1, 2, \dots, k; \quad \tau = 0, 1, 2, \dots \end{aligned} \quad (3-53)$$

Здесь через m_{β_i} обозначено математическое ожидание β_i , знание которого необязательно. Кроме того, будем считать, что $\sigma_{ii}^2 < \infty$ и $\lim_{\tau \rightarrow \infty} f_{ii}(\tau) = 0$.

Таким образом, рассматривается случай, когда дрейф описывается стационарным случайным процессом, а вся исходная информация заключена в корреляционных функциях $R_{ii}(\tau); i = 1, \dots, k$

этого процесса. В качестве критерия для выбора оптимального значения γ в алгоритме (3-46) используем математическое ожидание квадрата ошибки прогнозирования выхода объекта в момент $(n+\alpha)$ при n , стремящемся к бесконечности. Таким образом, критерий представляет собой среднее значение величины квадрата ошибки предсказания выходной переменной объекта в установившемся режиме слежения. При этом $\gamma_{\text{опт}}$ может быть найдено из условия

$$\min_{\gamma} Q(\gamma) = \min_{\gamma} \lim_{n \rightarrow \infty} M \{ [u(n+\alpha) - \widehat{B}^T(n+\alpha) x(n+\alpha)]^2 \}. \quad (3-54)$$

Как и ранее, будем полагать, что

$$y(i) = u(i) + e(i), \quad (3-55)$$

где $e(i)$ — последовательность независимых случайных величин с нулевым математическим ожиданием и дисперсией σ^2 , которую в данном случае считаем известной:

$$M[e(i)e(j)] = \begin{cases} 0 & \text{при } i \neq j; \\ \sigma^2 & \text{при } i = j. \end{cases} \quad (3-56)$$

Кроме того, предполагаем, что случайные процессы $x(i)$, $B(i)$ и $e(i)$ взаимонезависимы, т. е.

$$\left. \begin{aligned} M[x_j(i) \beta_v(\mu)] &= M[x_j(i)] M[\beta_v(\mu)]; \\ M[x_j(i) e(\lambda)] &= M[\beta_v(\mu) e(\lambda)] = 0 \end{aligned} \right\} \quad (3-57)$$

при $j, v = 1, 2, \dots, k$ и для любых i, μ, λ .

При этих условиях найдем выражение для $Q(\gamma)$ при $\alpha > 0$. Предварительно перепишем (3-46) следующим образом:

$$\begin{aligned} \widehat{B}(n+\alpha) &= \widehat{B}(n+\alpha-1) + \gamma x(n) \times \\ &\times \{ [B(n) - \widehat{B}(n+\alpha-1)]^T x(n) + e(n) \} = \\ &= (I_k - \gamma x(n) x^T(n)) \widehat{B}(n+\alpha-1) + \\ &+ \gamma x(n) x^T(n) B(n) + \gamma e(n) x(n). \end{aligned} \quad (3-58)$$

Используя (3-58), выразим $\widehat{B}(n+\alpha)$ через начальную оценку $\widehat{B}(\alpha)$ и значения $x(i)$, $B(i)$ и $e(i)$ при $i = 1, 2, \dots, n$:

$$\begin{aligned} \widehat{B}(n+\alpha) &= \prod_{i=1}^n (I_k - \gamma x(i) x^T(i)) \widehat{B}(\alpha) + \gamma \sum_{i=1}^n \prod_{j=i+1}^n (I_k - \\ &- \gamma x(j) x^T(j)) x(i) x^T(i) B(i) + \gamma \sum_{i=1}^n e(i) \times \\ &\times \prod_{j=i+1}^n (I_k - \gamma x(j) x^T(j)) x(i). \end{aligned} \quad (3-59)$$

Учитывая условие независимости случайных процессов $\mathbf{x}(i)$, $\mathbf{B}(i)$ и $e(i)$, найдем

$$M[\widehat{\mathbf{B}}(n+\alpha)] = (1-\gamma)^n \widehat{\mathbf{B}}(\alpha) + M[\mathbf{B}(n)] \gamma \sum_{i=1}^n (1-\gamma)^{n-i}. \quad (3-60)$$

$$\text{Из (3-60) следует, что при } |1-\gamma| < 1, \text{ т. е. при } 0 < \gamma < 2, \quad (3-61)$$

имеем

$$\lim_{n \rightarrow \infty} M[\widehat{\mathbf{B}}(n+\alpha)] = M[\mathbf{B}]. \quad (3-62)$$

Запишем теперь выражение для $Q(\gamma)$:

$$\begin{aligned} Q(\gamma) &= \lim_{n \rightarrow \infty} M\{[u(n+\alpha) - \widehat{\mathbf{B}}^T(n+\alpha) \mathbf{x}(n+\alpha)]^2\} = \\ &= \lim_{n \rightarrow \infty} M\{[\mathbf{B}(n+\alpha) - \widehat{\mathbf{B}}(n+\alpha)]^T \mathbf{x}(n+\alpha) \mathbf{x}^T(n+\alpha) [\mathbf{B}(n+\alpha) - \\ &\quad - \widehat{\mathbf{B}}(n+\alpha)]\} = \lim_{n \rightarrow \infty} M[\mathbf{B}^T(n+\alpha) \mathbf{B}(n+\alpha)] + \\ &\quad + \lim_{n \rightarrow \infty} M[\widehat{\mathbf{B}}^T(n+\alpha) \widehat{\mathbf{B}}(n+\alpha)] - \\ &\quad - 2 \lim_{n \rightarrow \infty} M[\mathbf{B}^T(n+\alpha) \widehat{\mathbf{B}}(n+\alpha)]. \end{aligned} \quad (3-63)$$

При заданной матрице (3-50), корреляционной функции (3-53) и условии независимости процессов $\mathbf{B}(n)$, $\mathbf{x}(n)$ и $e(n)$ нетрудно, используя выражение (3-59), найти

$$\begin{aligned} &\lim_{n \rightarrow \infty} M[\mathbf{B}^T(n+\alpha) \widehat{\mathbf{B}}(n+\alpha)] = \\ &= \sum_{i=1}^k \left[\sigma_{ii}^2 \gamma \sum_{j=0}^{\infty} (1-\gamma)^j f_{ii}(j+\alpha) + m_{\mathbf{B}_i}^2 \right]; \end{aligned} \quad (3-64)$$

$$\lim_{n \rightarrow \infty} M[\mathbf{B}^T(n+\alpha) \mathbf{B}(n+\alpha)] = \sum_{i=1}^k (\sigma_{ii}^2 + m_{\mathbf{B}_i}^2). \quad (3-65)$$

В выражении (3-63) осталось найти только второе слагаемое. Воспользуемся записью $\widehat{\mathbf{B}}(n+\alpha)$ в виде (3-46). Учитывая (3-50), (3-55) и условие независимости процессов $\mathbf{x}(i)$, $\mathbf{B}(i)$ и $e(i)$ (3-57), путем непосредственного перемножения получим:

$$\begin{aligned} M[\widehat{\mathbf{B}}^T(n+\alpha) \widehat{\mathbf{B}}(n+\alpha)] &= M[\widehat{\mathbf{B}}^T(n+\alpha-1) \widehat{\mathbf{B}}(n+\alpha-1)] - \\ &\quad - \gamma(2-\gamma\rho) M[\widehat{\mathbf{B}}^T(n+\alpha-1) \widehat{\mathbf{B}}(n+\alpha-1)] + \\ &\quad + \gamma^2 \rho M[\mathbf{B}^T(n) \mathbf{B}(n)] + 2\gamma(1-\gamma\rho) M \times \\ &\quad \times [\mathbf{B}^T(n) \widehat{\mathbf{B}}(n+\alpha-1)] + k\sigma^2\gamma^2. \end{aligned} \quad (3-66)$$

В выражении (3-66) через ρ обозначена следующая величина:

$$\rho = (k-1) + \mu, \quad (3-67)$$

где k — число входных переменных; μ представляет собой четвертый момент распределения переменных x_j ; $j=1, 2, \dots, k$. Величина ρ появляется при вычислении следующего выражения:

$$M[\mathbf{x}(n) \mathbf{x}^T(n) \mathbf{x}(n) \mathbf{x}^T(n)] = [(k-1) + \mu] \mathbf{I}_k = \rho \mathbf{I}_k. \quad (3-68)$$

В справедливости (3-68) легко убедиться, непосредственно производя умножение и имея в виду (3-50).

Так как процесс $\mathbf{B}(n)$ — стационарный, то воспроизводящий его процесс $\widehat{\mathbf{B}}(n)$ тоже является стационарным. Тогда для достаточно больших n (т. е. в установившемся режиме слежения) имеем:

$$M[\widehat{\mathbf{B}}^T(n+\alpha) \widehat{\mathbf{B}}(n+\alpha)] = M[\widehat{\mathbf{B}}^T(n+\alpha-1) \widehat{\mathbf{B}}(n+\alpha-1)]. \quad (3-69)$$

Заметим, что не всякие значения γ приводят к достижению установившегося режима слежения [Л. 32], т. е. не при всяком γ слежение будет устойчивым. Как показано в [Л. 48], в данном случае устойчивое слежение будет иметь место, если $\gamma < 2/\rho$.

Пользуясь (3-66) и (3-69), легко получить

$$\begin{aligned} \lim_{n \rightarrow \infty} M\{\widehat{\mathbf{B}}^T(n) \widehat{\mathbf{B}}(n)\} &= \frac{1}{2 - \gamma\rho} \{ \gamma \sigma^2 k + \gamma\rho \lim_{n \rightarrow \infty} M[\mathbf{B}^T(n) \mathbf{B}(n)] + \\ &+ 2(1 - \gamma\rho) \lim_{n \rightarrow \infty} M[\mathbf{B}^T(n) \widehat{\mathbf{B}}(n-1)] \}. \end{aligned} \quad (3-70)$$

Подставляя теперь (3-64), (3-65) в (3-70) и (3-63), после весьма несложных алгебраических преобразований получаем окончательный результат:

$$\begin{aligned} Q(\gamma) &= \frac{2}{2 - \gamma\rho} \left[\sum_{i=1}^k \sigma_{ii}^2 + \gamma(1 - \gamma\rho) \sum_{i=1}^k \sigma_{ii}^2 \sum_{j=0}^{\infty} (1 - \right. \\ &\left. - \gamma)^j f_{ii}(j+1) - 2(2 - \gamma\rho) \sum_{i=1}^k \sigma_{ii}^2 \sum_{j=0}^{\infty} (1 - \gamma)^j f_{ii}(j+\alpha) \right] + \frac{\gamma k \sigma^2}{2 - \gamma\rho}. \end{aligned} \quad (3-71)$$

Для нахождения $\gamma_{\text{опт}}$ необходимо определить минимум (3-71) по γ . Заметим, что величина среднего квадрата ошибки $Q(\gamma)$ не зависит от математического ожидания вектора коэффициентов модели $M[\mathbf{B}]$. Это связано с тем, что для процесса $\widehat{\mathbf{B}}(i)$ выполняется условие (3-62), т. е. математическое ожидание вектора $\widehat{\mathbf{B}}$ равно математическому ожиданию вектора $\mathbf{B}$.

