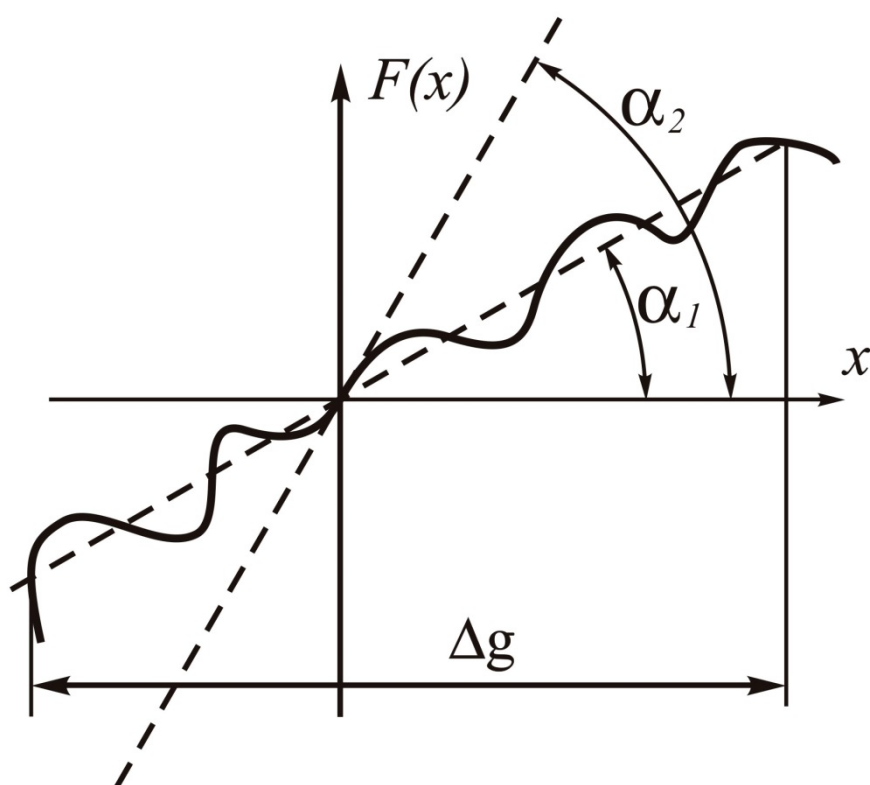


**В. Л. Бурковский,
П. В. Мурзинов,
И. М. Матвеевко**

МЕТОДЫ ИДЕНТИФИКАЦИИ ОБЪЕКТОВ СИСТЕМ УПРАВЛЕНИЯ

Учебное пособие



Воронеж 2022

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Воронежский государственный технический университет»

В. Л. Бурковский, П. В. Мурзинов, И. М. Матвеевко

МЕТОДЫ ИДЕНТИФИКАЦИИ ОБЪЕКТОВ СИСТЕМ УПРАВЛЕНИЯ

Учебное пособие

Воронеж 2022

УДК 681.324(075:8)

ББК 32.81я7

Б914

Рецензенты:

кафедра информационных технологий Воронежского института высоких технологий (зав. кафедрой, д-р техн. наук, доцент А. П. Преображенский);

С. Д. Кургалин, д-р физ.-мат. наук, профессор кафедры цифровых технологий Воронежского государственного университета

Бурковский, В. Л.

Б914 **Методы идентификации объектов систем управления:** учебное пособие [Электронный ресурс]: – Электрон. текстовые и граф. данные (2,9 Мб) / В. Л. Бурковский, П. В. Мурзинов, И. М. Матвеев; ФГБОУ ВО «Воронежский государственный технический университет». – Воронеж: Изд-во ВГТУ, 2022. – 1 электрон. опт. диск (CD-ROM): цв. – Систем. требования: ПК 500 и выше; 256 Мб ОЗУ; Windows XP; SVGA с разрешением 1024x768; Adobe Acrobat; CD-ROM; мышь. – Загл. с экрана

ISBN 978-5-7731-1018-7

Рассматриваются вопросы математического описания процессов функционирования объектов систем управления и их элементов, а также планирования, организации и обработки результатов экспериментов. Подробно изложены методы параметрической идентификации объектов систем управления различной сложности при заданных значениях структурных параметров.

Пособие предназначено для студентов бакалавриата, обучающихся по направлению подготовки 27.03.04 «Управление в технических системах», а также других специальностей при изучении методов моделирования и идентификации объектов различной физической природы.

Ил. 33. Табл. 10. Библиогр.: 14 назв.

УДК 681.324(075:8)

ББК 32.81я7

*Издается по решению редакционно-издательского совета
Воронежского государственного технического университета*

ISBN 978-5-7731-1018-7

© Бурковский В. Л., Мурзинов П. В.,
Матвеев И. М., 2022

© ФГБОУ ВО «Воронежский
государственный технический
университет», 2022

ВВЕДЕНИЕ

Отрасли науки и техники для своего нормального функционирования нуждаются в методах построения и практического использования сложных объектов, которые могут быть различной природы, иметь различное функциональное строение. Их автоматизация требует внедрения так называемых «интеллектуальных» систем управления. Такие системы характеризуются возможностью оценивать ненаблюдаемые переменные объекта, прогнозировать состояние объекта при выбираемых типах управления и автоматически синтезировать оптимальные стратегии управления.

В решениях этих задач применяют различные подходы, среди которых используются аналитические решения. Однако, многие аналитические решения очень сложны, и не всегда удается найти их решения. Вместо аналитических подходов в этих случаях используют данные экспериментальных исследований, которые требуют много времени на подготовку и проведения самих исследований. Кроме того, эти испытания требуют больших финансовых затрат и сам процесс исследования содержит вероятностную черту. Избежать этих трудностей можно, если использовать метод математического моделирования объектов, участвующих в экспериментах, и перенесение этих моделей в вычислительную технику. Поэтому идентификация и моделирование объектов в современной теории управления играют первостепенную роль.

Математическое моделирование на базе вычислительной техники позволяет исследовать поведение сложных объектов, для которых составлены соотношения и уравнения, описывающие поведение этих объектов. Изучение поведения сложных объектов на математической модели намного удобнее, чем изучать реальный физический объект.

Исследование математической модели сложного объекта называется вычислительным экспериментом. Вычислительный эксперимент позволяет получать данные о параметрах объекта для выполнения его оптимизации, прогнозирования его состояния, получать информацию скрытых параметрах, которые в реальном, физическом объекте получить невозможно. Человек, проводящий вычислительный эксперимент, является таким же исследователем, как и человек, работающий с реальным, физическим объектом.

Вычислительный эксперимент допускает изменение входных параметров, позволяет менять структуру сложного объекта. На основе выдвинутых гипотез добавляет к объекту новые элементы и исследует поведение всего объекта в целом.

Полученные количественные данные в результате вычислительного эксперимента позволяют перейти к реальному объекту и, настроив входные параметры, получить оптимальное функционирование сложного объекта. При этом математическая модель и реальный объект становятся тождественными друг другу.

Математическое моделирование и численный эксперимент могут применяться в любой области деятельности человека. Каждая область научных исследований вносит свои особенности в решении поставленных задач. При этом

используются знания из разных областей. Этими областями являются такие направления, как методы теории вероятностей и теории случайных процессов, методы математической статистики, подходы математического анализа, положения дискретной математики. Особое место занимают решения дифференциальных уравнений и решения на основе методов оптимизации.

Оформление рисунков в тексте было выполнено авторами с помощью различных графических редакторов.

1. СВЕДЕНИЯ О МЕТОДАХ МОДЕЛИРОВАНИЯ И ИДЕНТИФИКАЦИИ

1.1. Понятие модели, типы моделей и их использование

Правильным решением сложной задачи по описанию объекта является адекватный выбор необходимой модели, которая по своей внутренней структуре такая же, как и реальный объект. Модель внешне отличается от реального объекта, но имеет с позиции управления похожие входные и выходные параметры и соотношения между параметрами, имеет такую же форму.

Модели имеют различную степень сложности, и эта сложность определяется степенью точности описания объекта. Но степень точности описания объекта зависит от многих факторов: количество реальных условий, воздействующих на объект; потребности в точности решений; учет возможности цифрового управления.

Операторы, работающие с системами, используют наглядные модели, у которых отсутствуют математические соотношения. Эти модели представляют собой алгоритмы, описывающие этапы функционирования системы. Так же могут быть представлены в виде таблиц или графиков. В этом случае их называют графическими моделями. Системы автоматического управления могут представляться импульсными реакциями на единичный импульс. Такое графическое представление используется при проектировании систем автоматического управления.

Сложные системные задачи требуют применения математических моделей, которые устанавливают связь между параметрами системы в виде различных уравнений. Такой подход к системным задачам называется аналитическое моделирование. Эти модели могут быть линейными, нелинейными, дискретными, непрерывными, детерминированными и т.д.

Машинное моделирование – это программа, установленная в вычислительной технике. В этом случае допускается установка нескольких программ, взаимодействующих между собой и моделирующих различные параметры всей системы. Такой подход обеспечивает получение оптимальных решений для объекта исследования.

В настоящее время используются следующие модели: реальные или натурные модели; математические или символьные модели; статистические и динамические модели; детерминированные и динамические модели; дискретные и непрерывные модели; сосредоточенные и распределенные модели.

Вид моделей может быть следующий:

а) совокупность управляемых входных воздействий на объект:

$$u_i \in U, i = \overline{1, n}; \vec{u}(t) = (u_1(t), u_2(t), \dots, u_n(t));$$

б) совокупность неуправляемых входных воздействий:

$$z_i \in Z, i = \overline{1, m}; \vec{z}(t) = (z_1(t), z_2(t), \dots, z_m(t));$$

в) совокупность внутренних (собственных) параметров объекта:

$$x_i \in X, i = \overline{1, r}; \vec{x}(t) = (x_1(t), x_2(t), \dots, x_r(t));$$

г) совокупность выходных характеристик объекта (переменных состояния):

$$y_i \in Y, i = \overline{1, l}; \vec{y}(t) = (y_1(t), y_2(t), \dots, y_l(t));$$

Структура моделируемого объекта имеет вид представленный на рис. 1.1.

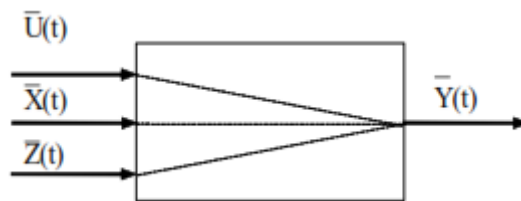


Рис. 1.1

Входные переменные $\vec{u}, \vec{x}, \vec{z}$ являются независимыми (экзогенными), а выходные \vec{y} – зависимыми (эндогенными) переменными.

Процесс функционирования объекта описывается во времени оператором F , который преобразует независимые переменные в зависимые:

$$y(t) = F(\vec{x}, \vec{u}, \vec{z}, t). \quad (1.1)$$

Совокупность зависимостей выходных характеристик объекта от времени $y_i(t), i = \overline{1, l}$ называется выходной траекторией $\vec{y}(t)$.

Соотношения (1.1) определяет закон действия объекта. Действие объекта может моделироваться различными функциями, функционалами, логическими условиями, сводом некоторых правил и т.д. Важное условие использования соотношения (1.1) – это получение выходной характеристики $\vec{y}(t)$, учитывающей входные параметры $\vec{u}(t), \vec{x}(t), \vec{z}(t)$.

Соотношения (1.1) – это математическая модель объекта, которая отражает его динамические свойства. В этом случае модель называют *динамической*. Например,

$$\frac{d\vec{y}(t)}{dt} = f(\vec{u}(t), \vec{x}(t), \vec{z}(t), \vec{y}(t)). \quad (1.2)$$

Такие модели часто возникают при проектировании технологических процессов изготовления изделий и, соответственно, средств и автоматических систем управления для этого процесса. При этом задача сводится к решению дифференциальных уравнений.

Статические модели описывают процессы, которые не изменяются во времени. Это модели, которые описывают его функционирование в установившемся режиме. Например,

$$\vec{y} = f(\vec{x}, \vec{u}, \vec{z}). \quad (1.3)$$

Статические модели используют, как правило, при проектной оптимизации объекта.

Обычно динамическая модель задается в виде дифференциальных уравнений, а статическая – в виде алгебраических или трансцендентных.

Модели, у которых существует жесткая связь между переменными, называют *детерминированными*. Такие модели не содержат случайных факторов, и значения выходных переменных однозначно определяются значениями входных переменных.

Стохастическая (вероятностная) модель отражает воздействие случайных факторов. Поэтому между входными и выходными переменными существует не функциональная зависимость (детерминированная модель), а вероятностная. Обычно переменные состояния объекта оцениваются в терминах математического ожидания, а входные воздействия – вероятностными законами распределения.

Непрерывная модель описывает непрерывные изменения переменных объекта в течении определенного промежутка времени, например,

$$\frac{dy(t)}{dt} = f(u(t), x(t), z(t), y(t)), t_j \leq t \leq t_{j+1}.$$

Дискретная модель описывает зависимость между переменными объекта в дискретные моменты времени, например,

$$y(t_{j+1}) = f(x(t_j), u(t_j), z(t_j), y(t_j)),$$

где t_j – начало j -ой стадии моделирования объекта;

t_{j+1} – ее конец, т.е. состояние объекта в момент времени t_{j+1} определяется по известному его состоянию в момент t_j при условии, что $u(t_j), z(t_j), x(t_j)$ известны и остаются постоянными.

У *линейной* модели существует пропорциональная связь между входными и выходными переменными. Модели, не удовлетворяющие этому условию, являются *нелинейными*.

Модели, у которых параметры зависят от времени, называют *динамической* моделью. Если в этой модели выходная величина зависит от одной переменной, то эту модель считают моделью с *сосредоточенными* параметрами. Например,

$$F[t, x(t), u(t), z(t), y(t), y'(t), \dots, y^{(n)}(t)] = 0. \quad (1.4)$$

Полная математическая модель наряду с дифференциальным уравнением (1.4) при решении практических задач содержит также некоторые дополнительные условия (например, значения искомым переменных y) в начальный момент времени t_0 , называемыми *начальными условиями*:

$$y(t_0) = y_0, y'(t_0) = y'_0, \dots, y^{(n-1)}(t_0) = y^{(n-1)}_0.$$

Во многих практических задачах искомая задача зависит от нескольких переменных. В этом случае математическая модель содержит частные производные и называется *моделью с распределенными параметрами*.

Если одной из независимых переменных является время t , то такая модель дает описание динамики процесса как во времени, так и в пространстве. Полная математическая модель содержит дифференциальное уравнение в частных производных, начальные условия и граничные условия если математическая модель определена в ограниченном пространстве. Примером такой модели может служить модель теплопроводности или диффузии (параболическое уравнение):

$$\frac{\partial y(t, x)}{\partial t} = a \frac{\partial^2 y(t, x)}{\partial x^2},$$

где y – параметр состояния (температура или концентрация);

t – время;

x – пространственная координата (толщина материала);

a – константа, при заданных начальных и граничных условиях.

В современном мире вопросы моделирования находят применение во многих областях. Это так же относится к области управления. Например:

– управление сложными объектами и системами (техническими, экономическими, социальными и т.д.);

– проектирование технических объектов и систем;

– прогнозирование и диагностика с использованием модели объекта;

– создание средств обучения и тренажа;

– постановка численных экспериментов на имитационной модели объекта.

Математическое моделирование позволяет оптимально управлять различными объектами, получать правильное решение в конструкторских задачах, указывать хорошее распределение ресурсов и т.д. Кроме того, математическое моделирование используется в задачах имитационного моделирования, получения информации о поведении объектов во времени, что используется в промышленности, экономике, экологии и т.д.