Из (3-71) следует, что в случае, когда дрейф отсутствует (т. е. коэффициенты $\mathbf{B}$ постоянны и $\sigma_{ii}^2 = 0$ для $i=1, 2, \dots, k$), оптимальное значение $\gamma_{\text{опт}} = 0$. Этот результат становится понятным, если вспомнить, что $Q(\gamma)$ характеризует ошибку в установившемся режиме слежения. Действительно, при числе экспериментов, стремящемся к бесконечности, и отсутствии дрейфа оценки коэффициентов становятся равными действительным значениям, и изменение оценок не должно иметь места (т. е. γ должно быть равно 0). Заметим еще,

что $\gamma_{\text{опт}}$ стремится к нулю при возрастании n также в случае, когда корреляционная функция процесса $B(i)$ затухает слишком быстро, так что $R_{jj}(\tau) = 0$ при $\tau \geq 1$ для $j=1, 2, \dots, k$. При этом старые наблюдения не содержат никакой информации о значениях коэффициентов модели в последующие моменты времени, и наилучшие результаты дает прогнозирование по средним значениям коэффициентов модели.

Приведем выражение для среднего квадрата ошибки предсказания ($\alpha > 0$), когда корреляционные функции $R_{11}(\tau) = R_{22}(\tau) = \dots = R_{kk}(\tau) = \sigma_p^2 e^{-\lambda \tau}$, где $\lambda > 0$.

Выберем $\gamma < 2/\rho$. Заметим, что величина ρ больше единицы, т. е. при таком выборе γ всегда выполняется условие (3-61). Бесконечные суммы, входящие в (3-71), легко раскрываются по элементарным формулам сумм убывающих геометрических прогрессий:

$$\sum_{j=0}^{\infty} (1-\gamma)^j e^{-\lambda(j+\alpha)} = e^{-\lambda\alpha} \frac{1}{1 - e^{-\lambda}(1-\gamma)}. \quad (3-72)$$

После простых алгебраических преобразований получаем:

$$Q(\gamma) = k \left\{ 2\sigma_p^2 \left[\frac{1 - e^{-\lambda}}{(2-\gamma\rho) [1 - e^{-\lambda}(1-\gamma)]} + \frac{\gamma(e^{-\lambda} - e^{-\alpha\lambda})}{1 - e^{-\lambda}(1-\gamma)} \right] + \frac{\gamma\sigma^2}{2 - \gamma\rho} \right\}. \quad (3-73)$$

Обозначим через γ' величину

$$\gamma' = \frac{\gamma}{2/\rho}. \quad (3-74)$$

На рис. 3-6 построены кривые для $Q(\gamma')/k\sigma_p^2$ в функции от γ' при $\alpha = 1$ для случая, когда x_j , $j=1, 2, \dots, k$ равномерно распределены в диапазоне от $-\sqrt{3}$ до $+\sqrt{3}$. Нетрудно убедиться (см. пример 3-3), что при этом $\sigma_{x_j}^2 = 1$ и $\mu_j = 1,8$. Кривые на рис. 3-6 построены для различных значений k , λ и отношения σ^2/σ_p^2 .

Исследование выражения (3-73) позволяет сделать некоторые практически полезные рекомендации относительно выбора $\gamma'_{\text{опт}}$:

1. Максимальное значение $\gamma'_{\text{опт}}$ достигается при $\sigma^2 = 0$ и при $\alpha = 1$ равно:

$$(\gamma'_{\text{опт}})^* = 0,5 - \frac{\rho(1 - e^{-\lambda})}{4e^{-\lambda}}. \quad (3-75)$$

Из (3-75) следует, что

$$0 < \gamma'_{\text{опт}} < 0,5. \quad (3-76)$$

2. При $\sigma^2 < \sigma_p^2$ и достаточно медленном изменении $B(n)$, т. е. при малых λ , функция $Q(\gamma')$ весьма незначительно изменяется вблизи $\gamma'_{\text{опт}}$ (кривые 1-3). Это позволяет без существенных потерь

выбирать вне зависимости от конкретных значений k и λ

$$\gamma'_{\text{опт}} \approx 0,25.$$

3. Увеличение σ^2 приводит к уменьшению $\gamma'_{\text{опт}}$, причем при $\sigma^2 \gg \sigma_{\beta}^2$ зависимость $Q(\gamma')$ приобретает более выраженный экстремальный характер (кривая 5).

Пример 3-3. Модель объекта имеет вид, показанный на рис. 3-7. Измеренное значение выхода объекта в момент i

$$y(i) = u(i) + e(i) = \beta(i) x(i) + e(i), \quad (3-77)$$

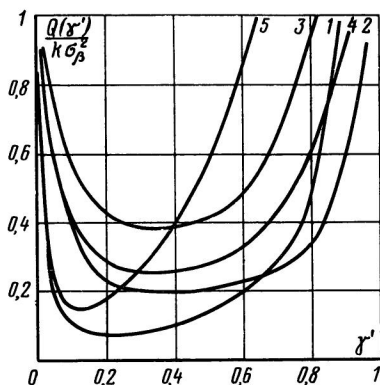


Рис. 3-6. Зависимость среднего квадрата ошибки прогнозирования от величины шага коррекции при $\alpha=1$.

1 — $k=1$; 2 — $k=10$; 3 — $k=1$; 4 — $k=1$;
5 — $k=1$; $\lambda=0,01$; $\lambda=0,01$; $\lambda=0,1$; $\lambda=0,01$;
 $\lambda=0,01$; $\sigma^2/\sigma_{\beta}^2=0,2$; $\sigma^2/\sigma_{\beta}^2=0,2$;
 $\sigma^2/\sigma_{\beta}^2=0,2$; $\sigma^2/\sigma_{\beta}^2=1$; $\sigma^2/\sigma_{\beta}^2=1$.

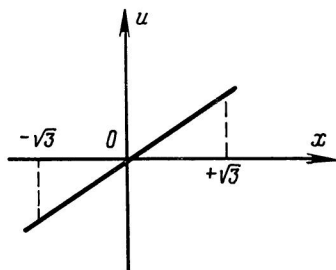


Рис. 3-7. Вид модели объекта для примера 3-3.

где $\beta(i)$, $x(i)$, $e(i)$ — взаимонезависимые стационарные случайные процессы, причем

$$\left. \begin{aligned} R_{\beta}(\tau) &= e^{-0,01\tau}; \\ M[e(i)] &= 0; \\ M[e^2(i)] &= \sigma^2 = 1; \\ M[x(i)x(j)] &= 0 \text{ при } i \neq j. \end{aligned} \right\} \quad (3-78)$$

Пусть значения x равномерно распределены в диапазоне от $-\sqrt{3}$ до $+\sqrt{3}$. Тогда

$$\left. \begin{aligned} M[x(i)] &= 0; \\ \sigma_x^2 &= \frac{1}{2\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} x^2 dx = 1; \\ M(x^4) &= \frac{1}{2\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} x^4 dx = 1,8; \\ \mu &= (k-1) + \mu = 1,8. \end{aligned} \right\} \quad (3-79)$$

Необходимо найти значение параметра γ алгоритма (3-46), обеспечивающее минимальную величину среднего квадрата ошибки прогнозирования выхода объекта на один шаг (т. е. при $\alpha=1$). Выражение для $Q(\gamma)$ в данном случае принимает вид:

$$Q(\gamma) = \frac{2(1 - e^{-0,01})}{(2 - 1,8\gamma)[1 - e^{-0,01}(1 - \gamma)]} + \frac{\gamma}{2 - 1,8\gamma}. \quad (3-80)$$

Рассмотренному примеру соответствует кривая 5 на рис. 3-6. Из рис. 3-6 находим:

$$\gamma'_{\text{опт}} = \frac{\gamma_{\text{опт}}}{2/\rho} = \frac{1,8\gamma_{\text{опт}}}{2} \approx 0,1,$$

откуда

$$\gamma_{\text{опт}} \approx 0,111.$$

Рекуррентный алгоритм прогнозирования процесса $\beta(i)$ на один шаг при этом определяется выражением

$$\hat{\beta}(i+1) = \hat{\beta}(i) + 0,111x(i)(y(i) - \hat{\beta}(i)x(i)). \quad (3-81)$$

Средний квадрат ошибки прогнозирования выхода объекта равен:

$$Q(\gamma_{\text{опт}}) = 0,17. \quad (3-82)$$

ГЛАВА ЧЕТВЕРТАЯ

КУСОЧНО-ЛИНЕЙНАЯ АППРОКСИМАЦИЯ ХАРАКТЕРИСТИК ДРЕЙФУЮЩИХ ОБЪЕКТОВ

При экспериментальном построении математического описания часто модель объекта представляют в виде степенной функции входных переменных x . При этом иногда оказывается затруднительным заранее подобрать вид и степень полинома, с помощью которого такую аппроксимацию можно было бы построить достаточно хорошо во всей области изменения x .

В таких ситуациях используют метод ступенчатой [Л. 45] или кусочно-линейной аппроксимации [Л. 39]. В этом параграфе мы остановимся на применении кусочно-линейной аппроксимации при построении математического описания дрейфующих объектов. Подобные задачи рассматривались в [Л. 44, 46], где в отличие от содержания этой главы использовалась ступенчатая аппроксимация характеристики объекта.

4-1. ОПИСАНИЕ МЕТОДА КУСОЧНО-ЛИНЕЙНОЙ АППРОКСИМАЦИИ

Сущность метода кусочно-линейной аппроксимации заключается в следующем. Область изменения входных переменных x разбивается на ряд подобластей. Для каждой подобласти строится мате-

математическое описание в виде линейного полинома от входных переменных путем обработки данных, относящихся к данной подобласти. Исходной предпосылкой при использовании этого метода является допущение возможности достаточно хорошего представления искомой зависимости сложного вида набором относительно несложных (в данном случае — линейных) функций входных переменных. Сказанное выше поясняет рис. 4-1. Здесь действительная характеристика объекта (кривая 1) представлена с помощью трех отрезков прямых (отрезки 2, 3 и 4), аппроксимирующих кривую 1 в соответствующих диапазонах изменения x . Такая аппроксимация называется

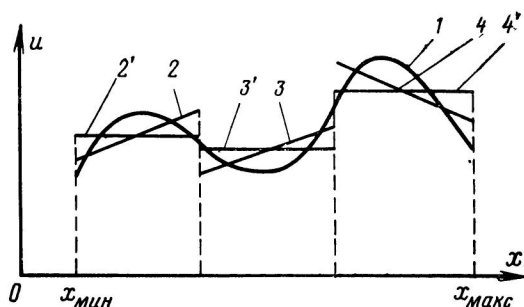


Рис. 4-1. Кусочно-линейная аппроксимация характеристики объекта.

ся кусочно-линейной. В случае ступенчатой аппроксимации для каждого из рассмотренных диапазонов выходная величина u принимается постоянной (отрезки 2', 3' и 4').

Пусть аппроксимация характеристики объекта (модель объекта) для всякой подобласти выбрана линейного или ступенчатого вида. Тогда всякий способ разбиения области изменения x на подобласти будет характеризоваться некоторой величиной показателя точности слежения за дрейфующими коэффициентами модели. Этот показатель достигнет максимального значения при некотором вполне определенном способе разбиения области изменения x .