Модели широко применяются в качестве средств профессиональной подготовки и обучения лиц, которые должны уметь справляться с всевозможными случайностями до возникновения реальной критической ситуации. Широко известны такие применения моделей, как натурные макеты или модели космических летательных аппаратов, используемые для тренировки космонавтов, тренажеры для обучения водителей, деловые игры для обучения персонала, принимающего решения.

Применение моделей позволяет проводить контролируемые эксперименты в ситуациях, когда экспериментирование на реальных объектах практически невозможно или экономически нецелесообразно. При экспериментировании с моделью сложной системы мы часто можем узнать больше о ее внутренних взаимодействующих факторах, чем могли бы узнать, проведя эксперименты с реальной системой. Это становится возможным благодаря наблюдаемости переменных структурных элементов модели, благодаря тому, что мы можем контролировать ее поведение при различных внешних воздействиях, легко изменять ее параметры.

Резюмируя изложенное выше, отметим, что модель может служить для достижения одной из двух основных целей: либо описательной, если модель служит для объяснения и (или) лучшего понимания объекта, либо предписывающей, когда модель позволяет предсказать и (или) воспроизвести характеристики объекта, определяющие его поведение.

1.2. Идентификация как метод построения моделей

Источником для построения моделей используются данные, полученные в результате наблюдений. При этом реализуется два способа проведения наблюдений.

К первому способу относятся методы, которые могут расчленять систему на подсистемы с известными свойствами, определенными из предыдущих опытов. Это расчленение базируется на законах природы, на известных соотношениях, на опытных данных. Обладая этой информацией, можно построить математическую модель всей системы. Этот подход называется аналитическим методом. Модель, построенная таким образом, не нуждается в реальном эксперименте для проверки своей адекватности. Реальный вид процедуры моделирования определяется прикладной областью, для которой решается определенная задача. При этом осуществляется процесс структуризации процедуры построения и получается блок-схема, каждый элемент которой представляет собой простое звено. Полученную модель закладывают в вычислительную машину и получают электронную модель.

Второй способ – это использование экспериментальных данных. Эти данные представляют собой входные и выходные сигналы, обработка которых позволяет получить модель системы. Процесс получения модели таким способом называется идентификацией.

Задача идентификации. Задача идентификации формулируется следующим образом: по результатам наблюдений за входными и выходными переменными объекта построить оптимальную в некотором смысле его модель. При этом объект находится в нормальном режиме функционирования (т. е. в обстановке случайных возмущений и помех). Иными словами, если объект описывается некоторым неизвестным оператором F_0 , то, имея измеренные значения входа и выхода, необходимо построить оценку F оператора объекта, оптимальную в смысле некоторого критерия.

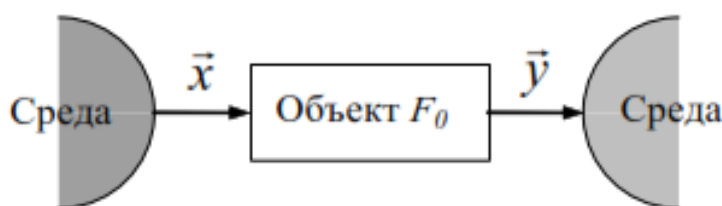


Рис. 1.2

Рис. 1.2 иллюстрирует взаимодействие идентифицируемого объекта со средой. Это взаимодействие происходит по каналам \vec{X} и \vec{Y} . По каналу \vec{X} среда воздействует на объект, а по каналу \vec{Y} объект воздействует на среду. Задача идентификации сводится к определению оператора модели F , связывающего вход и выход объекта:

$$\vec{Y} = F(\vec{X}). \quad (1.6)$$

Получить модель среды не всегда удаётся, но можно получить параметры входные и выходные, а так же ненаблюдаемые случайные факторы $\vec{Z}(t)$ возмущающего воздействия. Вход можно рассматривать как случайную функцию $\vec{X} = \vec{X}(t)$, а выходную – как функцию $\vec{Y}(t)$.

Пусть x_1, \dots, x_n – наблюдения входа объекта, а y_1, \dots, y_n – соответствующие им наблюдения его выхода в дискретные моменты времени t_1, t_2, \dots, t_n . Эти наблюдения связаны неизвестным оператором объекта F_0 , т.е. $y_i = F_0(x_i)$ ($i = \overline{1, n}$). Задача идентификации заключается в построении (синтезе) модельного оператора F , т.е. в получении оценки F по наблюдениям x_i, y_i в дискретные моменты времени t_i (в общем случае реализации $x(t), y(t)$ – непрерывны).

Таким образом, идентификация – это синтез оптимального модельного оператора F исследуемого объекта с использованием результатов наблюдений за его входными и выходными переменными.

Классификация идентификации. В соответствии с современной теорией можно предложить следующую классификацию идентификации:

1) по конечному результату идентификации:

- структурная;
- параметрическая;

2) по способу изучения объекта идентификации:

- активная;
- пассивная;

3) по типу идентифицируемой модели:

- линейная и нелинейная;
- детерминированная и стохастическая;
- с непрерывным и дискретным временем;
- стационарная и нестационарная;
- одномерная и многомерная;
- статическая и динамическая;
- с сосредоточенными и распределёнными параметрами.

Удачная идентификация какого-либо объекта определяется объемом информации о структуре этого объекта и объемом измерительной информации. Структура объекта содержит определенное количество входов и выходов, а так же вид связи между ними. Этот вид идентификации называется *структурной идентификацией*.

При структурной идентификации объем априорной информации об объекте весьма ограничен. Поэтому необходимо решить следующие задачи:

- выделение объекта из среды;
- задание класса моделей;
- определение характера связи между входом и выходом модели объекта;
- определение рационального числа информативных переменных (входов и выходов объекта), учитываемых в модели;
- определение возможности представления модели с требуемой точностью в классе линейных операторов и другие.

Объем измерительной информации позволяет решать задачу определения параметров системы. Этот объем измерительной информации называется *параметрической идентификацией*. Идентификация, реализуемая человеком, осуществляется при помощи процесса подачи входного сигнала $X(t)$ заданной формы для испытания системы. Форма этого сигнала может быть самой разнообразной: скачкообразный сигнал в виде ступеньки, сигнал в виде единичного импульса, сигнал в виде гармонических колебаний и т.д. При этом наблюдают выходной сигнал $Y(t)$, который является реакцией на входное воздействие.

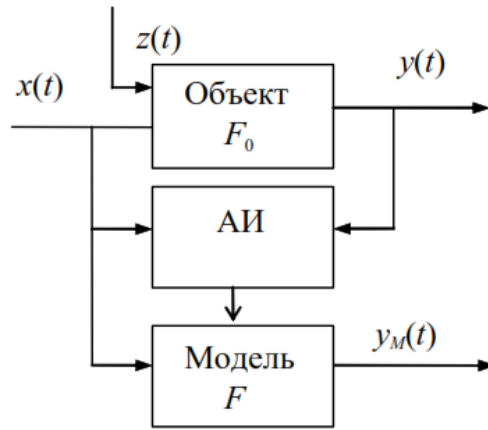


Рис. 1.3

Пассивный способ идентификации допускает принятия в качестве информации о состоянии объекта по входным $X(t)$ и выходным $Y(t)$ параметрам в процессе его функционирования.

Критерии идентификации. Формирование критерия качества, характеризующего адекватность модели реальному объекту, является одним из основных этапов идентификации.

Пусть реальный объект описывается оператором F_0 , т.е. $y(t) = F_0(x(t), z(t))$, который нельзя найти, но можно сделать его оценку. Применяя некоторый алгоритм идентификации (АИ), необходимо построить модель $y_m(t) = F(x(t))$ с оптимальным оператором F , достаточно близким к F_0 .

Но близость операторов непосредственно оценить трудно или просто невозможно, тем более что часто об операторе объекта мало известно. В связи с этим естественно оценивать близость операторов по их реакциям на одно и то же входное воздействие $x(t)$, т.е. по выходам объекта:

$$y(t) = F_0(x(t), z(t))$$

и модели $y_M(t) = F(x(t))$, где $z(t)$ – ненаблюдаемое возмущение (рис.1.3).

В общем случае $x(t)$ и $y(t)$ могут быть как детерминированными, так и случайными функциями времени.

Оптимальный оператор F модели ищется в смысле некоторого критерия, связанного с выходами $y(t)$ и $y_M(t)$. С этой целью вводится понятие функции потерь (функции невязки) $\rho[y(t), y_M(t)]$, которая в любой фиксированный момент времени t зависит от выходов объекта и модели и не зависит от оператора. Эта скалярная неотрицательная функция векторных аргументов – выходов объекта и модели. В процессе идентификации эта функция минимизируется.

Во многих практических задачах возникает проблема оценки полученной идентификации на временном отрезке $0 \leq t \leq T < \infty$. При этом добиваются ми-

нимального расхождения модельного оператора с реальным оператором, принадлежащим объекту. В результате решают задачу минимизации наибольшего отклонения модели от реального объекта. Задача минимизации может быть записана в виде (1.7)

$$\rho(y, y_M) = \max_{t \in [0, T]} |y(t) - y_M(t)| \rightarrow \min. \quad (1.7)$$

При наличии случайных возмущений и шумов, действующих на объект, в качестве критерия выбирается не само отклонение, которое так же является случайным, а его математическое ожидание (стохастический случай):

$$\rho(y, y_M) = M |y(t) - y_M(t)| \rightarrow \min. \quad (1.8)$$

Значение критериев (1.7) и (1.8) едва ли оспоримо, однако в практических расчётах они почти не используются. Это связано с недостаточной разработанностью аналитических приёмов.

Наиболее часто функция потерь используется в виде квадрата отклонения:

$$\rho(y, y_M) = \max_{t \in [0, T]} (y(t) - y_M(t))^2 \rightarrow \min; \quad (1.9)$$

$$\rho(y, y_M) = M (y(t) - y_M(t))^2 \rightarrow \min. \quad (1.10)$$

В случае наличия непрерывных реализаций входа и выхода объекта в течение интервала времени $[0, T]$:

$$\rho(y, y_M) = \int_0^T (y(t) - y_M(t))^2 dt \rightarrow \min. \quad (1.11)$$

Если реализации входа $x(t)$ и выхода $y(t)$ объекта получены в дискретные моменты $t_i (i = \overline{1, n})$, тогда

$$\rho(y, y_M) = \sum_{i=1}^n (y(t_i) - y_M(t_i))^2 \rightarrow \min. \quad (1.12)$$

На практике для сравнения, например, качества управления, применяют характеристики: временные, частотные, дискретные и т.д. В качестве критерия оценивают величину рассогласования выходных характеристик. В самонастраивающихся системах применяют специальные измерители динамики системы или объекта, что значительно усложняет систему и увеличивает время переходных процессов в системе измерения. Следовательно, предпочтительно использовать критерии, показанные выше.

Математический аппарат теории идентификации. Для идентификации детерминированных объектов принимают регулярные функции, связывающие входы и выходы объекта. Это обстоятельство и породило первый подход теории идентификации, который появился в математическом анализе в виде *теории приближения* функций многочленами и ведёт своё начало от трудов Чебышева. Это направление связано с представлением функции в виде разложения по некоторой системе функций (чаще всего по системе полиномов).

Теория приближения имеет два направления. Одно направление – это теория аппроксимации, а второе направление – это теория интерполяции. Теория интерполяции позволяет получить совпадение модели и оригинала в заданных точках.

Теория оценивания используется для идентификации стохастических объектов. При этом выполняется оценка параметров объекта, когда присутствуют случайные помехи.

Другим направлением математической статистики для целей идентификации статических стохастических объектов стала *теория планирования экспериментов*, которая рассматривает активные эксперименты с целью повышения эффективности идентификации.

Третьим подходом к решению задач идентификации являются методы *теории систем автоматического управления*. Эта теория породила специальные методы идентификации динамических объектов управления в режиме нормальной эксплуатации (т.е. в обстановке случайных возмущений и помех). Именно к этим методам вначале был применён термин «идентификация».

Основные подходы к построению алгоритмов идентификации. Возможны два подхода к построению алгоритмов идентификации:

В первом подходе в начале собирается массив данных (реализации входа и выхода объекта на рабочем отрезке времени $[0, T]$) $x_i, y_i, i = \overline{1, n}$ и оценки параметров модели получается после обработки этого массива (идентификация по массиву или ретроспективная идентификация). Место сбора данных и обработки при этом могут быть территориально разобщены.

В другой постановке оценки параметров модели уточняется на каждом шаге поступления новых измерений, т.е. по ходу процесса. Соответствующие алгоритмы и оценки называются рекуррентными (пошаговыми).

В случае, если осуществляется некоторый пошаговый процесс вычисления без поступления новых измерений, такие процедуры будем называть итерационными (их называют так же рекурсивными или последовательными приближениями).

Рекуррентные алгоритмы характерны для адаптивных систем управления, когда вследствие дрейфа параметров или недостаточной исходной точности необходимо уточнение параметров по ходу процесса.

Методы идентификации, реализующие некоторый критерий идентификации, могут быть построены как по рекуррентному, так и по ретроспективному алгоритму.

Алгоритм ретроспективный идентификации с целью оценивания вектора

параметров A может быть (для дискретных отсчетов времени $t_i, i = \overline{1, n}$) записан в форме

$$\tilde{A} = \Phi(x_0^n, y_0^n), \quad (1.13)$$

где (x_0^n, y_0^n) – массив измерений от 0 до n ;

\tilde{A} – оценка вектора A , полученная на полной выборке $0, 1, \dots, n$.

Рекуррентный алгоритм записывается в виде

$$\tilde{A}_{K+1} = \Phi_1(\tilde{A}_K, x_{K+1}, y_{K+1}), \quad (1.14)$$

где \tilde{A}_K – оценка вектора параметров на K -ом шаге;

x_{K+1}, y_{K+1} – измерения на $(K+1)$ -ом шаге.

1.3. Структурная идентификация

Структурная идентификация объекта требует определения структуры оператора F модели объекта. Если структура модели определена, то переходят к параметрической идентификации. При этом определяют параметры структуры, так как этот вид идентификации проще, чем структурная идентификация.

Структура модели – это оператор F , для которого определены его коэффициенты, совпадающие с коэффициентами структуры объекта. Но в действительности структура модели и структура объекта различаются. Кроме того, структура модели может содержать меньше каналов входа и выхода.

Входные и выходные параметры связаны между собой соотношениями, вид которых имеет различное представление. Ниже показаны виды различных соотношений, связывающие входные и выходные параметры.

Всякий одномерный статический непрерывный объект определяется функцией $y = F(x)$. Модель этого объекта можно представить в виде разложения

$$F(x) = \sum_{i=1}^k c_i \cdot f_i(x) \quad (1.15)$$

по определенной системе функций $\Phi(x) = [f_1(x), f_2(x), \dots, f_k(x)]$.

Здесь структура модели задается системой функций $\Phi(x)$ и числом k , а ее параметрами являются коэффициенты разложения c_1, c_2, \dots, c_k . Идентификация структуры такого объекта заключается в отыскании удовлетворительной системы функций $\Phi(x)$, а параметрическая идентификация сводится к определению параметров $c_i (i = 1, 2, \dots, k)$ при заданной системе функций.

Поведение детерминированного динамического одномерного объекта

удобно описывать оператором

$$F = \frac{B(p)}{A(p)}, \quad (1.16)$$

где $A(p) = 1 + \sum_{i=1}^m a_i \cdot p^i$; $B(p) = \sum_{j=1}^l b_j \cdot p^j$, $p^r = \frac{d^r}{dt^r}$ – оператор дифференцирования.