Действительно, увеличение размера подобластей приводит к увеличению ошибки слежения за счет того, что при выбранном виде аппроксимации модель оказывается неадекватной действительному характеру зависимости во всякой подобласти. С другой стороны, уменьшение размера подобластей ведет к увеличению их числа. При этом коррекция математического описания для всякой подобласти производится менее часто, что ухудшает качество слежения за дрейфом характеристики объекта. Поэтому для всякой конкретной задачи существует свое наилучшее разбиение на подобласти, и способ разбиения требует, вообще говоря, настройки при использовании метода. Заметим, что в некоторых практических случаях целесообразное разбиение области изменения x определяется из технологических соображений. Ниже будут рассматриваться только такие задачи, где способ разбиения на подобласти заранее выбран.

Пусть область изменения x разбита некоторым образом на m подобластей $\omega_1, \omega_2, \dots, \omega_m$. Введем систему функций $\varphi_j(x)$, которые определим так:

$$\varphi_j(x) = \begin{cases} 1 & \text{при } x \in \omega_j; \\ 0 & \text{при } x \notin \omega_j, \end{cases} \quad (4-1)$$

где $j = 1, 2, \dots, m$.

Система функций $\varphi_j(x)$ является ортонормированной, так как

$$\varphi_j(x) \varphi_\mu(x) = \begin{cases} 1 & \text{при } j = \mu; \\ 0 & \text{при } j \neq \mu. \end{cases} \quad (4-2)$$

Используя систему функций $\varphi_j(x)$, можно весьма компактно записать выражение для модели объекта, построенной с использованием кусочно-линейной аппроксимации. Положим, что для j -й под-области модель имеет вид:

$$u(i) = B^{(j)T}(i) x(i) \quad \text{при } x(i) \in \omega_j; \quad j = 1, 2, \dots, m. \quad (4-3)$$

Тогда всю характеристику объекта запишем следующим образом:

$$u(i) = \sum_{j=1}^m B^{(j)T}(i) x(i) \varphi_j(x(i)), \quad (4-4)$$

где $\varphi_j(x(i))$, $j=1, 2, \dots, m$ — система функций, определяемая из (4-1).

В этом параграфе рассматриваются способы нахождения текущих значений коэффициентов модели, вид которой задан с помощью выражения (4-4). Постановка рассматриваемой здесь задачи ничем не отличается от описанной в § 3-1. Как и выше, полагаем, что коэффициенты $B^{(j)}(i)$ изменяются во времени, причем не будет налагаться никаких дополнительных условий на значения коэффициентов $B^{(j)}(i)$, которые могут быть выведены из требования непрерывности характеристики объекта.

При решении задачи воспользуемся подходом, общая характеристика которого дана в § 3-1. Выберем структуру алгоритма решения для всякой подобласти ω_j в виде рекуррентного соотношения типа (3-5). Для обеспечения оптимальных свойств выбранного алгоритма будем искать оценки параметров γ_j , $j=1, 2, \dots, m$ из условия максимальной точности слежения за действительным выходом объекта. Ниже рассмотрен способ оценки параметров алгоритма $\gamma_1, \dots, \gamma_m$ (совокупность этих параметров обозначим вектором Θ) по конечной выборке наблюдений, и получены выражения для аналитической настройки алгоритма в случае известных корреляционных функций случайного дрейфа коэффициентов модели и заданной дисперсии ошибки σ^2 .

4.2. ОЦЕНКА ПАРАМЕТРОВ АЛГОРИТМА ПО КОНЕЧНОЙ ВЫБОРКЕ НАБЛЮДЕНИЙ

Пусть экспериментально получена выборка наблюдений входа и выхода объекта в равноотстоящие моменты времени $1, 2, \dots, n$. Требуется оценить значение коэффициентов модели объекта, представленной в виде (4-4), в момент $(n+\alpha)$, где $\alpha > 0$.

Будем искать решение для любой подобласти ω_j заданного разбиения области изменения $\mathbf{x}$ с помощью рекуррентного алгоритма типа (3-40), который в данном случае можно записать следующим образом:

$$\hat{B}^{(j)}(i) = \begin{cases} \hat{B}^{(j)}(i-1) & \text{при } \mathbf{x}(i-\alpha) \notin \omega_j; \\ \hat{B}^{(j)}(i-1) + \frac{\mathbf{x}(i-\alpha)}{\mathbf{x}^T(i-\alpha)\mathbf{x}(i-\alpha) + \gamma_j} \times \\ \times [y(i-\alpha) - \hat{B}^{(j)T}(i-1)\mathbf{x}(i-\alpha)] & \\ \text{при } \mathbf{x}(i-\alpha) \in \omega_j. \end{cases} \quad (4-5)$$

Параметрами алгоритма, требующими оценки по имеющейся выборке, здесь являются $(\gamma_1, \gamma_2, \dots, \gamma_m) = \mathfrak{G}$.

Для оценки параметров алгоритма воспользуемся критерием (3-42), который в данном случае будет иметь следующий вид:

$$\min S_{\alpha}(\mathfrak{G}) = \min_{\mathfrak{G}} \sum_{\mu=\alpha}^n \left[y(\mu) - \sum_{j=1}^m \hat{B}^{(j)T}(\mu) \mathbf{x}(\mu) \varphi_j \mathbf{x}(\mu) \right]^2, \quad (4-6)$$

где $\varphi_j(\mathbf{x}(\mu))$ — система ортонормированных функций (4-1), определяемая заданным разбиением области изменения $\mathbf{x}$; $\hat{B}^{(j)}(\mu)$ — оценка вектора коэффициентов для подобласти ω_j , вычисляемая в соответствии с алгоритмом (4-5). Раскрывая выражение, записанное в правой части (4-6), и имея в виду свойство ортонормированности функций $\varphi_j(\mathbf{x})$, легко получить:

$$S_{\alpha}(\mathfrak{G}) = \sum_{j=1}^m S_{\alpha}(\gamma_j), \quad (4-7)$$

где

$$S_{\alpha}(\gamma_j) = \sum_{\mu=\alpha}^n [y(\mu) - \hat{B}^{(j)T}(\mu) \mathbf{x}(\mu)]^2 \varphi_j(\mathbf{x}(\mu)). \quad (4-8)$$

Таким образом, задача получения оценки $\mathfrak{G}_{\text{опт}}$ для заданного разбиения области изменения $\mathbf{x}$ распадается на m несвязанных задач вычисления $\hat{\gamma}_{j\text{опт}}$ для всякой подобласти ω_j . Алгоритм решения такой задачи описан в § 3-4.

4-3. ОПТИМАЛЬНЫЙ ВЫБОР ПАРАМЕТРОВ АЛГОРИТМА ПРИ ИЗВЕСТНЫХ КОРРЕЛЯЦИОННЫХ ФУНКЦИЯХ ДРЕЙФУЮЩИХ КОЭФФИЦИЕНТОВ

В этом параграфе рассматривается решение задачи, аналогичной изложенной в § 3-5 для случая, когда модель объекта представлена с помощью набора линейных поверхностей (4-4). Будем исходить из того, что разбиение области изменения вектора переменных $\mathbf{x}$ задано, и требуется определить наилучшее значение параметров алгоритма. Во всякой подобласти аппроксимация ищется в виде линейного полинома, а в качестве алгоритма решения выбран алгоритм (3-46). В [Л. 46] рассмотрено решение близкой задачи, но

для случая, когда аппроксимация строится с помощью ступенчатых функций, а способ разбиения определен лишь с точностью до значения некоторого параметра (величины шага равномерного разбиения области x). Этот параметр в упомянутой работе так же, как и ω , является параметром алгоритма, требующим настройки.

Пусть область изменения входных переменных x разбита на m подобластей ω_j , а модель объекта представляется выражением

$$u(n) = \sum_{j=1}^m B^{(j)T}(n) (x(n) - \bar{x}^{(j)}) \varphi_j(x(n)). \quad (4.9)$$

Здесь n — дискретные моменты времени съема информации со входа и выхода объекта:

$$\varphi_j(x(n)) = \begin{cases} 1 & \text{при } x(n) \in \omega_j; \\ 0 & \text{при } x(n) \notin \omega_j; \end{cases}$$

$\bar{x}^{(j)}$ — среднее значение вектора x для подобласти ω_j , определяемое по формуле

$$\bar{x}^{(j)} = M \left[\frac{x}{x \in \omega_j} \right]. \quad (4.10)$$

В общем случае аппроксимация модели объекта для всякой подобласти ω_j должна содержать еще свободный член d_j (или некоторую постоянную составляющую). Чтобы не усложнять анализа (здесь, как и в § 3-5), будем считать, что значение свободного члена для всякой подобласти ω_j известно и не изменяется во времени. Тогда используемые при построении математического описания значения $y(n)$ оказываются равными:

$$y(n) = \sum_{j=1}^m (y'(n) - d_j) \varphi_j(x(n)), \quad (4.11)$$

где $y'(n)$ — фактически измеренное значение выхода объекта; d_j — известное значение постоянной составляющей для подобласти ω_j , $j=1, 2, \dots, m$.

Пусть во всякий момент i событие $x(i) \in \omega_j$ происходит с вероятностью q_j , где $j=1, 2, \dots, m$ и q_j не зависит от момента времени i .

Рассмотрим теперь процесс $x^{(j)}(n)$, который получается из $x(i)$ путем центрирования (вычитания $\bar{x}^{(j)}$) и удаления интервалов времени, когда $x(i)$ не принадлежит подобласти ω_j , т. е. $x(i) \notin \omega_j$. Это преобразование процесса $x(i)$ иллюстрирует рис. 4-2. Здесь рассмотрен случай, когда $x(i)$ — одномерный вектор: 1 — реализация процесса $x(i)$; ω_j — диапазон изменения x , соответствующий j -й подобласти разбиения; 2 — процесс $x^{(j)}(n)$, который получен из $x(i)$ путем удаления интервалов времени, когда $x(i) \notin \omega_j$, и центрирования (вычитания $\bar{x}^{(j)}$) — математического ожидания процесса $x^{(j)}(n)$.

Относительно процесса $x^{(j)}(n)$ примем те же предположения, что и для процесса $x(n)$ в задаче, рассмотренной в § 3-5. Полагаем, что математическое ожидание $M[x^{(j)}(n)] = 0$, а последовательность значений $x_{\mu}^{(j)}(n)$ независима, т. е.

$$M[x_{\mu}^{(j)}(n) x_{\mu}^{(j)}(n+\tau)] = 0 \quad (4-12)$$

при $\tau \neq 0$ для $\mu = 1, 2, \dots, k$; $j = 1, 2, \dots, m$.

Ковариационная матрица вектора $\mathbf{x}^{(j)}$ в любой момент n имеет вид:

$$M[\mathbf{x}^{(j)}(n) (\mathbf{x}^{(j)}(n))^T] = \mathbf{I}_k, \quad (4-13)$$

где $\mathbf{I}_k$ — единичная матрица размерности k ; $j = 1, 2, \dots, m$.

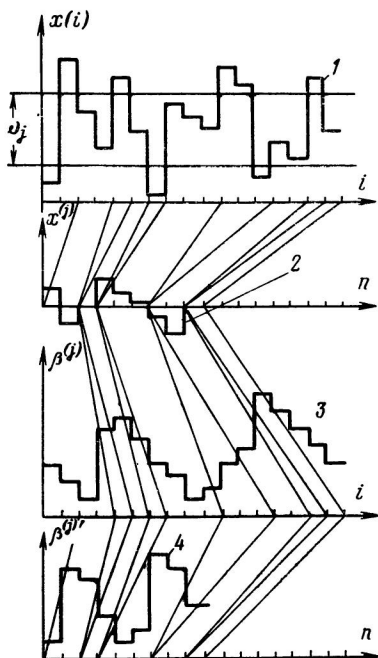


Рис. 4-2. Схема преобразования процесса $x(i)$ и $\beta^{(j)}(i)$.

Кроме того, четвертые моменты распределения переменных $x_{\nu}^{(j)}$ примем равными:

$$M[(x_{\nu}^{(j)})^4] = \mu_{\nu j} \text{ для } \nu = 1, 2, \dots, k. \quad (4-14)$$

Положим, что $\mathbf{B}^{(j)}(i)$ есть k -мерный стационарный случайный процесс, причем известны корреляционные функции:

$$R_{ii}^{(j)}(\tau) = (\sigma_{ii}^{(j)})^2 f_{ii}^{(j)}(\tau), \quad (4-15)$$

где $(\sigma_{\mu\nu}^{(j)})^2 < \infty$; $\lim_{|\tau| \rightarrow \infty} |f_{ii}^{(j)}(\tau)| = 0$; $i = 1, 2, \dots, k$; $j = 1, 2, \dots, m$.

Пусть измеряемое значение выхода объекта в момент n

$$y(n) = u(n) + e(n), \quad (4-16)$$

где $e(n)$ — последовательность независимых случайных величин с нулевым математическим ожиданием и известной дисперсией σ^2 . Кроме того, примем, как и в § 3-5, что случайные процессы $x^{(j)}(n)$, $B^{(j)}(n)$ ($j=1, 2, \dots, m$) и $e(n)$ взаимонезависимы.

Допустим, что для прогнозирования значений коэффициентов модели используется алгоритм типа (3-46), который для случая кусочно-линейной аппроксимации можно записать следующим образом:

$$\hat{B}^{(j)}(n + \alpha + 1) = \begin{cases} \hat{B}^{(j)}(n + \alpha) & \text{при } x(n + 1) \notin \omega_j; \\ \hat{B}^{(j)}(n + \alpha) + \gamma_j x(n + 1) [y(n + 1) - x^T(n + 1) \hat{B}^{(j)}(n + \alpha)] & \text{при } x(n + 1) \in \omega_j. \end{cases} \quad (4-17)$$

Выражение (4-17) записано для j -й подобласти разбиения. Аналогичные выражения могут быть записаны для любой подобласти $j=1, 2, \dots, m$.

Требуется найти значение вектора $\mathfrak{G}$ с координатами $\gamma_1, \dots, \gamma_m$, которое обеспечивает минимум среднего квадрата ошибки прогнозирования действительного выхода объекта в установившемся режиме слежения:

$$\min_{\mathfrak{G}} Q(\mathfrak{G}) = \min_{\mathfrak{G}} \lim_{n \rightarrow \infty} M \left\{ \left[u(n) - \sum_{j=1}^m B^{(j)T}(n) (x(n) - \bar{x}^{(j)}) \varphi_j(x(n)) \right]^2 \right\} = \min_{\mathfrak{G}} \lim_{n \rightarrow \infty} M \left\{ \sum_{j=1}^m \varphi_j(x(n)) [u(n) - \hat{B}^{(j)T}(x(n) - \bar{x}^{(j)})]^2 \right\}. \quad (4-18)$$

Так как в любой момент n только один член суммы в (4-18) отличен от нуля, причем $\varphi_j=1$ с вероятностью q_j , то выражение (4-18) можно представить в виде

$$\min_{\mathfrak{G}} Q(\mathfrak{G}) = \min_{\mathfrak{G}} \sum_{j=1}^m q_j Q_j(\gamma_j), \quad (4-19)$$

где $Q_j(\gamma_j)$ — средний квадрат ошибки прогнозирования выхода объекта для подобласти ω_j .

Найдем выражения для $Q_j(\gamma_j)$. Для этого воспользуемся формулой (3-71). Это можно сделать, если предварительно свести постановку задачи вычисления среднего квадрата ошибки прогнозирования выхода объекта для j -й подобласти к постановке задачи

§ 3-5. Дело в том, что изменение вектора оценок коэффициентов $\hat{B}^{(j)}(i)$ для подобласти ω_j происходит в соответствии с алгоритмом (4-17) только в те моменты i , когда $x(i) \in \omega_j$, в отличие от алгоритма (3-46), для которого получено выражение (3-71). Таким образом, для области ω_j в какие-то моменты времени изменения оценок не происходит; эти моменты времени как бы «пропускаются». Результатом таких пропусков является отличие последовательности значений $B^{(j)'}(n)$, рассматриваемой только в моменты i , когда происходит изменение оценок $\hat{B}^{(j)}$, от последовательности $B^{(j)}(i)$, для которой определена корреляционная функция $R_{ii}^{(j)}(\tau)$; $i=1, 2, \dots, k$. Поэтому, пользуясь для вычисления $Q_j(\gamma_j)$ формулой (3-71), нужно исходить из корреляционной функции процесса $B^{(j)'}(n)$.

Рассмотрим механизм образования процесса $B^{(j)'}(n)$ и способ вычисления корреляционных функций $R_{ii}^{(j)'}(\tau)$. Вернемся к рис. 4-2.

Процесс 4 получен из процесса 3 $\beta^{(j)}(i)$ путем удаления интервалов времени, когда $x(i) \notin \omega_j$. Пусть известна корреляционная функция $R^{(j)}(\tau)$ процесса $\beta^{(j)}(i)$. Необходимо определить корреляционную функцию $R^{(j)'}(\tau)$ процесса $\beta^{(j)'}(n)$, если известно, что вероятность события $x(i) \in \omega_j$ во всякий момент i равна q_j .

Решение можно получить с помощью следующих рассуждений [Л. 46]. Если $\beta^{(j)'}(n) = \beta^{(j)}(i)$, то $\beta^{(j)'}(n+\tau) = \beta^{(j)}(i+\tau+\theta)$, где θ — случайная величина.

Тогда

$$\beta^{(j)'}(n) \beta^{(j)'}(n+\tau) = \beta^{(j)}(i) \beta^{(j)}(i+\tau+\theta). \quad (4-20)$$

Усредняя (4-20) по множеству реализаций последовательности x (усреднение по x обозначим символом M_x), получаем:

$$M_x [\beta^{(j)'}(n) \beta^{(j)'}(n+\tau)] = \sum_{\theta=0}^{\infty} P_j(\theta, \tau) \beta^{(j)}(i) \beta^{(j)}(i+\tau+\theta), \quad (4-21)$$

где $P_j(\theta, \tau)$ — вероятность того, что за интервал $[\tau+\theta]$ x попадает в рассматриваемую область ω_j точно τ раз, причем в $(i+\tau+\theta)$ -й момент x также попадает в эту область. Усредним (4-21) по множеству реализаций процесса $\beta^{(j)}(i)$. Тогда получим:

$$\begin{aligned} M_{\beta^{(j)}} M_x [\beta^{(j)'}(n) \beta^{(j)'}(n+\tau)] &= M [\beta^{(j)'}(n) \beta^{(j)'}(n+\tau)] = \\ &= \sum_{\theta=0}^{\infty} P_j(\theta, \tau) M [\beta^{(j)}(i) \beta^{(j)}(i+\tau+\theta)] = \\ &= \sum_{\theta=0}^{\infty} P_j(\theta, \tau) R^{(j)}(\tau+\theta) + (m^{(j)})^2, \end{aligned} \quad (4-22)$$

где $R^{(j)}(\tau)$ и $m^{(j)}$ — корреляционная функция и математическое ожидание процесса $\beta^{(j)}(i)$ соответственно.

Так как

$$M [\beta^{(j)'}(n+\tau)] = M \left[\sum_{\theta=0}^{\infty} P_j(\theta, \tau) \beta^{(j)}(i+\tau+\theta) \right] = m^{(j)}, \quad (4-23)$$

то из (4-22) следует, что

$$R^{(j)'}(\tau) = \sum_{\theta=0}^{\infty} P_j(\theta, \tau) R^{(j)}(\tau + \theta). \quad (4-24)$$

Значение $P_j(\theta, \tau)$, когда вероятность события $\in \tau \omega_j$ в любой момент времени равна q_j , записывается следующим образом [Л. 47]:

$$P_j(\theta, \tau) = C_{\tau+\theta-1}^{\theta} q_j^{\tau} (1 - q_j)^{\theta}, \quad (4-25)$$

где $C_{\tau+\theta-1}^{\theta}$ — число сочетаний из $(\tau + \theta - 1)$ по θ .

Таким образом, корреляционная функция последовательности $\beta^{(j)'}(n)$ определяется путем следующего преобразования корреляционной функции последовательности $\beta^{(j)}(i)$:

$$R^{(j)'}(\tau) = \sum_{\theta=0}^{\infty} C_{\tau+\theta-1}^{\theta} q_j^{\tau} (1 - q_j)^{\theta} R^{(j)}(\tau + \theta). \quad (4-26)$$

Определим, как влияет преобразование (4-26) на вид корреляционной функции. Положим

$$R^{(j)}(\tau) = (\sigma_{\beta}^{(j)})^2 e^{-\lambda |\tau|}; \lambda > 0. \quad (4-27)$$

Тогда для $R^{(j)'}(\tau)$ имеем:

$$\begin{aligned} R^{(j)'}(\tau) &= \sum_{\theta=0}^{\infty} C_{\tau+\theta-1}^{\theta} q_j^{\tau} (1 - q_j)^{\theta} (\sigma_{\beta}^{(j)})^2 e^{-\lambda (\tau + \theta)} = \\ &= (\sigma_{\beta}^{(j)})^2 e^{-\lambda \tau} q_j^{\tau} \sum_{\theta=0}^{\infty} C_{\tau+\theta-1}^{\theta} [(1 - q_j) e^{-\lambda}]^{\theta}. \end{aligned} \quad (4-28)$$

Сумму в (4-28) можно посчитать, если воспользоваться формулой бинома Ньютона

$$(1 - g)^{-\tau} = \sum_{\theta=0}^{\infty} \left\| \begin{matrix} -\tau \\ \theta \end{matrix} \right\| (-g)^{\theta}, \quad (4-29)$$

где

$$\begin{aligned} \left\| \begin{matrix} -\tau \\ \theta \end{matrix} \right\| &= \frac{-\tau (-\tau - 1) (-\tau - 2) \dots (-\tau - \theta + 1)}{\theta!} = \\ &= (-1)^{\theta} \frac{(\tau + \theta - 1) (\tau + \theta - 2) \dots \tau}{\theta!} = (-1)^{\theta} C_{\tau+\theta-1}^{\theta}. \end{aligned} \quad (4-30)$$