Здесь структура оператора модели определяется линейностью оператора и числами m и l . Из физических соображений $l < m$. Коэффициенты a_i ($i = \overline{1, m}$) и b_j ($j = \overline{1, l}$) являются неструктурными параметрами модели.

Оператор (1.16) эквивалентен обыкновенному дифференциальному уравнению вида

$$y + \sum_{i=1}^m a_i \frac{d^i y}{dt^i} = \sum_{j=1}^l b_j \frac{d^j x}{dt^j}. \quad (1.17)$$

Непрерывная модель (1.17) преобразуется в форму разностного уравнения, если отсчет времени происходит в дискретные равноотстоящие моменты. При таком отсчете времени непрерывные функции, описывающие поведение переменных, превращаются в решетчатые.

Для дискретного времени в момент времени n разностное уравнение записывается в форме

$$A_1(q^{-1})y[n] = B_1(q^{-1})x[n], \quad (1.18)$$

где $A_1(q^{-1}) = a_{1m}q^{-m} + a_{1(m-1)}q^{-m+1} + \dots + 1$; $B_1(q^{-1}) = b_{1l}q^{-1} + b_{1(l-1)}q^{-l+1} + \dots + b_0$;

$$q^{-r} y[n] = y[n-r].$$

При переходе от непрерывной модели (1.17) к дискретной (1.18) порядок уравнения m и l сохраняется, но значения коэффициентов для непрерывной и дискретной форм моделей различны. Если известны коэффициенты в одной из форм и интервал дискретизации, значения коэффициентов в другой форме могут быть вычислены.

В теории автоматического управления широко используются также следующие записи связи между переменными на входе и выходе:

1. Интеграл свертки (интеграл Дюамеля):

$$y(t) = \int_{-\infty}^t x(\tau)w(t-\tau)d\tau = \int_{-\infty}^t w(\tau)x(t-\tau)d\tau, \quad (1.19)$$

где $x(t), y(t)$ – входная и выходная переменные;

$w(t)$ – импульсная переходная (весовая) функция объекта, т.е. сигнал $y(t)$ на выходе объекта, если входной сигнал $x(t)$ представляет собой дельта-функцию $\delta(t)$.

$$\delta(t) = \begin{cases} 0, & t \neq 0; \\ \infty, & t = 0. \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(t)dt = 1.$$

Если $x(t) = 0$ при $t < 0$, то уравнение (1.19) можно переписать в форме

$$y(t) = \int_0^t x(\tau)w(t-\tau)d\tau = \int_0^t w(\tau)x(t-\tau)d\tau. \quad (1.20)$$

Для дискретного времени это уравнение записывается в форме

$$y[n] = \sum_{i=1}^n x[i]w[n-i] = \sum_{i=1}^n w[i]x[n-i]. \quad (1.21)$$

2. Запись связи в операторном виде

$$y(p) = w(p)x(p),$$

где $x(p), y(p)$ – изображения по Лапласу входного и выходного сигналов;

$w(p)$ – передаточная функция объекта, $w(p) = \frac{B(p)}{A(p)}$.

В более общей форме модель динамики детерминированного нелинейного непрерывного одномерного объекта может быть представлена в виде разложения по системе операторов

$$F(x) = \sum_{i=1}^k c_i F_i(x),$$

где $F_i(x) (i = \overline{1, k})$ – система нелинейных операторов.

Определение $F_i(x)$ является основной целью структурной идентификации.

Отыскание чисел $c_i (i = \overline{1, k})$ составляет задачу параметрической идентификации.

Выбор вида модели для идентификации представляет собой сложную задачу. Существует большое количество видов моделей, которые могут быть использованы для идентификации. Но должна быть выбрана такая модель, которая обеспечит наилучшее приближение к свойствам реального объекта. В этом направлении большую роль играет субъективный подход – это интуиция разработчика модели, его опыт, его способности и т.д.

При решении задач идентификации, как правило, существуют две возможности выбора структуры модели. Суть одной из них заключается в том, что после тщательного изучения объекта моделирования на основе законов физики и других достоверных знаний формируется структура модели. Такие модели являются неформальными, включают неизвестные физические параметры, параметрическая идентификация сводится к определению их значений. Другая возможность состоит в том, чтобы без всякого физического обоснования использовать стандартные формальные модели. Множество таких моделей, у которых параметры рассматриваются прежде всего как варьируемые средства подстройки моделей к имеющимся данным и не отражают физики процесса, называется черным ящиком. Множества моделей с настраиваемыми параметрами, допускающими физическую интерпретацию, называют серыми ящиками.

Неформальные модели лучше представляют свойства реального объекта, так как в этих моделях учитываются реальные физические закономерности, знание которых позволяет построить основные соотношения, являющиеся базой для создания модели. Исследователь должен знать такие процессы, как тепловые, гидравлические, пневматические, электрические и т.д.

Общий подход к составлению моделей заключается в том, что сложный реальный объект виртуально расчлениают на отдельные части, имеющие более простую структуру, чем весь объект. Для каждой части составляют модель, в которой взаимодействие с другими частями заменяют эквивалентными воздействиями.

Рассмотрим пример составления математической модели технологического объекта на основе изучения физических процессов.

Модель динамики резистивной печи.

Тепловой баланс печи резистивного нагрева складывается из изменения накопленной теплоты Q , теплового потока от нагревателя q_1 и теплового потока теплотеря q_2 вследствие теплоотдачи и конвекции:

$$\frac{dQ}{dt} = q_1 - q_2, \quad (1.24)$$

$$Q = \sum_i c_i m_i \theta, \quad (1.25)$$

где c_i – удельная теплоемкость материала i -го конструктивного элемента печи;

m_i – масса i -го элемента печи;

θ – температура.

Тепловые потери q_1, q_2 в общем случае определяются формулами:

$$q_1 = \frac{U^2}{R}; \quad q_2 = K(\theta - \theta_1), \quad (1.26)$$

где U – напряжение на нагревателе;

R – сопротивление нагревателя;

K – коэффициент теплоотдачи;

θ_1 – температура окружающей среды.

Из (1.24), (1.25) и (1.26) получаем уравнение

$$\frac{d\left(\sum_i c_i m_i \theta\right)}{dt} = \frac{U^2}{R} - K(\theta - \theta_1). \quad (1.27)$$

Учитывая, что m_i, c_i постоянные величины, из (1.27) находим уравнение

$$\frac{dQ}{dt} = \frac{U^2}{R \sum_i m_i c_i} - \frac{K}{\sum_i m_i c_i} (\theta - \theta_1), \quad (1.28)$$

которое может быть линеаризовано.

Пусть θ_0 – температура заданного режима нагрева, θ_2 – средняя температура окружающей среды. Тогда

$$\theta = \theta_0 + \Delta\theta; \quad \theta_1 = \theta_2 + \Delta\theta; \quad U = U_0 + \Delta U,$$

где U_0 – постоянные значения напряжения, необходимого для поддержания заданного постоянного значения температуры θ_0 .

В установившемся режиме

$$\frac{d\theta}{dt} = 0, \quad \Delta\theta = \Delta\theta_i = \Delta U = 0.$$

Тогда из (1.28) следует

$$0 = \frac{U^2}{R \sum_i m_i c_i} - \frac{K}{\sum_i m_i c_i} (\theta_0 - \theta_2). \quad (1.29)$$

Вычитая (1.29) из (1.28), находим

$$\frac{d\Delta\theta}{dt} = \frac{U^2 - U_0^2}{R \sum_i m_i c_i} - \frac{K}{\sum_i m_i c_i} [(\theta - \theta_0) - (\theta_1 - \theta_2)].$$

Учитывая, что

$$\begin{aligned} U^2 - U_0^2 &= (U - U_0) \cdot (U + U_0) = \Delta U (U_0 + \Delta U + U_0) = \\ &= \Delta U (2U_0 + \Delta U) = 2U_0 \Delta U + \Delta U^2, \end{aligned}$$

и пренебрегая величиной ΔU^2 , получаем

$$\begin{aligned} \frac{d\Delta\theta}{dt} &= \frac{2U_0}{R \sum_i m_i c_i} \Delta U - \frac{K}{\sum_i m_i c_i} (\Delta\theta - \Delta\theta_1); \\ \frac{d\Delta\theta}{dt} + \frac{K}{\sum_i m_i c_i} \Delta\theta &= \frac{2U_0}{R \sum_i m_i c_i} \Delta U + \frac{K}{\sum_i m_i c_i} \Delta\theta_1. \end{aligned} \quad (1.30)$$

Обозначим:

$$a_0 = \frac{K}{\sum_i m_i c_i}; \quad b_0 = \frac{2U_0}{R \sum_i m_i c_i}; \quad b_1 = \frac{K}{\sum_i m_i c_i}. \quad (1.31)$$

Тогда получаем следующую модель объекта

$$\frac{d\Delta\theta}{dt} + a_0 \Delta\theta = b_0 \Delta U + b_1 \Delta\theta_1, \quad (1.32)$$

где a_0, b_0, b_1 – параметры модели, причем, как следует из (1.30), коэффициент b_0 зависит от режима печи (от величины U_0).

Анализ работы тепловой печи позволяет сделать предположение, что температура по всем точкам печи имеет одинаковое значение. Следовательно, решение ищется относительно одной переменной, в данном случае, это температура. Поэтому печь является объектом с сосредоточенными параметрами.

Определение параметров модели расчетным путем по формулам (1.31) практически не представляется возможным, поскольку характеристики m_i, c_i i -го конструктивного элемента неизвестны, а иногда трудно поддаются к учету. Отметим еще раз, что линеаризованная модель (1.32) справедлива лишь при малых значениях ΔU (когда допустимо пренебрежение величиной ΔU^2). Параметры модели a_0, b_0, b_1 теперь могут быть определены параметрической идентификацией после выбора структуры модели (1.32).

1.4. Методы упрощения моделей

Часто построенная модель в виде дифференциального уравнения имеет сложное решение, а иногда и не разрешимое уравнение. Тогда вводят различные допущения, упрощающие построенную модель, и получают решение, которое может отличаться от свойств реального объекта. В задачах, связанных с идентификацией, поступают несколько иначе: строят достаточно простую модель, проверяют её. При неудовлетворительном результате модель начинают усложнять таким образом, чтобы можно было получить решение. Этот процесс идет до тех пор, пока свойства полученной модели не приблизятся к свойствам реального объекта.

Наиболее распространенными являются следующие методы упрощения моделей:

1. расчленение сложной системы на ряд более простых подсистем (декомпозиция);
2. выделение существенных свойств и воздействий и учет остальных в параметрической форме (метод макромоделирования);
3. линеаризация нелинейных процессов в некоторой области изменения переменных;
4. приведение систем с распределенными параметрами к системам с сосредоточенными параметрами (введение более жестких предположений и ограничений);
5. пренебрежение динамическими свойствами процессов.

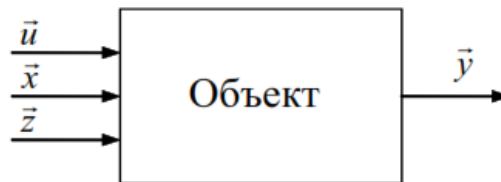


Рис. 1.4

Декомпозиция. В общем случае конечной целью декомпозиции является разбиение пространства переменных объекта (рис. 1.4.) $\{y_1, y_2, \dots, y_n, u_1, u_2, \dots, u_r, x_1, x_2, \dots, x_m, z_1, z_2, \dots, z_i\}$ на q подпространств меньшей размерности, в которых учитывается только связь данного выхода y_i с соответствующими переменными. Если любой выход имеет связь с остальными выходами, то декомпозиция практически невозможна. Если общая модель объекта имеет вид неявного выражения достаточно большой размерности

$$F(\vec{y}, \vec{u}, \vec{x}, \vec{z}) = 0 \quad (1.33)$$

и выходы y_i с объекта не имеют связи между собой, то сложную модель (1.33.) можно представить в виде совокупности эквивалентных ей n более простых частных моделей для каждого из выходов

$$\vec{y} = F(\vec{y}, \vec{u}, \vec{x}, \vec{z}), \quad (1.34)$$

где $\vec{F} = (F_1, F_2, \dots, F_n)$.

Благодаря проведенной декомпозиции системы значительно облегчается задача ее теоретического исследования.

Макромоделирование. При использовании метода макромоделирования в исходном пространстве переменных оставляются (т. е. учитываются) только те из них, которые влияют на выходные переменные наиболее сильно. Остальные неучтенные воздействия могут быть учтены в параметрической форме путем изменения коэффициентов при учтенных переменных (в случае мультипликативных воздействий) либо путем введения свободных членов (для аддитивных воздействий).

При построении упрощенных моделей с учетом только существенных факторов широко используется метод адаптивной модели, т.е. модели, коэффициенты которой подстраиваются таким образом, чтобы некоторая мера расхождения (невязка) выходов модели и объекта принимала допустимые (минимальные) значения. Для этого используют критерии минимизации невязок. Те переменные, которые стабилизируются и не приводят к изменению выходных переменных, в модели не отражаются.

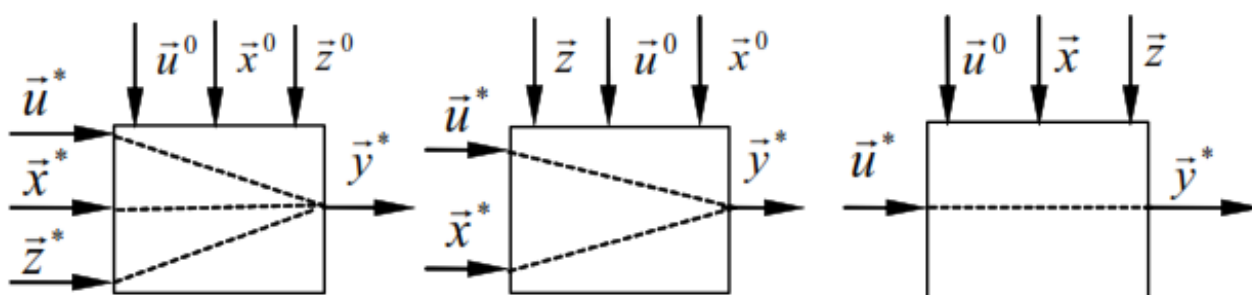


Рис. 1.5

Структура упрощенной модели, называемой макромоделью, может быть трехканальной с каналом управления u и каналами контролируемых x и неконтролируемых z воздействий, двухканальной и одноканальной (рис. 1.5).

Учет возмущений \vec{z} в двухканальных и \vec{x} , и \vec{z} в одноканальных моделях производится параметрически за счет подстройки коэффициентов оставшихся каналов.

Полная математическая модель

$$F(\vec{y}, \vec{u}, \vec{x}, \vec{z}) = 0. \quad (1.35)$$

Трехканальная макромодель

$$F(\vec{y}^*, \vec{u}^*, \vec{x}^*, \vec{z}^*, \vec{\alpha}) = 0. \quad (1.36)$$

Двухканальная макро модель

$$F(\vec{y}^*, \vec{u}^*, \vec{x}^*, \vec{\alpha}) = 0. \quad (1.37)$$

Одноканальная макро модель

$$F(\vec{y}^*, \vec{u}^*, \vec{\alpha}) = 0, \quad (1.38)$$

где $\vec{y}^* = (y_1, y_2, \dots, y_{n1})$, $\vec{u}^* = (u_1, u_2, \dots, u_{r1})$, $\vec{x}^* = (x_1, x_2, \dots, x_{m1})$, $\vec{z}^* = (z_1, z_2, \dots, z_{l1})$ – векторы контролируемых переменных, причем выполняются условия $n_1 < n$, $m_1 < m$, $r_1 < r$, $l_1 < l$, свидетельствующие о сокращении числа переменных в макро модели;

$\vec{\alpha} = (\alpha_1, \alpha_2, \dots)$ – вектор настраиваемых коэффициентов;

$\vec{u}^0, \vec{x}^0, \vec{z}^0$ – векторы неконтролируемых переменных.