Подставляя (4-30) в (4-29) и приняв $g = (1 - q_j) e^{-\lambda}$, получим:

$$\sum_{\theta=0}^{\infty} C_{\tau+\theta-1}^{\theta} [(1 - q_j) e^{-\lambda}]^{\theta} = \frac{1}{1 - (1 - q_j) e^{-\lambda}}, \quad (4-31)$$

Тогда

$$R^{(i)}(j)' = (\sigma_{\beta}^{(j)})^2 \left[\frac{q_j e^{-\lambda}}{1 - (1 - q_j) e^{-\lambda}} \right]^{\tau}. \quad (4-32)$$

Из (4-32) следует, что преобразование (4-25) при любом $q_j < 1$ приводит к более быстрому затуханию корреляционной функции, причем при $q_j = 1$ корреляционная функция не изменяется, а при уменьшении q_j скорость затухания увеличивается. Этот факт иллюстрирует рис. 4-3, где кривая 1 соответствует исходной корреляционной функции $f(\tau) = e^{-0,1\tau}$, а кривые 2 и 3 — корреляционным функциям, преобразованным по формуле (4-32) при $q_j = 0,2$ и $0,02$

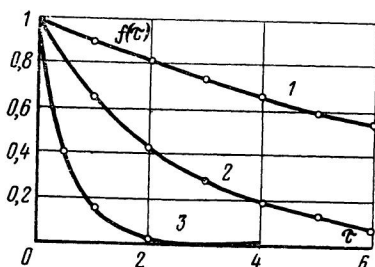


Рис. 4-3. Преобразование корреляционной функции процесса $\beta^{(j)}(i)$ при кусочно-линейной аппроксимации модели объекта.

соответственно. Имея в виду выражение (4-26), нетрудно вычислить:

$$R_{ii}^{(j)'}(\tau) = [\sigma_{ii}^{(j)'}]^2 f_{ii}^{(j)'}(\tau); \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, m. \quad (4-33)$$

Теперь можно воспользоваться формулой (3-71) и записать выражение для среднего квадрата $Q_j(\gamma_j)$ ошибки прогнозирования для подобласти ω_j :

$$\begin{aligned} Q_j(\gamma_j) = & \frac{2}{2 - \gamma_j \rho_j} \left[\sum_{i=1}^k (\sigma_{ii}^{(j)'})^2 + \gamma_j (1 - \gamma_j \rho_j) \sum_{i=1}^k (\sigma_{ii}^{(j)'})^2 \times \right. \\ & \times \sum_{v=0}^{\infty} (1 - \gamma_j)^v f_{ii}^{(j)'}(v+1) - \gamma_j (2 - \gamma_j \rho_j) \times \\ & \times \sum_{i=1}^k (\sigma_{ii}^{(j)'})^2 \sum_{v=0}^{\infty} (1 - \gamma_j)^v f_{ii}^{(j)'}(v+\alpha) \left. \right] + \frac{\gamma_j k \sigma^2}{2 - \gamma_j \rho_j}, \end{aligned} \quad (4-34)$$

где

$$\rho_j = (k-1) + \mu_j. \quad (4-35)$$

Для определения $\gamma_{\text{опт}}$ необходимо найти минимум этого выражения по γ_j .

Решая эту задачу последовательно для всех подобластей ω_j ; $j=1, 2, \dots, m$, находим оптимальное значение вектора $\Phi_{\text{опт}}$.

Пример 4-1. Характеристика объекта представляется двумя отрезками (рис. 4-4):

$$u(i) = \begin{cases} \beta^{(1)}(i) (x(i) - \bar{x}^{(1)}) & \text{при } x(i) \in \omega_1; \\ \beta^{(2)}(i) (x(i) - \bar{x}^{(2)}) & \text{при } x(i) \in \omega_2, \end{cases} \quad (4-36)$$

где $x(i) \in \omega_1$, когда $0 < x(i) \leq 2\sqrt{3}$, и $x(i) \in \omega_2$ при $0 > x(i) \geq -2\sqrt{3}$.

Положим, что изменение характеристики объекта во времени происходит за счет поворота отрезков вокруг точек $\bar{x}^{(1)}$ и $\bar{x}^{(2)}$, так

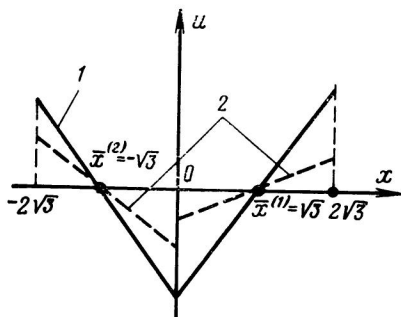


Рис. 4-4. Вид модели объекта для примера 4-1.

что в некоторый момент i характеристика объекта может выглядеть как зависимости 1 или 2 на рис. 4-4.

Пусть последовательность значений $x(i)$ независима и x равномерно распределен в интервале от $-2\sqrt{3}$ до $+2\sqrt{3}$. Известны дисперсия ошибки эксперимента $\sigma^2=1$ и корреляционные функции коэффициентов $\beta^{(1)}(i)$ и $\beta^{(2)}(i)$:

$$\left. \begin{aligned} R^{(1)}(\tau) &= e^{-0,1|\tau|}; \\ R^{(2)}(\tau) &= e^{-0,01|\tau|}. \end{aligned} \right\} \quad (4-37)$$

Для прогнозирования значений коэффициентов объекта используется алгоритм типа (4-17). Требуется определить значения параметров алгоритма (4-17) γ_1 и γ_2 , обеспечивающие минимальную величину среднего квадрата ошибки прогнозирования действительного выхода объекта на один шаг (т. е. при $\alpha=1$). Нетрудно убедиться, что в рассматриваемом случае (см. пример 3-3)

$$\left. \begin{aligned} \sigma_{x(1)}^2 &= \sigma_{x(2)}^2 = 1; \\ \mu_1 &= \mu_2 = 1,8. \end{aligned} \right\} \quad (4-38)$$

Кроме того,

$$q_1 = q_2 = \frac{1}{2}. \quad (4-39)$$

Найдем корреляционные функции процессов $\beta^{(1)'}(i)$ и $\beta^{(2)'}(i)$. Пользуясь формулой (4-32), получим:

$$\left. \begin{aligned} R^{(1)'}(\tau) &= \left(\frac{e^{-0,1}}{2 - e^{-0,1}} \right)^\tau = (0,826)^\tau; \\ R^{(2)'}(\tau) &= \left(\frac{e^{-0,01}}{2 - e^{-0,01}} \right)^\tau = (0,98)^\tau. \end{aligned} \right\} \quad (4-40)$$

Теперь запишем выражения для $Q_1(\gamma_1)$ и $Q_2(\gamma_2)$, подставляя (4-40) в (4-34) и имея в виду, что $\rho_1 = \rho_2 = 1,8$:

$$\left. \begin{aligned} Q_1(\gamma_1) &= \frac{2(1 - 0,826)}{(2 - 1,8\gamma_1)[1 - 0,826(1 - \gamma_1)]} + \frac{\gamma_1}{2 - 1,8\gamma_1}; \\ Q_2(\gamma_2) &= \frac{2(1 - 0,98)}{(2 - 1,8\gamma_2)[1 - 0,98(1 - \gamma_2)]} + \frac{\gamma_2}{2 - 1,8\gamma_2}. \end{aligned} \right\} \quad (4-41)$$

Определим значения $\gamma_{1\text{опт}}$ и $\gamma_{2\text{опт}}$, минимизирующие (4-41): $\gamma_{1\text{опт}} = 0,178$; $\gamma_{2\text{опт}} = 0,15$.

Алгоритм решения задачи определяется следующими рекуррентными соотношениями:

$$\begin{aligned} \hat{\beta}^{(1)}(i) &= \begin{cases} \hat{\beta}^{(1)}(i-1) & \text{при } x(i-1) \in \omega_2; \\ \hat{\beta}^{(1)}(i-1) + 0,178(y(i-1) - \hat{\beta}^{(1)}(i-1)x(i-1))x(i-1) & \text{при } x(i-1) \in \omega_1; \end{cases} \\ \hat{\beta}^{(2)}(i) &= \begin{cases} \hat{\beta}^{(2)}(i-1) & \text{при } x(i-1) \in \omega_1; \\ \hat{\beta}^{(2)}(i-1) + 0,15(y(i-1) - \hat{\beta}^{(2)}(i-1)x(i-1))x(i-1) & \text{при } x(i-1) \in \omega_2. \end{cases} \end{aligned}$$

Средний квадрат ошибки прогнозирования выхода при этом равен:

$$Q(\mathbb{G}) = \frac{1}{2} 0,832 + \frac{1}{2} 0,225 = 0,5285.$$

В двух последних главах были изложены методы построения математического описания объекта по экспериментальным данным в случаях, когда коэффициенты модели изменяются во времени. Задача построения математического описания дрейфующего объекта рассматривалась как задача прогнозирования значений коэффициентов модели объекта для момента времени $(n+a)$ по наблюдениям за входом и выходом объекта до момента n .

Для решения задачи были использованы рекуррентные алгоритмы, совпадающие по структуре с алгоритмами метода стохастической аппроксимации. Оптимальные свойства используемых алгоритмов достигаются в рамках заданной структуры путем выбора

наилучших значений параметров алгоритма γ . В работе приведен алгоритм нахождения оценки $\hat{\gamma}_{opt}$ по конечной выборке наблюдений и показана состоятельность получаемой оценки. Этот алгоритм рассматривался как некоторое дополнение к методу множественного регрессионного анализа. Применяя его, можно получить более точные оценки коэффициентов объекта для момента времени $(n+\alpha)$ в случаях, когда имеет место дрейф. Кроме того, показано, как можно аналитически выбрать оптимальные значения параметров алгоритма γ_{opt} при решении задачи прогнозирования выхода объекта при наличии дрейфа, описываемого стационарным случайным процессом. Решение этой задачи требует знания корреляционных функций дрейфующих коэффициентов и дисперсии ошибки эксперимента. Такое рассмотрение позволяет заключить, насколько хорош может быть выбранный алгоритм во всякой ситуации, насколько целесообразно его использовать для прогнозирования дрейфующих параметров.

Способы решения задачи прогнозирования дрейфующих коэффициентов модели, безусловно, не ограничиваются методами, рассмотренными в этой работе. Достоинство приводимых здесь методов заключается прежде всего в их простоте и совершенно очевидном физическом смысле. Вероятно, далеко не всегда простота метода является достоинством, особенно если она идет в ущерб точности. Однако в условиях априорной неопределенности, когда требуется производить настройку алгоритма в ходе решения задачи (как это делается, например, в § 3-4), необходимость настройки малого числа параметров алгоритма, безусловно, является положительным свойством метода.

ПРИЛОЖЕНИЕ

Некоторые сведения из матричной алгебры

Пусть A — матрица размером $(n \times k)$, т. е. имеющая n строк и k столбцов с элементами a_{ij} , где $i=1, \dots, n$; $j=1, \dots, k$. Транспонированная матрица A^T имеет k строк и n столбцов.

Квадратная матрица A называется симметричной, если $A=A^T$.

Квадратная матрица, у которой все элементы, расположенные вне главной диагонали, равны нулю, называется диагональной матрицей.

Квадратная матрица I_k размером $(k \times k)$, у которой на главной диагонали стоят единицы, а все остальные элементы равны нулю, называется единичной матрицей размера k (или порядка k).

Если A — матрица размером $(n \times k)$, а B — матрица размером $(n \times k)$, то сумма $(A+B)$ есть матрица C размером $(n \times k)$ с элементами c_{ij} ; $i=1, \dots, n$; $j=1, \dots, k$. Сумма $(A+B)$ определена, если матрицы A и B одинакового размера.

Пусть A — матрица размером $(n \times m)$, а B — матрица размером $(m \times k)$. Произведение AB есть матрица C размером $(n \times k)$ с элемен-

тами $c_{jq} = \sum_{s=1}^n a_{js} b_{sq}$; $j=1, \dots, m$; $q=1, \dots, k$.

Произведение AB определено, если число столбцов A равно числу строк B .

Умножение матриц обладает следующими свойствами:

- 1) ассоциативностью: $A(BC) = (AB)C$;
- 2) дистрибутивностью относительно сложения: $(A+B)C = AC + BC$;
- 3) ассоциативностью относительно умножения на числа: $a(bA) = (ab)A$ и $a(AB) = (aA)B$, где a, b — числа; A, B — матрицы.

Умножение матриц не коммутативно, т. е. $AB \neq BA$.
Если I_k — единичная матрица размером k , A — матрица $(k \times k)$, то $I_k A = A I_k = A$. В частности, имеет место равенство $\rho I_k B = \rho I_k AB$, где ρ — число:

$$(A^T)^T = A, (A^{-1})^{-1} = A.$$

Символы транспонирования и обращения можно менять местами:

$$(A^T)^{-1} = (A^{-1})^T.$$

Если ранг матрицы A размером $(n \times n)$ равен $\text{rg } A = m$, где $m < n$, то число линейно независимых строк матрицы A , так же как и число линейно независимых столбцов, равно m .

Следом квадратной матрицы A размером $(n \times n)$ называется сумма ее диагональных элементов: $\text{sp}(A) = \sum_{j=1}^n a_{jj}$.

Если I_k — единичная матрица размером k , то $\text{sp}(I_k) = k$, а $\text{sp}(ABC) = \text{sp}(CAB) = \text{sp}(BCA)$.

Квадратная матрица A называется идемпотентной, если выполняются следующие условия: $A = A^T$, $A = A^2$. Если A — идемпотентная матрица ранга r , то $\text{sp}(A) = r$.

Последовательность k чисел, взятых в определенном порядке, будем называть k -мерным вектором (или точкой k -мерного пространства). Любой элемент k -мерного пространства может быть представлен в виде матрицы Y размера $(k \times 1)$, которую будем называть вектором-столбцом.

Скалярным произведением z двух k -мерных векторов Y_1 и Y_2 называется величина $z = Y_1^T Y_2 = Y_2^T Y_1$.

Если Y — матрица размером $(k \times 1)$, т. е. k -мерный вектор, то $YY^T YY^T = Y^T YY^T$.

Длиной l вектора Y в k -мерном пространстве называется положительное значение корня квадратного из скалярного произведения $Y^T Y$, т. е. $l = \sqrt{Y^T Y}$.

Под нормой вектора Y будем понимать положительное значение корня квадратного из скалярного произведения $Y^T Y$, т. е. длину l вектора Y .

Если Y — n -мерный вектор, а A — матрица размером $(n \times n)$, то квадратичная форма $Y^T A Y$ определяется как $\sum_{j=1}^n \sum_{q=1}^n y_j y_q a_{jq}$. Если

$$A = I_n, \text{ то } Y^T A Y = \sum_{j=1}^n y_j^2.$$

Если A — вектор-столбец $(n \times 1)$ и Y — вектор-столбец $(n \times 1)$, то $z = A^T Y = A Y^T$ и $\partial z / \partial Y = A$.

Если A — матрица $(n \times n)$, Y — n -мерный вектор и $z = Y^T A Y$, то $\partial z / \partial Y = 2A Y$.

Некоторые сведения из математической статистики

Величина ξ , которая при повторяющихся условиях опыта принимает случайные значения, называется случайной величиной.

Совокупность всех возможных значений случайной величины для всех возможных условий опыта называется генеральной совокупностью.

Совокупность ограниченного числа значений случайной величины, полученных в результате эксперимента, называется выборкой из генеральной совокупности. Если n выборочных значений $z_1, \dots, z_n$ случайной величины ξ получены для n независимо изменявшихся условий опыта, то $z_1, \dots, z_n$ можно рассматривать как n независимых случайных величин. Такая выборка называется случайной выборкой объема n .

Свойства генеральной совокупности описываются функцией $f(z)$ распределения вероятностей возможных значений z_i , которые мо-

жет принимать случайная величина ξ . Значение $f(z_1)$ функции распределения есть вероятность события $\{\xi < z_1\}$, т. е. $f(z_1) = P\{\xi < z_1\}$.

Производная $\varphi(z)$ от функции распределения $f(z)$ называется плотностью вероятности случайной величины ξ . Из соотношения $f'(z) = \varphi(z)$ следует:

$$f(z) = \int_{-\infty}^z \varphi(z) dz.$$

Вероятность того, что значение ξ лежит в пределах между z_1 и $z_1 + dz$, равна:

$$P\{z_1 < \xi < z_1 + dz\} = \int_{z_1}^{z_1 + dz} \varphi(z) dz.$$

Совместная функция распределения m случайных величин $\xi_1, \dots, \xi_m$ определяет вероятность события $\{\xi_1 < z_1, \dots, \xi_m < z_m\}$. Случайные величины $\xi_1, \dots, \xi_m$ называются независимыми, если $f(z_1, \dots, z_m) = f(z_1) \cdot \dots \cdot f(z_m)$.

Центр группирования значений случайной величины характеризуется математическим ожиданием $M[\xi] = \int_{-\infty}^{\infty} z \varphi(z) dz$, которое обладает следующими свойствами:

- 1) $M[c] = c$, если $c = \text{const}$;
- 2) $M[c\xi] = cM[\xi]$;
- 3) $M[\xi + \eta] = M[\xi] + M[\eta]$;
- 4) $M[a + b\xi] = a + bM[\xi]$,

где a и $b = \text{const}$.

Если Ξ — случайная матрица (т. е. матрица, элементами ξ_{ij} которой являются случайные величины), то

$$M[\Xi] = \Psi,$$

где Ψ — матрица с элементами $\Psi_{ij} = M[\xi_{ij}]$.

Степень рассеяния значений случайной величины вокруг центра группирования характеризуется дисперсией

$$D[\xi] = \int_{-\infty}^{\infty} (z - M[\xi])^2 \varphi(z) dz,$$

которая обладает следующими свойствами:

- 1) $D[c] = 0$, если $c = \text{const}$;
- 2) $D[c\xi] = c^2 D[\xi]$;
- 3) $D[a + b\xi] = b^2 D[\xi]$, где a и $b = \text{const}$.

Дисперсию можно рассматривать как математическое ожидание квадратов отклонений, т. е. $D[\xi] = M[(z - M[\xi])^2]$.

Величина, равная квадратному корню из дисперсии, называется среднеквадратическим отклонением.

Дисперсия случайной величины ξ является центральным моментом второго порядка. Иногда используются моменты более высокого

порядка. Например, центральный момент случайной величины ξ порядка k равен:

$$\mu_k = M[\xi^k] = \int_{-\infty}^{\infty} (z - M[\xi])^k \varphi(z) dz.$$

Если значения случайной величины есть результат действия многочисленных независимых и примерно одинаковых по силе факторов, то случайная величина имеет нормальный закон распределения. Плотность нормального распределения имеет вид:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}};$$

соответственно функция распределения

$$f(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz.$$

Числовые параметры μ и σ^2 полностью характеризуют генеральную совокупность нормально распределенной случайной величины, причем μ — математическое ожидание генеральной совокупности, а σ^2 — ее дисперсия. Таким образом, $\xi \sim N_1(\mu, \sigma^2)$.

Всякое нормальное распределение можно свести к нормированному нормальному распределению с $\mu=0$ и $\sigma^2=1$, если вместо случайной величины ξ рассматривать ее преобразование

$$\xi' = \frac{\xi - \mu}{\sigma}.$$

Пусть случайные величины $\xi_1, \dots, \xi_m$ имеют совместное нормальное распределение. Функция плотности вероятности совместного нормального распределения имеет вид:

$$\varphi(\mathbf{Z}) = (2\pi)^{-\frac{m}{2}} |D|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{Z}-\mathbf{M})^T \mathbf{D}^{-1}(\mathbf{Z}-\mathbf{M})},$$

где $\mathbf{Z} = (m \times 1)$ — вектор значений случайных величин $\xi_1, \dots, \xi_m$; $\mathbf{M}$ — вектор математических ожиданий $\mu_1, \dots, \mu_m$ случайных величин $\xi_1, \dots, \xi_m$; $\mathbf{D}$ — симметричная матрица размером $(m \times m)$, являющаяся ковариационной матрицей совокупности случайных величин $\xi_1, \dots, \xi_m$; $|D|$ — определитель ковариационной матрицы.

Диагональные элементы ковариационной матрицы представляют собой дисперсии соответствующих случайных величин:

$$d_{jj} = D[\xi_j] = \sigma_j^2 = M[(z_j - M[\xi_j])^2]; \quad j = 1, \dots, m,$$

а недиагональные элементы ковариационной матрицы есть ковариации соответствующих случайных величин:

$$d_{jq} = d_{qj} = \text{cov}[\xi_j, \xi_q] = M[(z_j - M[\xi_j])(z_q - M[\xi_q])];$$

$$j, q = 1, \dots, m; \quad j \neq q.$$

Ковариации отличаются от коэффициентов корреляции нормировкой

$$r_{jq} = r_{qj} = \frac{d_{jq}}{\sigma_j \sigma_q}; \quad j \neq q; \quad r_{jj} = 1.$$

Оценки параметров распределения, полученные по данным выборки, называются выборочными оценками. Например, в качестве выборочной оценки математического ожидания используют среднее значение

$$z = \frac{1}{n} \sum_{i=1}^n z_i, \quad \text{а для дисперсии — выборочную дисперсию } \hat{\sigma}^2 =$$

$$= \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2. \quad \text{Выборочные оценки будут оптимальными, если}$$

они обладают рядом свойств, важнейшими из которых являются состоятельность, несмещенность, эффективность.

Оценка $\hat{\theta} = \psi(z_1, \dots, z_n)$ параметра θ называется состоятельной, если при неограниченном увеличении n ее значение с вероятностью единица стремится к истинному значению параметра θ , т. е.

$$P \left\{ \lim_{n \rightarrow \infty} \psi(z_1, \dots, z_n) = \theta \right\} = 1.$$

Оценка $\hat{\theta} = \psi(z_1, \dots, z_n)$ называется несмещенной, если при любом n ее математическое ожидание равно истинному значению параметра θ , т. е. $M[\hat{\theta}] = \theta$. Требование несмещенности устраняет систематическую погрешность оценки, обусловленную ограниченностью выборки. Для состоятельной оценки эта погрешность стремится к нулю при $n \rightarrow \infty$.