Рассмотрим в качестве примера идею одного из методов адаптации моделей (компенсационный метод).

На рис. 1.6:

$\varepsilon = y - y_M$ – сигнал рассогласования выходов объекта и модели;

АИ – алгоритм идентификации.

Алгоритм идентификации позволяет настраивать модель объекта по критерию минимума ошибки ε путем изменения параметров $\alpha_1, \alpha_2, \dots$ модели (α_i^* – оптимальное значение параметра α_i).

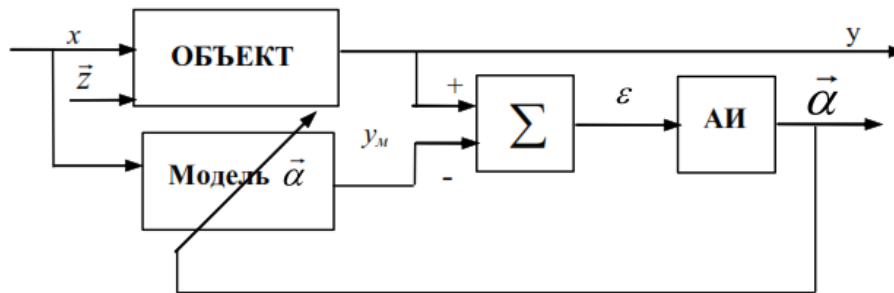


Рис. 1.6

Приведенная выше схема работает хорошо, если на ее выход подаётся сигнал без помех. При наличии шума на входе ставится задача подавления помех, которую обычно решают с помощью дифференциальных схем, содержащих полосовые фильтры (дифференциальный метод).

Линеаризация. Линеаризация исходной нелинейной модели облегчает решение конкретной задачи исследования. Поэтому для упрощения моделиро-

вания и исследования, когда это возможно, желательно заменить нелинейное уравнение приближенным линейным, решение которого с достаточной степенью точности описывает свойство исходной нелинейной системы. Процесс замены нелинейной модели линейной называется *линеаризацией*.

Если дифференциальное уравнение объекта нелинейно из-за нелинейности его статической характеристики, то для линеаризации уравнения необходимо заменить нелинейную статическую характеристику $y = F(x)$ линейной функцией $y = a_0 + a_1x$.

Основное содержание идеи линеаризации состоит в том, что различие в решениях нелинейных уравнений и их линеаризованного представления не столь существенны, чтобы приводить к недопустимым ошибкам в смысле требований к точности решения поставленной задачи.

Для линеаризации нелинейной модели $y = F(x)$ чаще всего применяют общепринятый метод малых отклонений. Техника составления линеаризованных уравнений принципиально проста. Математическое обоснование этой процедуры заключается в требованиях к виду нелинейности функции $F(x)$. Для допустимости линеаризации достаточно, что $F(x)$ и $\frac{\partial F}{\partial x}$ существуют и непрерывны в некоторой окрестности точки (x_0, y_0) . Тогда линеаризация осуществляется при помощи разложения в ряд Тейлора функции $F(x, y)$ в окрестности точки (x_0, y_0) и отбрасыванием всех нелинейных членов этого ряда

$$y = F(x_0) + \left(\frac{\partial F}{\partial x} \right)_0 (x - x_0).$$

Индекс 0 означает, что производные берутся в точке (x_0, y_0) . Таким образом, исходная нелинейная модель заменяется линейной моделью вида

$$y = a_0 + a_1x,$$

где $a_0 = F(x_0) - \left(\frac{\partial F}{\partial x} \right)_0 x_0$; $a_1 = \left(\frac{\partial F}{\partial x} \right)_0$.

Такой способ линеаризации означает замену кривой $y = F(x)$ касательной в точке (x_0, y_0) .

В случае многомерной модели, т.е. модели вида $y = F(x_1, x_2, \dots, x_m)$ получим

$$y = F(x_{10}, x_{20}, \dots, x_{m0}) + \sum_{i=10}^m \left[\left(\frac{\partial F}{\partial x} \right)_0 (x_i - x_{i0}) \right]$$

или

$$y = a_0 + \sum_{i=1}^m a_i x_i,$$

где $a_0 = F(x_{10}, x_{20}, \dots, x_{m0}) - \sum_{i=10}^m \left[\left(\frac{\partial F}{\partial x} \right)_0 x_{i0} \right]$, $a_i = \left(\frac{\partial F}{\partial x} \right)_0$, $i = \overline{1, m}$.

В этом случае поверхность, описываемая нелинейной функцией $F(x)$ в пространстве переменных x_1, x_2, \dots, x_m и y , заменяется плоскостью, касательной к поверхности в точке $(x_{10}, x_{20}, \dots, x_{m0}, y)$.

Интуитивно ясно, что линеаризованная модель, полученная при помощи разложения в ряд Тейлора, может оказаться пригодной для описания процессов в нелинейном объекте, не связанных с большими изменениями переменных в окрестности точки (x_0, y_0) . Ошибка моделирования тем меньше, чем меньше отклонения переменных.

Рассмотрим процесс линеаризации нелинейной модели при помощи разложения в ряд Тейлора на примере объекта, поведение которого описывается нелинейным дифференциальным уравнением общего вида

$$F(y'', y', y, x) = 0. \tag{1.39}$$

Если x_0, y_0 – некоторое установившееся состояние, то координаты x и y можно записать в виде

$$x = x_0 + \Delta x, \quad y = y_0 + \Delta y,$$

где Δx и Δy – отклонения координат x и y от установившегося состояния. Уравнение (1.39) в приращениях имеет вид

$$F(\Delta y'', \Delta y', y_0 + \Delta y, x_0 + \Delta x) = 0. \tag{1.40}$$

Разложим левую часть уравнения (1.40) в ряд Тейлора относительно точки установившегося состояния $(0, 0, y_0, x_0)$ и, полагая $\Delta x, \Delta y$ и их производные по времени малыми, отбросим все нелинейные члены этого ряда. При этом получаем уравнение

$$\left(\frac{\partial F}{\partial y''} \right)_0 \Delta y'' + \left(\frac{\partial F}{\partial y'} \right)_0 \Delta y' + \left(\frac{\partial F}{\partial y} \right)_0 \Delta y + \left(\frac{\partial F}{\partial x} \right)_0 \Delta x = 0,$$

которое является линейным дифференциальным с постоянными коэффициентами. Линеаризованное уравнение обычно записывают в следующем виде:

$$a_0 \Delta y'' + a_1 \Delta y' + a_2 \Delta y = b_1 \Delta x. \quad (1.41)$$

Здесь:

$$a_0 = \left(\frac{\partial F}{\partial y''} \right)_0; \quad a_1 = \left(\frac{\partial F}{\partial y'} \right)_0; \quad a_2 = \left(\frac{\partial F}{\partial y} \right)_0; \quad b_1 = - \left(\frac{\partial F}{\partial x} \right)_0.$$

Необходимым условием линеаризации является разложимость в ряд Тейлора функции $F(y'', y', y, x)$ в окрестности точки, соответствующей установившемуся состоянию.

Уравнение (1.41) приближенно заменяет уравнение (1.39) лишь в некоторой малой окрестности точки $(0, 0, y_0, x_0)$. Величина этой окрестности зависит от вида функции $F(y'', y', y, x)$, т. е. от величин производных порядка выше первого в рассматриваемой точке. В большинстве случаев при помощи линеаризованной модели (1.41) можно исследовать поведение объекта лишь при малых отклонениях входной и выходной координаты.

Представление исходной нелинейной зависимости $F(\vec{x})$ в виде суммы линейных членов разложения в ряд Тейлора предполагает известным вид этой зависимости. Только в этом случае возможным окажется вычисление величин производных $\frac{\partial F}{\partial \vec{x}}$. Для этого необходимо построение сложной нелинейной модели и подробное выписывание формул в общем виде с целью последующего преобразования, что усложняет процесс идентификации и увеличивает объем вычислительных работ.

Возможна линеаризация нелинейной функции $F(\vec{x})$ при помощи секущей плоскости в многомерном пространстве, описываемой линейным уравнением

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m,$$

коэффициенты которого определяются методом наименьших квадратов (МНК) так, чтобы получить хорошее приближение исходной и линеаризованной модели в некоторой области возможных изменений переменных.

Применение МНК к вычислению коэффициентов линеаризации существенным образом влияет на процедуру составления линейных уравнений. При таком подходе достаточно задаться структурой линеаризованного уравнения, т. е. указать состав переменных, от которых должна зависеть исследуемая величина. Отсюда легко усматривается преимущество применения МНК для целей линеаризации: далеко не всегда можно дать точный вид функциональной связи между переменными в нелинейном виде. Для статистического подхода этого не требуется, достаточно знать даже не все, а лишь в основном определяющие переменные для описания свойств реального объекта. Это позволяет отказаться от процедуры построения нелинейных уравнений с последующим их преобразо-

ванием, так как можно сразу же отыскивать линейную связь между требуемыми по условию задачи переменными объектами.

На рис. 1.7 показана геометрическая интерпретация рассматриваемых способов линеаризации, где приняты следующие обозначения: α_1 – угол наклона секущей на интервале Δg (линеаризация МНК); α_2 – угол наклона касательной в точке $x = 0$ (линеаризация разложением в ряд Тейлора).

Из рис. 1.7 хорошо видно, что при использовании МНК обнаруживается возможность дать оценку допустимого диапазона линеаризации. Основой для этого служит свойство МНК определять наилучшие значения искомых коэффициентов линейной модели в принятом диапазоне.

Упрощение модели с распределёнными параметрами. Характеристики состояния объекта могут зависеть не только от времени, но и от пространственных координат. Из множества объектов с распределёнными параметрами можно выделить объекты, параметры которых приводимы к сосредоточенным. Это такие объекты, для которых достаточно знать значения входных и выходных переменных в конечном числе фиксированных точек пространства.

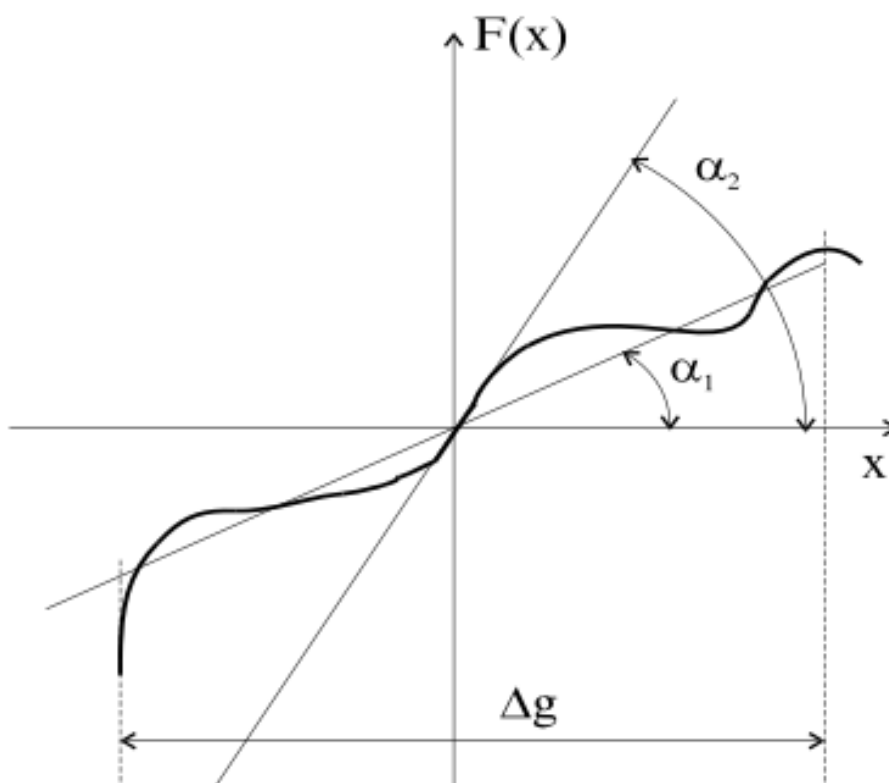


Рис. 1.7

Например, линейные объекты с распределёнными параметрами структурно могут быть представлены в виде многомерного линейного объекта с сосредоточенными параметрами. Тогда процессы в таких объектах будут описаны совокупностью математических моделей, определяющих изменения только во времени исследуемых выходных величин объектов в каждой фиксированной

точке пространства.

Контрольные вопросы

1. Дать определение модели. Классификация моделей. Описать основные типы моделей.
2. Что такое идентификация? Опишите постановку задачи идентификации.
3. Как классифицируется идентификация? Каковы критерии идентификации?
4. Опишите математический аппарат теории идентификации.
5. Какие существуют методы идентификации?
6. Опишите основные подходы к построению алгоритмов идентификации.
7. Как происходит идентификация структуры модели?
8. Какие существуют методы упрощения моделей?

2. ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ МОДЕЛИРОВАНИЯ

2.1. Адекватность модели и объекта

Объект может быть подвержен изучению непосредственно в реальных условиях. При этом на объект помещают датчики и фиксирующую аппаратуру. Снимают показания параметров объекта и проводят анализ полученной информации. Но это не всегда удобно по различным причинам. Поэтому стараются построить модель объекта такую, которая смоделирует требуемые свойства объекта. Работа с моделью проще, легче менять входные и выходные параметры, получать оптимальные решения для данной модели и, соответственно, для реального объекта.

При этом возникает проблема адекватности, то есть, насколько близко получилась модель по отношению к реальному объекту. Адекватность не может быть абсолютной, всегда будет присутствовать отличие модели от реального объекта. Поэтому полученные решения будут с некоторой погрешностью, и чем меньше погрешность, тем точнее построенная модель.

Адекватность модели оценивают на основе сравнения входных и выходных сигналов модели и реального объекта. При испытании на вход модели и реального объекта подают одинаковые сигналы, и фиксируют выходной сигнал модели и реального объекта. Если выходные сигналы близки друг другу или равны между собой, то процесс построения модели считается удачным.

Структурная схема вычисления оценки ошибки модели статического объекта приведена на рис. 2.1.



Рис. 2.1

Пусть проведено l опытов при различных уровнях входных воздействий $\vec{x}_j = (x_{1j}, x_{2j}, \dots, x_{nj})$ из области D_z их допустимых значений и получены реализации выходов объекта $\vec{y}_j = (y_{1j}, y_{2j}, \dots, y_{mj})$ и выходы модели $\vec{y}_j^M = (\vec{y}_{1j}^M, \vec{y}_{2j}^M, \dots, \vec{y}_{mj}^M), (j = \overline{1, l})$.

Ошибки модели $\vec{\Delta} = (\Delta_1, \Delta_2, \dots, \Delta_m)$ и $\vec{\delta} = (\delta_1, \delta_2, \dots, \delta_m)$ для оценки ее адекватности вычисляются по формулам:

$$\Delta_i = \max_{j=1, m} |y_{ij} - y_{ij}^M|, (i = \overline{1, m}),$$

$$\delta_i = \frac{\Delta_i}{\Delta y_i}, (i = \overline{1, m}),$$

$$\sigma_i = \sqrt{\frac{1}{l} \sum_{j=1}^l (y_{ij} - y_{ij}^M)^2}, (i = \overline{1, m}),$$

где $\Delta_i, \delta_i, \sigma_i$ – абсолютная, приведенная и среднеквадратичная ошибки модели по i -му выходу ($i = \overline{1, m}$);

y_{ij}, y_{ij}^M – значение i -го выхода объекта и модели в j -ом опыте ($i = \overline{1, m}; j = \overline{1, l}$);

Δy_i – максимальное изменение i -го выхода объекта ($i = \overline{1, m}$) при допустимых значениях входов x_k $k = \overline{1, n}$ из области D_x .