Оценка $\hat{\theta} = \psi(z_1, \dots, z_n)$ называется эффективной, если среди прочих оценок того же параметра она обладает наименьшей дисперсией.

Если вид функции распределения известен, то для получения выборочной оценки $\hat{\theta} = \psi(z_1, \dots, z_n)$ удобно воспользоваться методом максимального правдоподобия, который сводится к следующему.

Случайную выборку $z_1, \dots, z_n$ можно рассматривать как n -мерную случайную величину, имеющую совместную плотность распределения $L(z_1, \dots, z_n) = \varphi(z_1) \cdot \varphi(z_2) \cdot \dots \cdot \varphi(z_n)$, называемую функцией правдоподобия. Для любой выборки $z_1, \dots, z_n$ значение функции правдоподобия есть функция неизвестных параметров распределения $\theta_1, \dots, \theta_m$. По методу наибольшего правдоподобия в качестве оценки параметра θ_j выбирается такая функция $\hat{\theta}_j = \psi_j(z_1, \dots, z_n)$, которая максимизирует величину $L(z_1, \dots, z_n, \hat{\theta}_j)$. Практически удобнее пользоваться $\ln L(z_1, \dots, z_n)$ вместо $L(z_1, \dots, z_n)$, что правомерно, поскольку максимумы этих функций совпадают. Тогда для получения совокупности m наиболее правдоподобных оценок $\hat{\theta}_1, \dots, \hat{\theta}_m$ нужно решить систему из m уравнений:

$$\frac{\partial}{\partial \hat{\theta}_j} \ln L(z_1, \dots, z_n, \hat{\theta}_1, \dots, \hat{\theta}_m) = 0; \quad j = 1, \dots, m.$$

Оценки, полученные методом максимального правдоподобия, являются состоятельными и обладают рядом дополнительных оптимальных свойств.

Пример. Пусть $z_1, \dots, z_n$ — случайная выборка из нормальной генеральной совокупности с неизвестными параметрами μ и σ^2 . Функция правдоподобия имеет вид:

$$L(z_1, \dots, z_n, \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (z_i - \mu)^2}.$$

Для получения выборочной оценки математического ожидания нужно решить уравнение

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (z_i - \hat{\mu}) = 0,$$

откуда

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n z_i = \bar{z}.$$

Рассмотрим три важнейших статических распределения, основанных на нормальном законе.

Если элементы вектора $\mathbf{Y}^T = (y_1, \dots, y_n)$ представляют собой независимые нормированные нормально распределенные случайные величины, т. е.

$$\mathbf{Y} \sim N_n(0, \mathbf{I}_n),$$

то случайная величина $u = \mathbf{Y}^T \mathbf{Y} = \sum_{i=1}^n y_i^2$ имеет центральное χ^2 -распределение с n степенями свободы, т. е. $u \sim \chi^2(n)$.

Если случайные величины u и w независимы и $u \sim \chi^2(n)$, а $w \sim \chi^2(m)$, то случайная величина $v = \frac{u}{n} / \frac{w}{m}$ имеет центральное

F -распределение с n степенями свободы числителя и m степенями свободы знаменателя, т. е. $v \sim F(n, m)$.

Если ζ' — нормированная нормально распределенная случайная величина, т. е. $\zeta' \sim N_1(0, 1)$, а $w \sim \chi^2(m)$, то случайная величина $t = \zeta' / \sqrt{\frac{w}{m}}$ имеет центральное t -распределение с m степенями свободы, т. е. $t \sim t(m)$.

Теорема [Л. 7]. Если вектор $\mathbf{Y}$ независимых случайных величин $y_1, \dots, y_n$ имеет совместное нормальное распределение с параметрами $(\mathbf{M}, \sigma^2 \mathbf{I}_n)$, т. е. $\mathbf{Y} \sim N_n(\mathbf{M}, \sigma^2 \mathbf{I}_n)$, а идемпотентная матрица $\mathbf{A}(n \times n)$ имеет ранг m , то квадратичная форма $\mathbf{Y}^T \mathbf{A} \mathbf{Y} / \sigma^2$ имеет нецентральное $(\chi')^2$ -распределение с m степенями свободы и параметром нецентральности $\lambda = \frac{1}{2\sigma^2} \mathbf{M}^T \mathbf{A} \mathbf{M}$, т. е. $\mathbf{Y}^T \mathbf{A} \mathbf{Y} / \sigma^2 \sim (\chi')^2(m, \lambda)$.

Если $u \sim (\chi')^2(m, \lambda)$, а $w \sim \chi^2(n)$ и величины u и w независимы, то случайная величина $v = \frac{u}{m} / \frac{w}{n}$ имеет нецентрально F' -рас-

пределение с m и n степенями свободы и параметром нецентральности λ , т. е. $v \sim F'(m, n, \lambda)$.

Характеристики нецентральных t' -, F' - и $(\chi')^2$ -распределений, выраженные явно через гамма-функции, а также их вид можно найти в [Л. 7, 55].

Если вектор случайных величин $\mathbf{Y}^T = (y_1, \dots, y_n)$ имеет нормальное распределение, а $\mathbf{A}$ и $\mathbf{B}$ квадратные матрицы размером $(n \times n)$, то случайные величины $u = \mathbf{Y}^T \mathbf{A} \mathbf{Y}$ и $w = \mathbf{Y}^T \mathbf{B} \mathbf{Y}$ будут статистически независимы, если $\mathbf{AB} = \mathbf{0}$.

Некоторые сведения о случайных процессах

Случайный процесс $\xi(t)$ — это случайная функция времени t , т. е. такая функция, которая в каждый момент времени является случайной величиной. Всякое наблюдение случайного процесса $\xi(t)$ дает некоторую функцию $\psi(t)$, которая называется реализацией случайного процесса. Случайный процесс можно рассматривать либо как совокупность реализаций $\psi(t)$, либо как совокупность случайных величин, зависящих от параметра t .

Полным описанием случайного процесса являются совместные распределения случайных величин $\xi(t_1), \xi(t_2), \dots, \xi(t_n)$:

$$f_n(z_1, t_1; z_2, t_2; \dots; z_n, t_n) = P\{\xi(t_1) \leq z_1; \\ \xi(t_2) \leq z_2; \dots; \xi(t_n) \leq z_n\},$$

определенные для любого $n=1, 2, \dots$.

Простейший тип случайных процессов характеризуется независимостью значений $\xi(t)$ в различные моменты времени. Для такого процесса

$$f_n(z_1, t_1; z_2, t_2; \dots; z_n, t_n) = f_1(z_1, t_1) f_1(z_2, t_2) \dots f_1(z_n, t_n),$$

где $f_1(z_i, t_i)$ — одномерный закон распределения случайной величины в момент t_i .

Математическое ожидание случайного процесса в момент t_i определяется как математическое ожидание случайной величины $\xi(t_i)$:

$$M[\xi(t_i)] = \int_{-\infty}^{\infty} z_i \varphi(z_i) dz_i = m_{\xi}(t_i),$$

где $\varphi(z_i) = \frac{\partial f_1(z_i, t_i)}{\partial z_i}$ — плотность вероятности случайной величины $\xi(t_i)$.

Дисперсия случайного процесса в момент t_i

$$D[\xi(t_i)] = \int_{-\infty}^{\infty} \{z_i - M[\xi(t_i)]\}^2 \varphi(z_i) dz_i = \sigma_{\xi}^2(t_i).$$

Корреляционной функцией случайного процесса $\xi(t)$ называется среднее значение произведения случайных величин $[\xi(t_1) - m_\xi(t_1)]$ и $[\xi(t_2) - m_\xi(t_2)]$:

$$R_\xi(t_1, t_2) = M \{ [\xi(t_1) - m_\xi(t_1)] [\xi(t_2) - m_\xi(t_2)] \} = \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [z_1 - m_\xi(t_1)] [z_2 - m_\xi(t_2)] \varphi(z_1, t_1; z_2, t_2) dz_1 dz_2,$$

где

$$\varphi(z_1, t_1; z_2, t_2) = \frac{\partial^2 f_2(z_1, t_1; z_2, t_2)}{\partial z_1 \partial z_2}$$

— двумерная плотность вероятности случайных величин $\xi(t_1)$ и $\xi(t_2)$. Корреляционная функция представляет собой оценку связи между значениями случайного процесса в различные моменты времени.

Взаимная корреляционная функция $R_{\xi\eta}(t_1, t_2)$ случайных процессов $\xi(t)$ и $\eta(t)$ определяется выражением

$$R_{\xi\eta}(t_1, t_2) = M \{ [\xi(t_1) - m_\xi(t_1)] [\eta(t_2) - m_\eta(t_2)] \}.$$

Если $R_{\xi\eta}(t_1, t_2) = 0$ при любых t_1 и t_2 , то случайные процессы

$\xi(t)$ и $\eta(t)$ называются некоррелированными.

Случайный процесс называется стационарным, если его статистические характеристики не меняются при произвольном сдвиге Δ по оси t , т. е.

$$f_n(z_1, t_1; \dots; z_n, t_n) = f_n(z_1, t_1 + \Delta; \dots; z_n, t_n + \Delta),$$

где $-\infty \leq \Delta \leq \infty$ и $n=1, 2, \dots$

При этом математическое ожидание и дисперсия случайного процесса $\xi(t)$ не зависят от момента времени, т. е.

$$M[\xi(t_i)] = M[\xi(t_i + \Delta)] = m_\xi; D[\xi(t_i)] = \sigma_\xi^2,$$

а корреляционная функция зависит только от разности аргументов:

$$R_\xi(t_1, t_2) = R_\xi|t_1 - t_2| = R_\xi(\tau) = \sigma_\xi^2 f_\xi(\tau).$$

Заметим, что при $\tau = 0$, $f_\xi(\tau) = 1$ и $R_\xi(0) = \sigma_\xi^2$.

Стационарный процесс является аналогом установившегося процесса.

Стационарный случайный процесс $\xi(t)$ удовлетворяет гипотезе эргодичности (т. е. является эргодическим), если для любой функции наблюдений среднее по времени равно среднему по множеству наблюдений с вероятностью единица. Например, для математического

ожидания случайного процесса $\xi(t)$ с вероятностью единица имеем:

$$M[\xi(t)] = \int_{-\infty}^{\infty} z \varphi(z) dz = m_{\xi} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \xi(t) dt.$$

Аналогичные выражения имеют место с вероятностью единица для дисперсии и корреляционной функции эргодического процесса:

$$D[\xi(t)] = \int_{-\infty}^{\infty} (z - m_{\xi})^2 \varphi(z) dz = \sigma_{\xi}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (\xi(t) - m_{\xi})^2 dt;$$

$$\begin{aligned} R_{\xi}(\tau) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{\xi})(z_2 - m_{\xi}) \varphi(z_1, z_2, \tau) dz_1 dz_2 = \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (\xi(t) - m_{\xi})(\xi(t + \tau) - m_{\xi}) dt = \sigma_{\xi}^2 f_{\xi}(\tau). \end{aligned}$$

Достаточным условием справедливости эргодической гипотезы для процесса $\xi(t)$ является неограниченное убывание модуля $R_{\xi}(\tau)$ при $|\tau| \rightarrow \infty$.