Если величины этих ошибок меньше некоторого заданного положительного числа, то модель адекватна объекту.

Теперь рассмотрим оценку модели динамического объекта. Положим, что после идентификации получена модель одномерного объекта в форме линейного дифференциального уравнения вида

$$\sum_{k=0}^n a_k \frac{d^k y_M(t)}{dt^k} = \sum_{k=0}^m b_k \frac{d^k x_M(t)}{dt^k}, \quad (2.1)$$

где $x_M(t)$ – входной сигнал модели;
 $y_M(t)$ – выходной сигнал модели;
 n, m – высшие порядки производных;
 $y_M(t)$ и $x_M(t)$ соответственно ($m \leq n$).

Пусть получены реализации входа $x(t)$ и выхода $y(t)$ объекта на отрезке времени $t \in [0, T]$, где T – длина реализации (время наблюдения). Теперь качество модели можно оценить путем сравнения $y_M(t)$ и $y(t)$ либо непосредственно на графике (визуально), либо введя некоторую формальную меру расстояния между этими сигналами.

Выходные сигналы объекта и модели при одном и том же входном сигнале различаются, так как их дифференциальные уравнения и начальные состояния не идентичны. Для оценки адекватности модели и объекта введем критерий их близости по разности выходных сигналов, т. е. реакций на один и тот же входной сигнал $x(t)$, например, следующего вида

$$I_x = \int_0^T F(y(t) - y_M(t)) dt, \quad (2.2)$$

где F – некоторая выпуклая функция.

В частности

$$F(y(t) - y_M(t)) = (y(t) - y_M(t))^2. \quad (2.3)$$

В общем случае оценка адекватности проводится для различных форм входного сигнала $x(t)$. Отсюда следует идея необходимости усреднения по входным сигналам и начальным условиям, т. е. введения операции математического ожидания оценки I_x :

$$I = M[I_x] = M \left[\int_0^T F(y(t) - y_M(t)) dt \right]. \quad (2.4)$$

Выражение выходного сигнала имеет довольно сложный вид, что затрудняет аналитическое исследование зависимости I от коэффициентов модели, поэтому вводятся и другие критерии. В частности, если уравнение модели имеет вид

$$\sum_{k=0}^n a_k \frac{d^k y_M(t)}{dt^k} = x_M(t), \quad (2.5)$$

то для оценки близости модели и объекта удобным оказывается функционал от разности входных сигналов ($x_M(t) - x(t)$) модели и объекта, обеспечивающих один и тот же выходной сигнал

$$I = M \left[\int_0^T F(x_M(t) - x(t)) dt \right], \quad (2.6)$$

при условии, что $y_m(t) = y(t)$. В этом случае выходной сигнал модели и объекта будем обозначать $y(t)$. Тогда подставляя (2.5) в (2.6), имеем

$$I = M \left[\int_0^T F \left(\sum_{k=0}^n a_k \frac{d^k y(t)}{dt^k} - x(t) \right) dt \right], \quad (2.7)$$

т.е. функционал в явном виде зависит от коэффициентов модели, что удобно для аналитического исследования.

Развивая эту идею, можно формализовать удобный функционал для общего случая модели (2.1)

$$I = M \left[\int_0^T F \left(\sum_{k=0}^n a_k \frac{d^k y(t)}{dt^k} - \sum_{k=0}^n b_k \frac{d^k x(t)}{dt^k} \right) dt \right]. \quad (2.8)$$

Выражение

$$\Delta = \sum_{k=0}^n a_k \frac{d^k y(t)}{dt^k} - \sum_{k=0}^n b_k \frac{d^k x(t)}{dt^k}, \quad (2.9)$$

называется *обобщенной ошибкой* модели. В качестве функции $F(\bullet)$, как правило, принимают квадрат обобщенной ошибки

$$I = M \left[\int_0^T \Delta^2(t) dt \right]. \quad (2.10)$$

Этот функционал удобен тем, что в явном виде зависит от параметров модели и от доступных измерению входного и выходного сигналов объекта. Однако при вычислении этого функционала возникают определенные трудности, связанные с дифференцированием сигналов $x(t)$ и $y(t)$, а также с необходимостью выполнения операции математического ожидания. Структурная схе-

ма вычисления обобщенной ошибки и оценки критерия I представлена на рис. 2.2, где $p = \frac{d}{dt}$ – оператор дифференцирования.

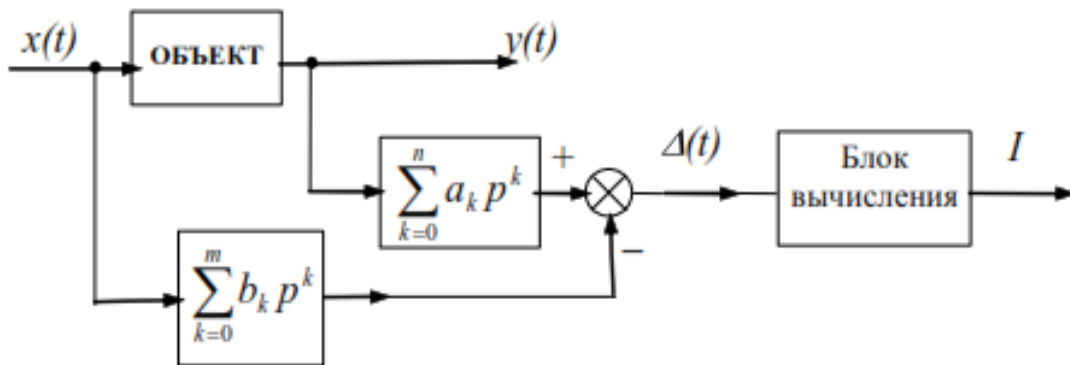


Рис. 2.2

Однако по условиям физической реализуемости можно создавать лишь устройства, порядок числителя которых меньше (или равно) порядка знаменателя, т.е. можно реализовать устройства с операторами

$$\sum_{k=0}^n a_k p^k / D(p) \text{ и } \sum_{k=0}^m b_k p^k / D(p),$$

где $D(p)$ – многочлен степени больше или равно n ; $m \leq n$.

Тогда структурная схема вычисления обобщенной ошибки $\tilde{\Delta}(t)$ и оценки критерия \tilde{I} будет иметь вид, представленной на рис. 2.3.

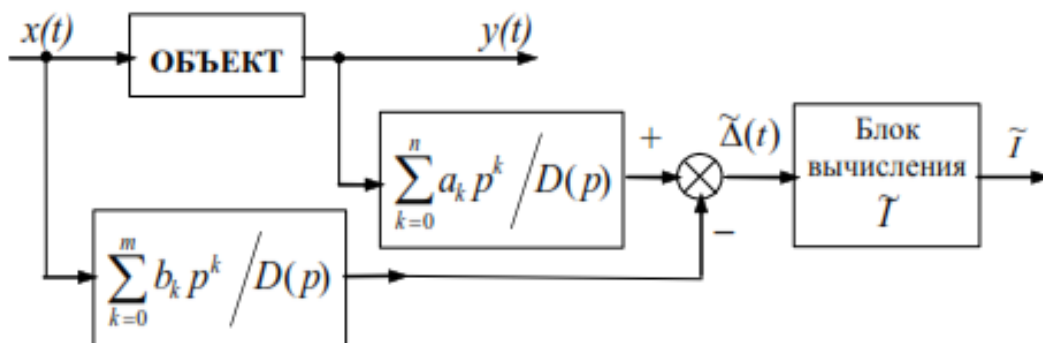


Рис. 2.3

$$\tilde{\Delta}(t) = \left(\sum_{k=0}^n a_k p^k / D(p) \right) y(t) - \left(\sum_{k=0}^m b_k p^k / D(p) \right) x(t);$$

$$\tilde{I} = M \left[\int_0^T \tilde{\Delta}^2(t) dt \right].$$

Структурной схеме, изображенной на рис. 2.3, эквивалентна схема, приведенная на рис. 2.4.

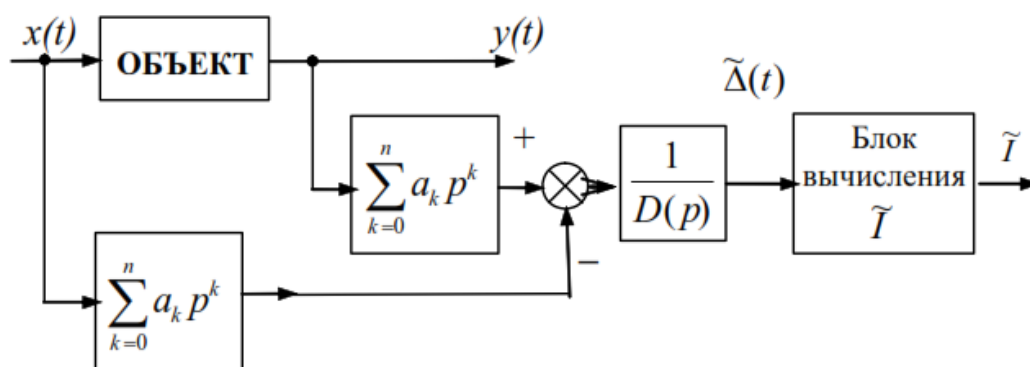


Рис. 2.4

Таким образом, обобщенная ошибка $\tilde{\Delta}(t)$, измеряемая с помощью физически реализуемых устройств, отличается от обобщенной ошибки $\Delta(t)$ тем, что $\tilde{\Delta}(t)$ является результатом преобразования $\Delta(t)$ фильтром с передаточной функцией $1/D(p)$. В силу конечности полосы пропускания этого фильтра происходят искажения сигнала обобщенной ошибки. Эти искажения будут тем меньше, чем больше полоса пропускания фильтра.

Если величина ошибок модели и оценок критериев приближения удовлетворяют требованиям к качеству модели, то модель считается адекватной объекту и может быть использована для решения задач моделирования, оптимизации и управления. В противном случае модель необходимо усовершенствовать путем изменения структуры и введения в нее неучтенных ранее факторов.

2.2. Характеристика проблем и методов моделирования объектов

Моделирование позволяет получать информацию, которая может быть перенесена на реальный объект. При этом параметры модели и объекта близки или могут отличаться на постоянную величину. Модель объекта всегда более удобна для исследования. Как правило, модель бывает отличной от оригинала физической природы. Модели выполняют с помощью вычислительной техники, специальных стендов, уменьшенных объектов, подобных реальным объектам и т.д.

В основе моделирования лежат информационные процессы, поскольку в процессе реализации модели получают информацию о данном объекте, одновременно в эксперимент с моделью вводится управляющая информация, далее происходит обработка полученных результатов, т.е. информация лежит в основе всего процесса моделирования.

В настоящее время получили применение системы моделирования, отличающиеся высокой степенью автоматизации процесса моделирования, когда наряду с программными средствами управления вычислительным экспериментом используется возможность диалогового общения исследователя с процессом мо-

делирования. Степень автоматизации процесса моделирования тесно связана с управляемостью модели. В этом смысле наличие многих управляемых параметров и переменных модели в реализованной системе моделирования дает возможность поставить широкий эксперимент и получить обширный спектр результатов.

Характеризуя проблему моделирования в целом, необходимо учитывать, что от постановки задачи моделирования до интерпретации полученных результатов существует большая группа сложных научно-технических проблем. К основным из них можно отнести:

- идентификацию реальных объектов;
- выбор вида моделей;
- получение реализаций переменных объекта;
- построение моделей и их реализацию на ЭВМ;
- взаимодействие исследователя с моделью в ходе вычислительного эксперимента;
- проверку правильности полученных в процессе моделирования результатов и выявление основных закономерностей.

В зависимости от объекта моделирования и вида используемой модели эти проблемы могут иметь разную значимость.

В одних случаях наиболее сложной оказывается идентификация, в других – проблема реализации модели на ЭВМ, особенно в случае имитационного моделирования сложных систем.

Для исследования объектов и систем управления используются различные методы моделирования:

- физическое моделирование;
- математическое моделирование на ЭВМ;
- полунатурное моделирование.

Каждый из этих методов имеет свои достоинства и недостатки. Применение того или иного метода определяется в каждом конкретном случае в зависимости от исследуемой системы и условий ее работы. При этом необходимо иметь набор правил и условий, выполнение которых обеспечивает требуемую точность изучения реального объекта по его модели. Эти правила и условия формулируются в теории подобия.

При использовании *физического моделирования* процессы, протекающие в модели и в оригинале, имеют одинаковую иногда и различную физическую природу. Физическая модель может отличаться от оригинала размерами, скоростью протекания и физической природой процессов или материалами, из которых она изготовлена. Примерами такого моделирования может служить изучение аэродинамических свойств самолета при помощи его макета в аэродинамической трубе, изучение механических колебаний с помощью электрической схемы.

Физическая модель может наиболее полно воспроизводить такие свойства оригинала, которые при теоретическом изучении не могут быть учтены в полной мере. Теоретической базой физического моделирования является теория подобия.

Теория подобия позволяет пересчитать количественные характеристики, полученные при изучении модели, в количественные характеристики оригинала.

При физическом моделировании необходимо для каждого конкретного объекта создать свою модель. Так как изготовление сложных объектов обходится, как правило, очень дорого, требует больших материальных и трудовых затрат, то изменение параметров модели часто нецелесообразно. Поэтому, несмотря на определенные преимущества, физическое моделирование находит ограниченное применение при исследовании сложных объектов и систем управления. Для этой цели широко применяется математическое моделирование.

При *математическом моделировании* исследование объекта осуществляется посредством модели, сформулированной на языке математики и реализованной, как правило, с помощью вычислительной техники.

При моделировании на ЭВМ в качестве объекта моделирования (оригинала) выступают исходные уравнения, представляющие математическую модель реального объекта, а в качестве модели – процессы, протекающие в соответствии с этими уравнениями и воспроизводимые на ЭВМ в виде «машинных решений» путем реализации программ.

Математическое моделирование объектов и систем управления осуществляется на АВМ и ЦВМ, поэтому часто такие способы называют аналоговым и цифровым моделированием.

В математическом моделировании выделяют *имитационное моделирование*, под которым понимается воспроизведение объектов с имитацией случайными величинами и случайными процессами элементов оригинала, которые не удается представить определенными математическими моделями. При этом процесс функционирования объекта (системы) представляется в виде определенного алгоритма, который реализуется на ЭВМ.

Имитационная модель имеет входные каналы, по которым подаются управляющие воздействия. Оператор обычно планирует, какой именно вид управляющего воздействия нужно подать. Если вариантов управляющих воздействий большое количество, то тогда выполняется оператором процедура поиска оптимального управляющего воздействия. Однако, если вариантов управляющих воздействий небольшое количество, то тогда используются поочередно все управляющие воздействия. В этом случае производят оценку каждого воздействия и наиболее плохие управляющие воздействия отбрасываются.

Таким образом, имитационное моделирование рассматривают так же, как управляемый эксперимент, производимый на ЭВМ. В таком эксперименте определенные математическими моделями части объекта моделирования взаимодействуют с имитирующими возмущающими воздействиями и некоторыми частями объекта генераторами случайных величин. Это взаимодействие проводится по определенным в эксперименте правилам, а результаты моделирования подвергаются статистической обработке.

Результаты имитационных экспериментов могут влиять на вид модели после завершения серии прогонов ее на ЭВМ (исключение малозначимых параметров и входов и т.д.).

Полунатурное моделирование — это моделирование с реальной аппаратурой, при котором часть системы моделируется, а остальная часть является реальной.

Применение такого метода моделирования становится необходимым в тех случаях, когда не удастся описать работу некоторых элементов системы математически. Например, при исследовании радиолокационных (РЛ) САУ летательных аппаратов в условиях помех процесс приема и передачи антенной отраженного сигнала трудно поддается математическому описанию. Поэтому приходится РЛ и другую аппаратуру, не поддающуюся математическому описанию, воспроизводить в натуре, а остальные звенья САУ — математически на ЭВМ. Естественно работа всей замкнутой схемы моделирования должна проходить в натуральном масштабе времени. Такое моделирование системы управления позволяет анализировать как всю систему в целом, так и отдельные ее элементы, исследовать влияния кинематики и динамики элементов реальной аппаратуры на работу САУ, не смотря на то, что их математическое описание отсутствует.

В процессе полунатурного моделирования удачно сочетаются достоинства математического и натурального моделирования, и может быть достигнуто оптимальное взаимодействие между вычислительными и натурными экспериментами. В настоящее время методы полунатурного моделирования эффективно применяются при проектировании разнообразных автоматических управляющих систем.

Необходимость применения таких методов возникает, если:

- объект управления находится еще в стадии проектирования, а элементы устройства управления реально существуют;
- объект не может испытываться в лабораторных условиях;
- элементы устройства управления имеют нелинейные характеристики, трения, помехи, которые при составлении их уравнений не учитывались;
- проведение натуральных экспериментов по настройке устройства управления на объекте дорого или вообще недопустимо.

Для проведения полунатурного моделирования необходимо иметь сопрягающие устройства для связи модели (ЭВМ) с внешними реальными элементами.

Структура полунатурного моделирования системы управления представлена на рис. 2.5.

Система включает также модель генератора возмущающих воздействий для выполнения имитационных экспериментов и устройство сопряжения с объектом (УСО), т.е. преобразующие элементы для связи модели объекта с реальной аппаратурой. Включение в схему моделирования дополнительных преобразующих элементов искажает результаты моделирования реального процесса. Поэтому основные свойства этих элементов должны быть учтены при обработке результатов моделирования или же при организации самого процесса моделирования. Например, если преобразующие элементы инерционны, то можно

хотя бы частично скомпенсировать инерционность введением в модель объекта форсирующих звеньев.

Цель моделирования при проектировании САУ – выбор структуры и параметров отдельных устройств автоматической системы, отвечающих требованию выполняемых ею функций и заданным показателям качества ее работы. При выбранной структуре автоматической системы моделированием определяются лишь оптимальные значения параметров устройства управления, в наибольшей степени влияющих на заданные показатели качества.

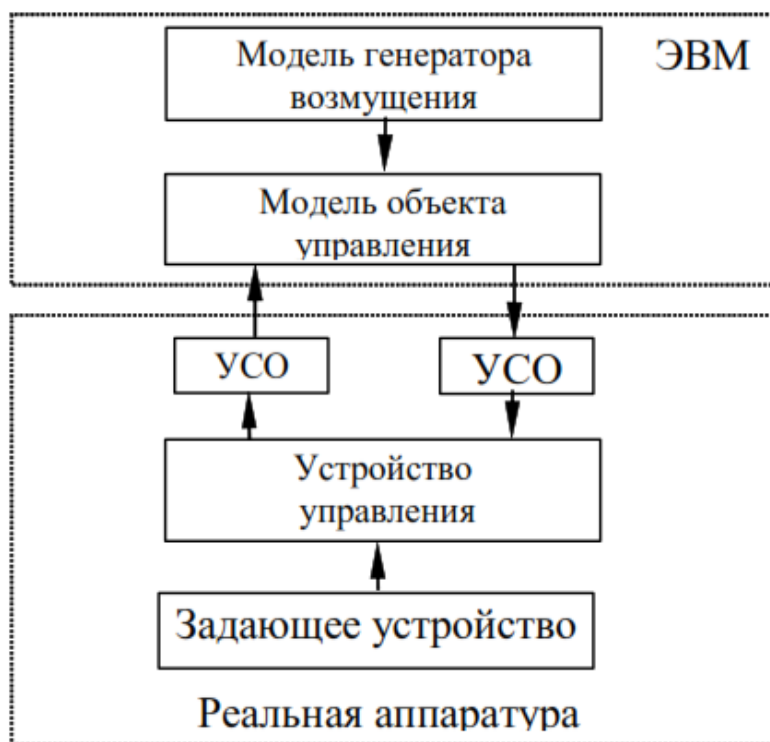


Рис. 2.5

Моделирование применяется также для исследования спроектированной автоматической системы с целью выяснения ее надежности, устойчивости, чувствительности и поведения при отказе тех или иных элементов, при различных (в том числе аварийных) возмущающих и задающих воздействиях.

При изучении процессов в реальных системах и их математических моделях различают три типа времени:

- 1) реальное, во время которого протекают процессы в реальных условиях и системах;
- 2) модельное (системное), во время которого протекают процессы в модели системы. Использование модельного времени позволяет анализировать процессы в моделях в ускоренном или замедленном темпе, а также останавливать решение для изучения процессов в какой-то момент времени;
- 3) машинное, во время которого происходит моделирование системы на ЭВМ, оно регистрируется таймером ЭВМ.

2.3. Этапы и принципы моделирования объектов и систем на ЭВМ

Имитационное моделирование, которое реализуется с помощью электронных вычислительных машин, помогает исследователям изучать сложные объекты и системы. Имитационные модели формируются последовательно, придерживаясь определенной схемы, которая имеет вид:

- 1) определение цели и постановка задачи исследования;
- 2) формализация объекта (системы) с целью построения его математической модели;
- 3) построение математической модели объекта;
- 4) разработка и составление моделирующего алгоритма;
- 5) построение машинной (имитационной) модели объекта (перевод математической модели объекта и моделирующего алгоритма в моделирующую программу и перенесение ее на машинные носители);
- 6) планирование имитационных экспериментов;
- 7) имитационные эксперименты с моделью объекта;
- 8) обработка и анализ результатов имитационных экспериментов;
- 9) документирование и практическое использование результатов моделирования.

На рис. 2.6 изображена структурная схема процесса моделирования, представляющего собой имитационное моделирование. По этой схеме можно определить, что построение модели и выполнение имитационных экспериментов не связаны между собой, так как модель, являющаяся машинной моделью, которая в процессе экспериментов выступает черным ящиком.

Результаты имитационных экспериментов могут влиять на вид модели объекта и, следовательно, на структуру моделирующего алгоритма. Например, если в процессе эксперимента выяснится, что выходные результаты слабо зависят от того или иного параметра, то это может послужить причиной упрощения модели (исключение данного параметра и соответствующее уменьшение размерности модели) и изменения моделирующего алгоритма.

В зависимости от степени формализации исследуемой системы различают:

- моделирование с применением численных методов;
- вероятностное (статистическое) моделирование с использованием специальных алгоритмических языков программирования.

Моделирование с применением численных методов осуществляют в тех случаях, когда систему удастся описать достаточно строгими математическими соотношениями. Его недостаток заключается в том, что структура моделирующего алгоритма и структура информации определяются численным методом, а не структурой и свойствами исследуемого процесса. Изменение требований к режиму моделирования, корректировка модели объекта, повышение требований к точности результатов могут привести к замене численного метода, а, следовательно, к замене алгоритма.

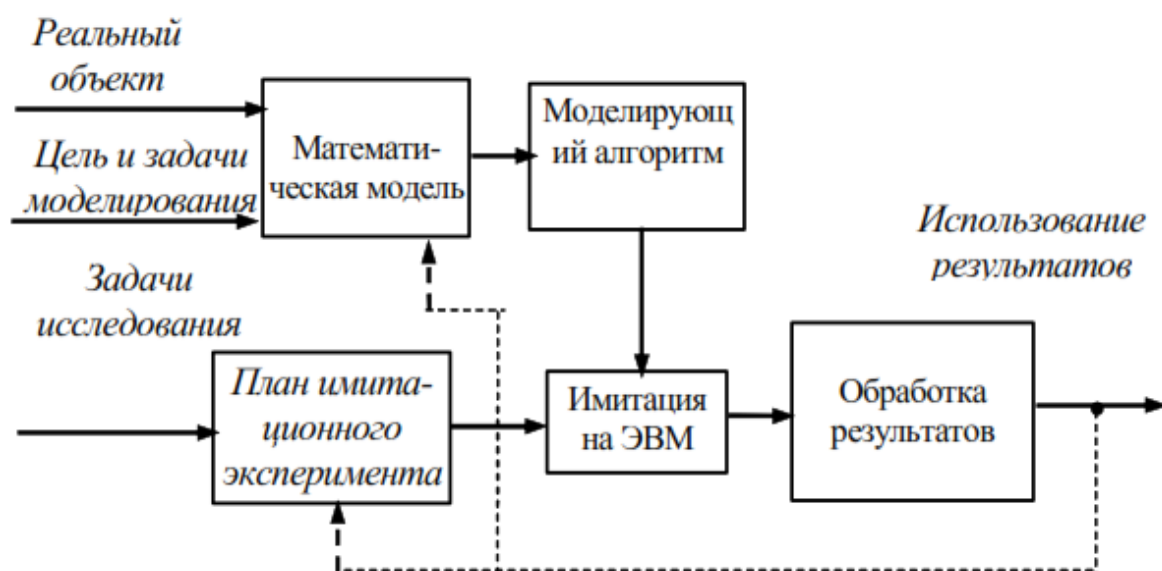


Рис. 2.6

Применение метода статистического моделирования делает моделирующий алгоритм по структуре близким к алгоритму функционирования реального объекта. Изменение условий моделирования не приводит к существенным изменениям алгоритма, который может просто дополняться новыми блоками или проигрываться большее число раз на ЭВМ.

Традиционная схема статистического моделирования включает три блока: блок имитации случайных процессов, действующих на систему; блок программы функционирования системы; блок статистической обработки результатов моделирования.

Статистические имитационные модели являются программным воспроизведением структуры реальной системы и тех элементарных действий, которые выполняют ее отдельные элементы. Недостаток статистического метода – необходимость большого количества машинного времени для его реализации.

Моделирующий алгоритм. Для моделирования процесса на ЭВМ необходимо преобразовать его математическую модель в специальный моделирующий алгоритм. Он представляет собой последовательность некоторых операций, выполняемых ЭВМ.

Запись алгоритма для моделирования сложного процесса сразу в виде программы, как правило, представляет значительные трудности. Кроме того, эта запись оказывается весьма неудобной. В самом деле, нагромождение всевозможных деталей, связанных с организацией вычислительной процедуры, делает ее мало обозримой и затрудняет ориентировку в структуре моделирующего алгоритма. Заметим, что построение моделирующего алгоритма является таким этапом исследования процесса, когда уже решены все принципиальные вопросы создания математического аппарата для исследования. Программирование задачи и решение ее на машине относятся скорее к технической части рабо-

ты. Поэтому желательно сделать запись алгоритма независимой от типа ЭВМ и языка программирования.

Как правило, моделирующие алгоритмы представляются в графической форме или в виде операторной схемы. Графической формой представления моделирующих алгоритмов являются структурные схемы.

Вся совокупность операторов, составляющих моделирующий алгоритм, может быть разделена на три группы: основные, вспомогательные и служебные операторы. Основные и вспомогательные операторы составляют главную часть моделирующего алгоритма.

Основные операторы реализуют соотношения математической модели, описывающие процессы функционирования реальных элементов системы. В отличие от них вспомогательные операторы не предназначены для имитации реального процесса. Они производят вычисление тех параметров и характеристик, которые необходимы для работы основных операторов.

Служебные операторы не связаны с соотношениями математической модели, а обеспечивают взаимодействие основных и вспомогательных операторов при моделировании процесса в автоматическом режиме, синхронизацию работы моделирующего алгоритма, фиксацию результатов и их обработку.

Каждый из рассмотренных операторов должен иметь достаточно ясный, наглядный физический смысл, связанный с природой моделируемого объекта. При этом он должен быть представлен последовательностью операций.

Операторы обозначают заглавными буквами с индексами, указывающими номер оператора. Применяется сквозная нумерация операторов, входящих в алгоритм.

Для изображения операторных схем алгоритмов удобно пользоваться операторами следующих типов:

1. **Вычислительные операторы** (A_i). Они описывают сколь угодно сложную вычислительную процедуру и должны удовлетворять следующим требованиям:

1) управление от операторов после выполнения предусмотренных им операций должно передаваться только одному оператору независимо от результатов;

2) все исходные данные, получаемые от других операторов алгоритма, к моменту начала их работы должны быть известны.

2. **Операторы формирования реализаций случайных процессов** (Φ_i). Они предназначены для имитации различных случайных факторов (случайных событий, величин, векторов, функций), сопровождающих исследуемый процесс. Исходным материалом для формирования реализаций случайных объектов на ЭВМ служат случайные числа, равномерно распределенные в интервале (0,1). Преобразование этих чисел в реализации случайных объектов с заданными свойствами осуществляет оператор (Φ_i).

3. **Операторы формирования неслучайных величин** (F_i). При моделировании сложных систем приходится формировать не только реализации случай-

ных процессов, но и различные константы и неслучайные функции. Сущность операций при этом принципиально не отличается от сущности обычных вычислительных операций, выполняемых оператором A_i . Однако, операторы A_i , как правило, являются вспомогательными, а F_i основными. Поэтому выделение операторов F_i в отдельный класс имеет смысл, поскольку они имитируют работу некоторых узлов сложных систем (чаще всего некоторого вычислительного и управляющего устройства системы).

4. **Счетчики** (K_i). Они относятся к классу служебных операторов по назначению и к арифметическим – по выполняемым функциям. Предназначены для подсчета количества различных объектов, обладающих заданными свойствами (количества циклов производственного процесса, количества деталей после определенной стадии обработки, количества бракованных изделий и т.д.). Результаты, выдаваемые счетчиками, часто являются исходными данными для логических служебных операторов, обеспечивающих синхронизацию моделирующего алгоритма. Содержимое счетчиков при моделировании может использоваться для получения вероятностных оценок исследуемых величин.

5. **Логические операторы** (P_i). Предназначены для проверки справедливости заданных условий и выработки признаков, обозначающих результат проверки. После реализации логического оператора управление передается одному из двух операторов алгоритма в зависимости от результата проверки.

6. **Оператор окончания вычислений** ($Я$). Часто с оператором $Я$ совмещают некоторые другие операции (выдачу результатов, переход к другому варианту задачи и т.д.).

При более детальном рассмотрении операторов моделирующего алгоритма могут быть выделены и другие типы операторов, обладающие некоторыми общими функциональными свойствами.

Раскроем смысловое содержание операторов, используя примеры:

A_{15} – арифметический (вычислительный) оператор № 15;

P_9 – логический оператор № 9;

A_{15}^{10} – оператор A_{15} после выполнения своей операции передает управление оператору с номером 10;

A_{16}^{30} – оператор A_{16} получает управление от оператора № 4, выполняет операцию и передает управление оператору № 30;

$A_{48}^{5,9}$ – оператор A_{48} управление передает от операторов № 5 и № 9;

$P_{35 \downarrow 12}^{\uparrow 10}$ – оператор P_{35} передает управление оператору № 10, если условие выполнено, или же оператору № 12, если оно не выполнено, получает управление от оператора № 5.

Если данный оператор получает управление от предшествующего оператора, или передает управление оператору, следующему непосредственно за ним, то верхний индекс опускается, например $A_5^6 \equiv A_5$; $A_{20} P_{21 \downarrow 30}^{\uparrow 4}$; $A_{10} P_{11 \downarrow 4} A_{12}$.