Случайный процесс $\Xi(t)$ называется k -мерным, если он представляется совокупностью k -случайных процессов $\xi_1(t), \xi_2(t), \dots, \xi_k(t)$. Математическим ожиданием k -мерного случайного процесса является k -мерный вектор $\mathbf{m}_{\xi}$, а корреляционная функция $\mathbf{R}(\tau)$ представляет собой матрицу размером $(k \times k)$ с элементами

$$R_{ij}(\tau) = M[(\xi_i(t) - m_i)(\xi_j(t + \tau) - m_j)].$$

Если время t меняется дискретно, т. е. $t=1, 2, \dots$, то говорят о случайном процессе с дискретным временем или случайной последовательности (дискретный случайный процесс). Все приведенные выше выражения и понятия остаются в силе для процессов с дискретным временем.

Аргумент τ корреляционной функции дискретного процесса принимает значения $\tau=0, 1, 2, \dots$. Если $R_{\xi}(\tau)=0$ при $\tau>0$, то случайная последовательность $\xi(t)$ называется независимой.

ЛИТЕРАТУРА

1. Налимов В. В., Чернова И. А., Статистические методы планирования экстремальных экспериментов, изд-во «Наука», 1964.
2. Планирование эксперимента, Сб. статей под ред. Г. К. Круга, изд-во «Наука», 1966.
3. Проблемы планирования эксперимента, Сб. статей под ред. Г. К. Круга, изд-во «Наука», 1969.
4. Новые идеи в планировании эксперимента, Сб. статей под ред. В. В. Налимова, изд-во «Наука», 1969.
5. Большаков И. А., Гуткин Л. С., Левин Б. Р., Стратонович Р. Л., Математические основы современной радиоэлектроники, вып. 2, изд-во «Советское радио», 1968.
6. Рао С. Р., Линейные статистические методы и их применение, изд-во «Наука», 1968.
7. Graybill F., An introduction to the linear statistical models, New York, 1961.
8. Крамер Г., Математические методы статистики, Изд-во иностр. лит., 1948.
9. Хальд А., Математическая статистика с техническими приложениями, Изд-во иностр. лит., 1956.
10. Gorman T. W., Toman R. J., Selection of variables for fitting equations to data, Technometrics, v. 8, 1966, № 1.
11. Draper N. R., Stoneman D. M., Testing for the inclusion of variables in linear regression by a randomization technique, Technometrics, v. 8, 1966, № 4.
12. Hocking R. R., Leslie R. N., Selection of the best subset in regression analysis, Technometrics, v. 9, 1967, № 4.
13. Newton R. G., Spurrell D. J., A development of multiple regression for the analysis of routine data, Applied Statistics, v. 6, 1967, № 1, 2.
14. Лившиц Н. А., Пугачев В. И., Вероятностный анализ систем автоматического управления, изд-во «Советское радио», 1963.
15. Кулик В. Т., Принципы алгоритмизации и построения управляющих машин, изд-во «Наукова думка», Киев, 1963.
16. Айвазян С. А., Применение методов корреляционного и регрессионного анализов к обработке результатов эксперимента, «Заводская лаборатория», 1964, № 7, 8.
17. Ицкович Э. Л., Статистические методы при автоматизации производства, изд-во «Энергия», 1964.
18. Бородюк В. П., Влияние погрешностей регистрации данных на точность регрессионного уравнения, «Заводская лаборатория», 1970, № 1.
19. Бородюк В. П., Построение статического математического описания динамического объекта, Труды МЭИ, вып. 68, 1969.

20. Бородюк В. П., Критерий для оценки работоспособности математического описания, Автоматизация химических и нефтехимических производств, 1969, вып. 2.
21. Филаретов Г. Ф., К вопросу о построении нелинейной регрессионной модели по данным пассивного эксперимента, сб. «Проблемы планирования эксперимента», изд-во «Наука», 1969.
22. Jennrich R. I., Missing data correlation computations, Math. Comput., v. 16, 1962, № 80.
23. Glasser M., Linear regression analysis with missing observations among the independent variables, J. Amer. Statist. Assoc., v. 59, 1964, № 307.
24. Haitovsky Y., Missing data in regression analysis, J. Roy. Statist. Soc., ser. B, v. 30, 1968, № 1.
25. Schultz G., Iterative Berechnung der reziproken Matrix, Zeitschr. f. angew. Math. und Mech., v. 13, 1933, № 1.
26. Hotelling H., Some new methods in matrix calculations, Ann. Math. Statist., v. 14, 1943, № 1.
27. Cochran W. G., The omission or addition of an independent variate in multiple linear regression, J. Roy. Statist. Soc., ser. B, 1938, № 5.
28. Левин Б. Р., Теоретические основы статистической радиотехники, изд-во «Советское радио», 1968.
29. Вопросы статистической теории радиолокации, под ред. Г. П. Тартаковского, т. 2, изд-во «Советское радио», 1964.
30. Маслов Е. П., Оценка параметров марковских объектов, «Автоматика и телемеханика», 1964, № 1.
31. Фельдбаум А. А., Основы теории оптимальных автоматических систем, изд-во «Наука», 1966.
32. Цыпкин Я. З., Адаптация и обучение в автоматических системах, изд-во «Наука», 1968.
33. Battin R., Optimizing a statistical navigation procedure for space flight, Amer. Rocket. Soc., v. 32, November 1962, p. 1681—1696.
34. Иванов А. З., Крут Г. К., Кушелев Ю. Н., Лецкий Э. К., Свечинский В. Б., Обучающиеся системы управления, «Автоматика и телемеханика», Труды МЭИ, вып. 44, 1962.
35. Перельман И. И., Текущий регрессионный анализ и его применение в некоторых задачах автоматического управления, Известия АН СССР, ОТН, «Энергетика и автоматика», т. XXIII, 1962, № 5.
36. Kalman R. E., Design of self optimizing control system, Trans. ASME, v. 80, № 2, 1958.
37. Ли Р., Оптимальные оценки, определение характеристик и управление, изд-во «Наука», 1966.
38. Лецкий Э. К., Последовательные алгоритмы вычисления коэффициентов регрессионной модели, Проблемы планирования эксперимента, изд-во «Наука», 1969.
39. Райбман Н. С., Чадеев В. М., Адаптивные модели в системах управления, изд-во «Советское радио», 1966.
40. Божанов Э. С., Некоторые оценки скорости сходимости алгоритмов стохастической аппроксимации в задачах идентификации, Доклады научно-технической конференции по итогам научно-исследовательских работ за 1967 г., подсекция «Автоматика и телемеханика», ч. 1, изд. МЭИ, 1968.
41. Уайлд Д. Дж., Методы поиска экстремума, изд-во «Наука», 1967.

42. Бирюков А. Д., Испирьян Э. М., Круг Г. К., Лецкий Э. К., Нейман Р. Э., Отыскание метода автоматического управления процессом выделения каучука и средств его реализации, Маномеры. Химия и технология СК, вып. 3, изд. ВГУ, Воронеж, 1964.

43. Ермуратский П. В., Эффективность градиентных методов поиска при наличии помех, «Доклады научно-технической конференции по итогам научно-исследовательских работ за 1967 г., подсекция «Автоматика и телемеханика», ч. 1, изд. МЭИ, 1968.

44. Лецкий Э. К., Методы восстановления текущей характеристики функционального преобразователя, «Доклады научно-технической конференции по итогам научно-исследовательских работ за 1967 г., подсекция «Автоматика и телемеханика», ч. 1, изд. МЭИ, 1968.

45. Перельман И. И., Поляков О. А., Идентификация объекта методами ступенчатой аппроксимации, «Автоматика и телемеханика», 1968, № 10.

46. Лецкий Э. К., Ермуратский П. Е., Табличный метод восстановления текущей характеристики функционального преобразователя, Известия вузов, Электромеханика, 1967, № 7.

47. Феллер В., Введение в теорию вероятностей и ее приложения, т. 1, изд-во «Мир», 1964.

48. Лецкий Э. К., Адаптивная система слежения за изменяющимися параметрами характеристики функционального преобразователя, «Автоматика и телемеханика», 1970, № 2.

49. Chien J. T., Fu K. S., Stochastic learning of time-varying parameters in random environment, AIEEE Trans. on Systems Science and Cybernetics, 1969, № 3.

50. Романовский В. И., Основные задачи теории ошибок, Гостехиздат, 1947.

51. Виленкин С. Я., Статистические методы исследования систем автоматического регулирования, изд-во «Советское радио», 1967.

52. Уилкс С., Математическая статистика, изд-во «Мир», 1968.

ОГЛАВЛЕНИЕ

Предисловие	3
Введение	4
Глава первая. Линейная модель множественной регрессии	9
1-1. Точечные оценки параметров модели	9
1-2. Интервальные оценки параметров модели	22
1-3. Проверка основных гипотез	28
Глава вторая. Применение множественной регрессии для построения математического описания промышленного объекта	39
2-1. Вводные замечания	39
2-2. Продолжительность эксперимента и интервал съема данных	42
2-3. Влияние погрешностей регистрации статистических данных на точность математического описания	45
2-4. Оценка работоспособности математического описания	53
2-5. Особенности вычислительной процедуры	56
Глава третья. Рекуррентные алгоритмы построения математического описания дрейфующих объектов	61
3-1. Постановка задачи и описание метода решения	63
3-2. Метод текущего регрессионного анализа	68
3-3. Алгоритмы стохастической аппроксимации	72
3-4. Оценка параметра алгоритма γ по конечной выборке наблюдений	74
3-5. Оптимальный выбор параметра алгоритма γ в случае известной корреляционной функции процесса $B(i)$ и дисперсии σ^2	79
Глава четвертая. Кусочно-линейная аппроксимация характеристик дрейфующих объектов	86
4-1. Описание метода кусочно-линейной аппроксимации	86
4-2. Оценка параметров алгоритма по конечной выборке наблюдений	88
4-3. Оптимальный выбор параметров алгоритма при известных корреляционных функциях дрейфующих коэффициентов	89
Приложение	99
Литература	108

Бородюк Виталий Павлович
Лецкий Эдуард Константинович

**Статистическое описание промышленных
объектов**

Редакторы: *Е. П. Маслов, Н. А. Медведева*

Технический редактор *Т. Н. Хромова*

Корректор *З. Б. Шлайфер*

Сдано в набор 31/III 1971 г.	Т-14371
Подписано к печати 14/IX 1971 г.	Формат 84×108 ¹ / ₃₂
Бумага типографская № 2	Усл. печ. л. 5,88
Уч.-изд. л. 7,33	Тираж 7 000 экз.
Цена 37 коп.	Зак. 136

Издательство „Энергия“, Москва, М-114,
Шлюзовая наб., 10.

Московская типография № 10 Главполиграфпрома
Комитета по печати при Совете Министров СССР.
Шлюзовая наб., 10.

Цена 37 коп.