Индекс сохраняется, если последующий оператор получает управление не только от предшествующего ему оператора, но и от другого (то же и при передаче управления), например ${}^{4,9}A_{10}; A_{10}A_{11}{}^{2,11}P_{12\downarrow 7}A_{13}; {}^2A_5P_6^{\uparrow 10}F_7$.

В качестве примера построения моделирующего алгоритма рассмотрим простейший вычислительный процесс табулирования функции $y(x)$

$$y = \begin{cases} f_1(x), x < 0; \\ f_2(x), 0 \leq x \leq a. \end{cases} \quad (2.13)$$

Модель, соответствующая соотношению (2.13), порождает структуру алгоритма. Для этого процесса необходимы операторы, показанные ниже.

F_1 – формирование очередного значения независимой переменной $x_k = x_{k-1} + \Delta x$ (Δx – шаг табуляции), подстановка $x = x_k$;

P_2 – проверка условия $x < a$;

P_3 – проверка условия $x \geq 0$;

A_4 – вычисление $y = f_1(x)$;

A_5 – вычисление $y = f_2(x)$;

F_6 – формирование таблицы значений x_k, y_k и переход к следующему шагу вычислений;

$Я_7$ – конец вычислений и выдача результатов.

Имея необходимый набор операторов, можно записать операторную схему моделирующего алгоритма в следующем виде:

$${}^6F_1P_{2\downarrow 7}P_3^{\uparrow 5}A_4A_5F_6^{1,2}Я_7.$$

На рис. 2.7 представлена структурная схема этого алгоритма, позволяющая получить наглядное представление о его работе.

На различных этапах моделирования составляются укрупненные и детальные структурные схемы моделирования, а также схемы программ.

Укрупненная схема моделирующего алгоритма задает общий порядок действий при моделировании системы без каких-либо уточняющих деталей, показывает, что необходимо выполнить на очередном шаге моделирования. Детальная схема моделирующего алгоритма содержит уточнения, отсутствующие в обобщенной схеме, и показывает не только, что следует выполнить на очередном шаге моделирования системы, но и как это выполнить.

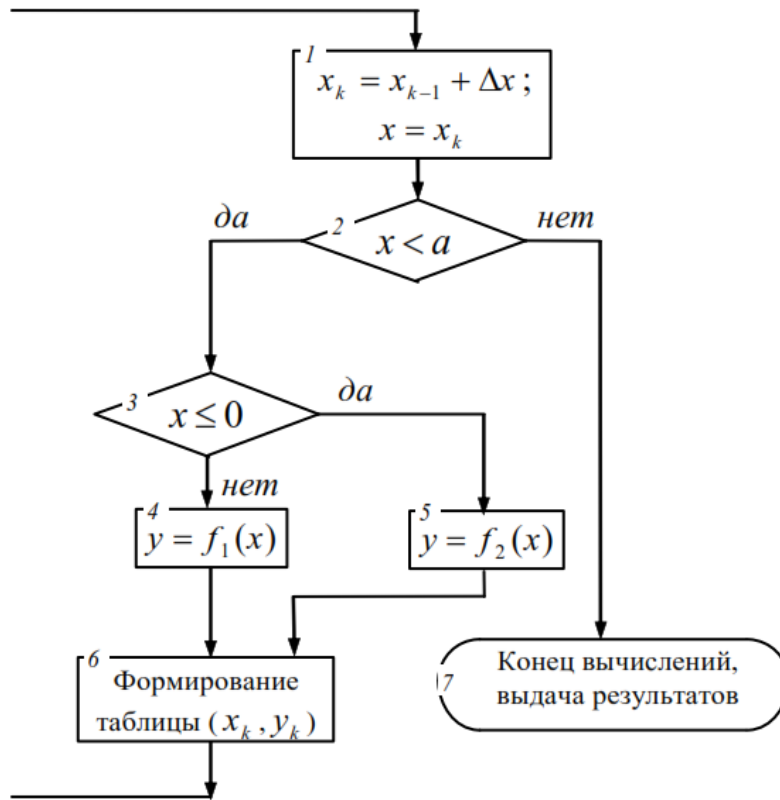


Рис. 2.7

Структурная схема программы отображает порядок программной реализации моделирующего алгоритма с использованием конкретного математического обеспечения. Схема программы представляет собой интерпретацию логической схемы моделирующего алгоритма разработчиком программы на базе конкретного алгоритмического языка и отражает логику машинной реализации модели с использованием конкретных программно-технических средств моделирования. Схема программы также может быть выполнена как в укрупненной, так и в детальной форме. Например, рис. 2.8.

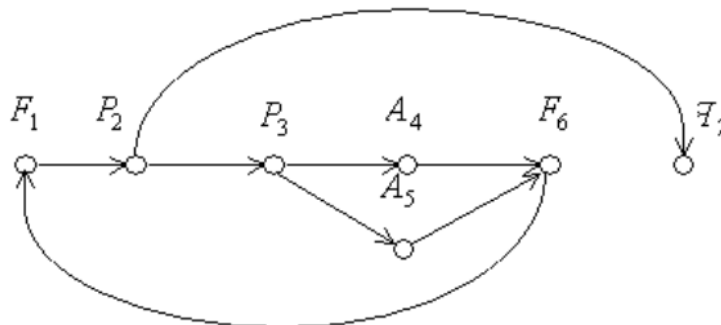


Рис. 2.8

Обычно структурная схема является наиболее удобной формой представления моделирующего алгоритма. В ряде случаев моделирующий алгоритм представляется в форме граф-схемы. В качестве пояснения к граф-схеме алгоритма в тексте дается раскрытие содержания операторов, что позволяет упростить представление алгоритма, но усложняет работу с ним. Для рассмотренного выше примера граф-схема алгоритма имеет вид, представленный на рис. 2.8. Обозначение операторов на граф-схеме соответствует обозначениям для операторной схемы.

2.4. Особенности фиксации и обработки результатов моделирования

При проведении имитационного эксперимента необходимо предусмотреть меры по организации эффективной обработки и представления его результатов. В результате эксперимента с моделью системы на ЭВМ вырабатывается информация о состояниях исследуемой системы. Эта информация является исходным материалом для определения оценок искомых характеристик. К качеству оценок, полученных в результате статистической обработки результатов моделирования, предъявляются следующие требования: оценки должны быть несмещенными, состоятельными и эффективными. Возможность получить при моделировании на ЭВМ большие выборки позволяет количественно оценить характеристики системы с учетом этих требований.

При моделировании сложных систем осуществляется многократная реализация на ЭВМ моделирующего алгоритма для получения статистически устойчивых оценок искомых характеристик. При этом накапливается значительный объем статистической информации, для хранения которой требуется большой объем памяти ЭВМ. С целью экономии памяти ЭВМ необходимо производить фиксацию результатов моделирования и их статистическую обработку непосредственно в процессе моделирования с получением текущих оценок для искомых характеристик, используя рекуррентные алгоритмы обработки, то есть без запоминания промежуточных результатов.

Если при моделировании процесса функционирования конкретной системы учитываются случайные факторы, то и среди результатов моделирования присутствуют случайные величины. В такой ситуации в качестве оценок для искомых характеристик рассчитывают средние значения, дисперсии, корреляционные моменты и другие вероятностные характеристики соответствующих случайных величин. При выборе способа фиксации и обработки результатов моделирования необходимо стремиться к использованию минимального объема памяти ЭВМ.

Для иллюстрации рациональной организации вычислительных процессов при моделировании систем на ЭВМ рассмотрим простейшие случаи определения оценок искомых характеристик систем.

Пусть в качестве искомой величины фигурирует вероятность некоторого события A . Предположим, что в результате воспроизведения N реализаций процесса имело место m случаев наступления события A . Тогда в качестве оценки

$\bar{P}(A)$ для искомой вероятности $P(A)$ можно использовать относительную частоту наступления события A

$$\bar{P}(A) = \frac{m}{N}.$$

Для оценки вероятности достаточно накапливать в памяти ЭВМ лишь числа m и N при обработке результатов моделирования.

Моделирование всегда дает результат, который подвергают обработке. Если в результате моделирования получены значения с вероятной составляющей, то возникает задача определения закона распределения случайной величины или определение функции плотности распределения случайной величины. Каждая случайная величина имеет область возможных своих значений z . Этот диапазон разбивают на n интервалов. Оценивают количество значений, попавших в каждый интервал, и для интервала с номером k вычисляют величину

$$P_k = \frac{m_k}{N}.$$

Таким образом, для оценки P_k достаточно фиксировать n значений m_k , то есть для обработки результатов моделирования требуется n ячеек памяти ЭВМ.

Для оценки среднего значения случайной величины z достаточно накапливать сумму возможных значений случайной величины z_k , ($k = \overline{1, N}$), которые она принимает при различных реализациях. Тогда среднее значение

$$\bar{z} = \frac{1}{N} \sum_{k=1}^N z_k.$$

В качестве оценки дисперсии случайной величины z при обработке результатов моделирования можно использовать

$$D_z = \frac{1}{N-1} \sum_{k=1}^N (z_k - \bar{z})^2. \quad (2.14)$$

Непосредственное вычисление дисперсии по этой формуле не рационально, так как среднее значение z изменяется в процессе накопления значений z_k . Это приводит к необходимости запоминания всех N значений, что нежелательно. Несложные преобразования (2.14) приводят к более удобной для рациональной организации вычислительного процесса формул

$$D_z = \frac{1}{N-1} \left[\sum_{k=1}^N (z_k)^2 - \frac{1}{N} \left(\sum_{k=1}^N (z_k) \right)^2 \right]. \quad (2.15)$$

Тогда для вычисления дисперсии достаточно накапливать для суммы значений z_k и их квадратов z_k^2 .

В случае оценки корреляционного момента K_{xy} для случайных величин X и Y с возможными значениями x_k, y_k ($k = \overline{1, N}$) используется выражение

$$K_{xy} = \frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{x})(y_k - \bar{y}). \quad (2.16)$$

Это выражение можно преобразовать к более удобному виду

$$K_{xy} = \frac{1}{N-1} \left(\sum_{k=1}^N x_k y_k - \frac{1}{N} \sum_{k=1}^N x_k \sum_{k=1}^N y_k \right), \quad (2.17)$$

требующих для вычисления K_{xy} накопления трех сумм.

Если при моделировании системы искомыми характеристиками являются математическое ожидание и корреляционная функция случайного процесса $z(t)$, то для нахождения оценок этих величин интервал моделирования $[0, T]$ разбивают на отрезки с постоянным шагом Δt и накапливают значения процесса $z(t)$ для фиксированных моментов времени $t = t_i = i\Delta t$.

При обработке результатов моделирования оценки для математического ожидания и корреляционной функции можно вычислить по формулам:

$$\bar{z}(t_i) = \frac{1}{N} \sum_{k=1}^N z_k(t_i), \quad (2.18)$$

$$R(u, s) = \frac{1}{N-1} \sum_{k=1}^N [z_k(u) - \bar{z}(u)][z_k(s) - \bar{z}(s)]. \quad (2.19)$$

где u и s пробегает все значения.

Для уменьшения затрат машинных ресурсов на хранение промежуточных результатов формулу (2.19) целесообразно преобразовать по аналогии с (2.17) и привести к следующему виду

$$R(u, s) = \frac{1}{N-1} \left[\sum_{k=1}^N z_k(u) z_k(s) - \frac{1}{N} \sum_{k=1}^N z_k(u) \sum_{k=1}^N z_k(s) \right]. \quad (2.20)$$

Следует отметить особенности фиксации и обработки результатов моделирования, связанные с оценкой характеристик стационарных случайных процессов, обладающих эргодическим свойством. Пусть рассматривается процесс,

для которого верны упомянутые предположения. Тогда поступают в соответствии с правилом: «среднее по времени равно среднему по множеству». Это означает, что для оценки искомым характеристик выбирается одна достаточно продолжительная реализация процесса, для которой целесообразно фиксировать результаты моделирования. Как известно, в рассматриваемом случае математическое ожидание и корреляционная функция процесса выражаются формулами:

$$\bar{z} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T z(t) dt, \quad (2.21)$$

$$R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T - \tau} \int_0^{T-\tau} z(t)z(t + \tau) dt - \bar{z}^2. \quad (2.22)$$

На практике при моделировании на ЭВМ системы интервал $[0, T]$ оказывается ограниченным и, кроме того, значения $z(t)$ удается определить только для конечного набора моментов времени t_i . Поэтому для получения оценок \bar{z} и $R(\tau)$ можно использовать приближенные формулы:

$$\bar{z} = \frac{\Delta t}{N} \sum_{i=1}^m z(t_i), \quad (2.22a)$$

$$R(\tau) = \frac{\Delta t}{T - \tau} \sum_{i=1}^{m-1} z(t_i)z(t_i + \tau) - \bar{z}^2, \quad (2.23)$$

где $m = \frac{T}{\Delta t}$; $l = \frac{\tau}{\Delta t}$; $\tau = l \cdot \Delta t$; $l = 0, 1, 2, \dots$

Исходя из соотношений (2.22a) и (2.23) можно установить целесообразный порядок фиксации и накопления результатов моделирования.

Мы затронули только простейшие случаи эффективной организации вычислительного процесса для оценок искомым характеристик при моделировании на ЭВМ.

Контрольные вопросы

1. Что такое адекватность модели и объекта? Какими параметрами она характеризуется?
2. Опишите основные методы моделирования объектов. Какие проблемы возникают при их использовании.
3. Опишите этапы и принципы моделирования объектов и систем на ЭВМ.
4. Опишите особенности обработки результатов моделирования.

3.АЛГОРИТМИЗАЦИЯ ПРОЦЕДУР ИДЕНТИФИКАЦИИ ОБЪЕКТОВ УПРАВЛЕНИЯ

3.1. Постановка задачи идентификации объектов управления

Объект идентификации в общем случае представляется в виде многополюсника, изображенного на рис. 3.1, где x_1, x_2, \dots, x_n – наблюдаемые входы объекта; z_1, z_2, \dots, z_k – ненаблюдаемые входы объекта; y_1, y_2, \dots, y_m – наблюдаемые выходы объекта.

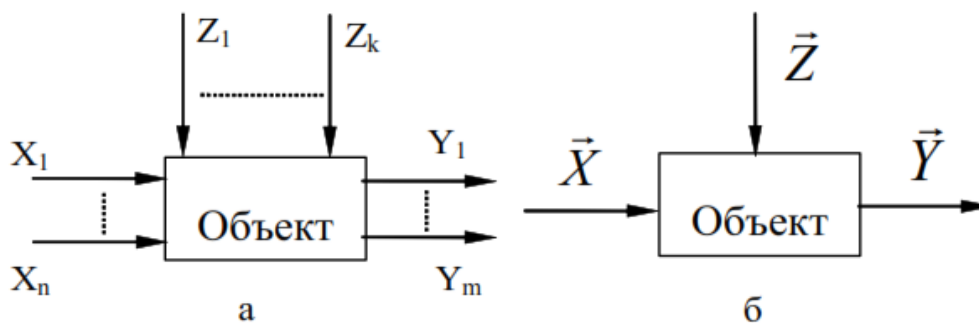


Рис. 3.1

Многомерный объект удобно представлять в векторной форме (рис. 3.1,б) где

$$\vec{x} = (x_1, x_2, \dots, x_n);$$

$$\vec{z} = (z_1, z_2, \dots, z_k);$$

$$\vec{y} = (y_1, y_2, \dots, y_m).$$

В общем случае переменные $\vec{x}, \vec{y}, \vec{z}$ являются случайными функциями времени $\vec{x} = \vec{x}(t), \vec{y} = \vec{y}(t), \vec{z} = \vec{z}(t)$.

Объект связывает входы \vec{x} и \vec{z} с выходом \vec{y} некоторым априори неизвестным оператором F_0

$$\vec{y} = F_0(\vec{x}, \vec{z}). \quad (3.1)$$

Однако идентифицируется не оператор F_0 , а оператор модели F , связывающий наблюдаемые входы и выходы:

$$\vec{y} = F(\vec{x}). \quad (3.2)$$

Ненаблюдаемый вход \vec{z} рассматривается как случайная помеха, затрудняющая определение оператора F .

Задачей идентификации является построение такого оператора модели F , который был бы в определенном смысле близок к оператору объекта F_0 , то есть $F \approx F_0$. Однако указанная близость весьма относительна, так как операторы F_0 и F могут иметь разную структуру и разное число входов. Поэтому близость операторов непосредственно оценить трудно или просто невозможно, тем более что часто об операторе F_0 мало что известно.

В связи с этим в теории идентификации близость операторов оценивают по их реакциям на одно и то же входное воздействие \vec{x}^0 , то есть по выходам объекта

$$\vec{y} = F_0(\vec{x}^0, \vec{z})$$

и модели

$$\vec{y}^M = F_0(\vec{x}^0),$$

где $\vec{y}^M = (y_1^M, y_2^M, \dots, y_m^M)$ – вектор выхода модели.

Степень близости этих реакций в каждый момент времени в общем случае оценивается функцией невязки ρ . Эта скалярная функция двух векторных аргументов – выходов объекта и модели

$$e(t) = \rho(\vec{y}(t), \vec{y}^M(t)), \quad (3.3)$$

которая обладает следующими свойствами:

$$\rho(\vec{y}(t), \vec{y}^M(t)) \geq 0 \text{ для любых } \vec{y}(t), \vec{y}^M(t);$$

$$\rho(\vec{y}(t), \vec{y}^M(t)) = 0 \text{ при } \vec{y}(t) = \vec{y}^M(t),$$

непрерывна и выпукла вниз по обоим аргументам.

Этим требованиям соответствует функция невязки

$$e(t) = |\vec{y}(t) - \vec{y}^M(t)|^2 = \sum_{i=1}^m (y_i(t) - y_i^M(t))^2, \quad (3.4)$$

которая чаще всего применяется в задачах идентификации.

Можно ввести единую меру близости выходов объекта и модели на всем интервале наблюдения $[0, T]$, а не только в каждой точке, как (3.3). Такой мерой в непрерывном случае может быть интеграл

$$E = \int_0^T \rho(\bar{y}(t), \bar{y}^M(t)) dt. \quad (3.5)$$

Значение E явно зависит от F :

$$E(F) = \int_0^T \rho(\bar{y}(t), F(\bar{x}(t))) dt, \quad (3.6)$$

и задача идентификации заключается в ее минимизации путем соответствующего выбора оператора модели F .

Следовательно, модельный оператор F должен быть таким, чтобы $\bar{y}^M \approx \bar{y}$, то есть выходы модели и объекта при одинаковых входных воздействиях должны быть эквивалентны.

Определение класса оператора, в котором ищется решение задачи идентификации, практически не поддается формализации и нуждается в эвристических решениях.

Для того, чтобы начать процедуру идентификации, необходимо иметь априорную информацию о структуре модели объекта и достаточном объеме измерительной (апостериорной) информации для определения параметров модели.

Априорная информация, которой необходимо располагать еще до наблюдения входов и выходов объекта, часто имеет качественный характер. Она должна ответить на вопрос, что представляет собой структура модели идентифицируемого объекта. Структура модели определяется в зависимости от основных свойств объекта.

Определение структуры оператора F модели составляет задачу структурной идентификации. Если же структура этого оператора определена и априори известна, то процесс идентификации сводится к определению параметров этой структуры по имеющейся измерительной информации, то есть к решению задач параметрической идентификации.

Встречаются также частные постановки задачи идентификации. Среди них чаще всего встречается задача определения характера и параметров распределения наблюдаемых сигналов Z . При этом возможны различные постановки задачи: определить, подчиняется ли наблюдаемый сигнал некоторому заданному закону распределения вероятностей; построить функцию распределения вероятностей сигнала или функцию плотности вероятности; аппроксимировать построенную функцию заданным аналитическим выражением (заданным видом закона распределения) и определить параметры этого выражения (найти функцию распределения вероятностей или плотности вероятности сигнала).

При решении задач идентификации широко используются процедуры минимизации функции или функционала и задачи оптимального планирования эксперимента.

Идентификации в данном случае подлежат вектор B и матрица A .

Модель (3.7) можно рассматривать как совокупность моделей с многомерным входом $x_j (j = \overline{1, n})$ и одномерным выходом $y, (m = 1)$.

Поэтому рассмотрим один выход объекта, то есть случай $m = 1, n > 1$.

Модель такого объекта в векторной форме имеет вид

$$y = a_0 + \langle A, \vec{x} \rangle,$$

где $\langle A, \vec{x} \rangle$ – скалярное произведение векторов $A = (a_1, a_2, \dots, a_n)$ и $\vec{x} = (x_1, \dots, x_n)$:

$$\langle A, \vec{x} \rangle = \sum_{i=1}^n a_i x_i.$$

В скалярной форме модель объекта имеет вид

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n. \quad (3.9)$$

Модель имеет $(n + 1)$ неизвестных параметров $a_i (i = \overline{1, n})$, которые подлежат оценке на основе измерений входом и выходов объекта. Эта информация обычно представляется в виде N соответствующих пар значений (\vec{x}_j, y_j) , где $j = \overline{1, n}$, $\vec{x}_j = (x_{1j}, x_{2j}, \dots, x_{nj})$ – j -е состояние входа объекта, а y_j – реакция объекта на этот вход.

Обычным подходом к решению этой задачи является приравнивание выходов объекта и модели в N заданных точках (\vec{x}_j, y_j) , в результате чего получают следующую систему уравнений идентификации:

$$a_0 + \sum_{i=1}^n a_i x_{ij} = y_j, \quad (j = \overline{1, N}). \quad (3.10)$$

Полученные уравнения с $(n + 1)$ неизвестными $(N \geq n + 1)$ имеют однозначное решение, если матрица

$$A_1 = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{n1} \\ 1 & x_{12} & x_{22} & \dots & x_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1N} & x_{2N} & \dots & x_{nN} \end{bmatrix} \quad (3.11)$$

невырождена, т.е. $\det A_1 \neq 0$ и, следовательно, ранг матрицы равен $(n + 1)$. Это возможно в том случае, если найдется $(n + 1)$ линейно независимая строка матрицы (3.11).

Поэтому из N строк следует выбрать $(n+1)$ независимых строк $(1, x_{1i}, x_{2i}, \dots, x_{mi})$, где $i \in \{1, N\}$.

В этом случае из системы (3.10) будут выделены $(n+1)$ линейно независимых уравнений

$$a_0 + \sum_{i=1}^n a_i x_{ir} = y_r, \quad (r = \overline{1, n+1}), \quad (3.12)$$

совместное решение, которых гарантирует определение точных оценок $a_i^*, (i = \overline{0, n})$ идентифицируемых параметров $a_i, (i = \overline{0, n})$ объекта, если, разумеется, объект действительно линеен. Покажем это.

Подставим с систему уравнений (3.12) уравнение объекта

$$y_r = a_0^* + \sum_{i=1}^n a_i^* x_{ir} = y_r, \quad (r = \overline{1, n+1}),$$

где $a_i^*, (i = \overline{1, n})$ – оценки параметров объекта.

Введем невязку

$$\Delta_i = a_i - a_i^*, \quad (i = \overline{0, 1}).$$

Тогда система уравнений (3.12) запишется в виде

$$\Delta_0 + \sum_{i=1}^n \Delta_i x_{ir} = 0, \quad (r = \overline{1, n+1}). \quad (3.13)$$

Для того, чтобы решение системы (3.13) было нулевым, необходимо и достаточно, чтобы определитель этой системы не был равен нулю. Легко заметить, что матрица системы (3.13) такая же, что и матрица системы (3.12), содержит $(n+1)$ линейно независимых строк матрицы (3.11) и ее определитель не равен нулю.

В результате имеем:

$$\Delta_i, \quad (i = \overline{0, n}).$$

И, следовательно, решение системы (3.12) гарантирует точную идентификацию параметров объекта, т.е.

$$a_i^* = a_i, \quad (i = \overline{0, n}).$$

Однако возможно, что объект не строго линеен и существуют незначительные случайные возмущения. При этом может оказаться, что ранг матрицы (3.11) меньше $(n+1)$ и из системы (3.10) невозможно выделить $(n+1)$ линейно независимых уравнений (3.12) для определения коэффициентов a_i ($i = \overline{0, n}$) объекта. В этом случае возможны следующие подходы к идентификации:

1) повторить измерения входов и выхода объекта в надежде, что первый эксперимент был неудачным, т.е. состояния входа \vec{x}_j ($j = \overline{1, N}$) были недостаточно разнообразны. Если и на новом экспериментальном материале не выполнится указанное условие, то можно попытаться изменить структуру модели;

2) понизить число идентифицируемых параметров, т.е. исключить рассмотрение одного из входов, например, тот, который мало изменяется. Это означает, что число идентифицируемых параметров стало n (а не $(n+1)$). Сказанное следует делать до тех пор, пока ранг матрицы (3.11) не совпадет с ее размерностью. При выполнении этого условия из системы (3.10) всегда можно выделить линейно независимые уравнения (3.12) в количестве равном числу n_1 оставшихся коэффициентов a_i ($i = \overline{0, n_1 - 1}$). Совместно решая их, находят эти коэффициенты;

3) отказаться от метода интерполяции для определения независимых коэффициентов a_i ($i = \overline{0, n_1 - 1}$), который привел к несовместной системе линейных уравнений. Ввести суммарную невязку выходов модели и объекта.

$$E = \sum_{j=1}^N \left(a_0 + \sum_{i=1}^n a_i x_{ij} - y_j \right)^2 = E(\vec{a}). \quad (3.14)$$

Величина E характеризует степень несоответствия модели и объекта и зависит от параметров a модели.

Задачу оценки параметров a_i ($i = \overline{0, n}$) можно теперь представить как задачу минимизации невязки (3.14), например, методом наименьших квадратов, т.е. свести к решению системы линейных уравнений:

$$\begin{cases} \frac{\partial E}{\partial a_0} = 2 \sum_{j=1}^N \left(a_0 + \sum_{i=1}^n a_i x_{ij} - y_j \right) = 0; \\ \frac{\partial E}{\partial a_k} = 2 \sum_{j=1}^N \left[\left(a_0 + \sum_{i=1}^n a_i x_{ij} - y_j \right) x_{kj} \right] = 0; \quad (k = \overline{1, n}). \end{cases} \quad (3.15)$$

Система (3.15) имеет следующий развернутый вид:

Во-первых, многие задачи линейной оптимизации, идентификации линейных и нелинейных моделей статики, идентификации линейных моделей динамики (дифференциальных уравнений) объекта сводятся к решению систем линейных уравнений.

Во-вторых, большинство нелинейных задач в «малом» линейны, т.е. нелинейные модели в малой окрестности некоторого решения могут быть описаны линейными. Следовательно, первым шагом решения нелинейных задач является исследование линеаризованных моделей, что также связано с решением систем линейных уравнений.

Таким образом, численные методы решения систем линейных уравнений оказываются важным инструментом решения обширного круга научно-технических задач на ЭВМ.

3.3. Идентификация статики нелинейного детерминированного объекта

Предположим, что помехи Z , действующие на выходную переменную объекта, малы или совсем отсутствуют ($\vec{Z} = 0$ в схеме на рис. 3.1), т.е. вход \vec{x} и выход \vec{y} объекта являются детерминированными сигналами (без помех), наблюдаемыми точно. Тогда выходной сигнал \vec{y} полностью определяется только наблюдаемым входным сигналом. Поведение моделируемого объекта такого рода в общем случае описывается многомерной функцией многих переменных

$$\vec{y} = F_0(\vec{x}). \quad (3.18)$$

Рассмотрим одномерный случай. Поведение нелинейного одномерного детерминированного объекта описывается нелинейной функцией одной переменной общего вида

$$y = F_0(x). \quad (3.19)$$

Априорные сведения о виде функции F_0 (оператора реального объекта) могут быть различными по объему и содержанию. Трансформация этих сведений в определенную функцию

$$y = F(x, \vec{a}) \quad (3.20)$$

с неизвестными параметрами $\vec{a} = (a_0, \dots, a_m)$ составляет предмет структурной идентификации. Основные подходы к выбору структуры модели (3.20) опираются на кусочно-линейное и функциональное представление зависимости выхода объекта от его входа, причем частным случаем последнего является разложение в ряд по заданной системе функций.

Относительно функции $F(x, \vec{a})$ возможна различная степень информированности.

1. Вид функции $F(x, \vec{a})$ известен, тогда задача поиска математической модели сводится к определению неизвестных параметров (a_0, a_1, \dots, a_m) (параметрическая идентификация).

Функция $F(x, \vec{a})$ выбирается из совокупности функций $F_1(x, \vec{a}^{(1)}), F_2(x, \vec{a}^{(2)}), \dots, F_k(x, \vec{a}^{(k)})$, в которых размерность векторов $(a^{(1)}, a^{(2)}, \dots, a^{(k)})$ различная. Требуется определить, какая из перечисленных функций является искомой (структурная идентификация) и, соответственно, найти неизвестные параметры \vec{a} (параметрическая идентификация).

Рассмотрим общую для этих случаев задачу параметрической идентификации функциональной модели, когда функция задана в аналитической форме и непрерывно дифференцируема (отметим, что кусочно-линейная модель является частным случаем функциональной).

В этом случае задача идентификации сводится к определению параметров (a_0, a_1, \dots, a_m) на основе экспериментальных наблюдений, т.е. апостериорной информации. Для этого достаточно потребовать, чтобы в каждом из зафиксированных состояний модель соответствовала объекту. Наиболее распространенным методом решения такой задачи является интерполирование и приближение.

При интерполировании предполагают, что выходы объекта и модели являются близкими, если они совпадают на заданной системе значений входа (x_0, x_1, \dots, x_n) , т.е.

$$F(x_i, \vec{a}) - y_i = 0; \quad i = \overline{1, n}, \quad (3.21)$$

где $n \geq m + 1$.

Выражение (3.21) в общем случае является системой нелинейных уравнений относительно неизвестных параметров. Наличие шумов и погрешностей измерения, как правило, делает систему уравнений (3.21) несовместной и затрудняет ее решение. Поэтому наиболее употребительным является метод приближения.

Для приближения характерна минимизация некоторого функционала, характеризующего различие $F(x, \vec{a})$ и $F_0(x)$ во всем промежутке изменения независимого переменного x . В практических приложениях чаще всего используется дискретное квадратичное приближение (метод наименьших квадратов), при котором задача идентификации представляется в виде минимизации функционала

$$E = \sum_{i=1}^n [F(x_i, \vec{a}) - y_i]^2 \rightarrow \min_{\vec{a}}, \quad (3.22)$$

равного сумме квадратов отклонений выходов объекта и модели на заданной системе точек (x_0, x_1, \dots, x_n) .

Очевидно, что величина E характеризует степень несоответствия модели и объекта и зависит от параметров модели (a_0, a_1, \dots, a_m) . Эти коэффициенты надо подобрать так, чтобы величина E была наименьшей. Минимум E при правильном выборе структуры модели F равен нулю, т.е. модель, может быть адекватной объекту. Если же структура модели отличается от структуры объекта, то минимум E не будет равен нулю, тем не менее задача синтеза модели по-прежнему сводится к задаче минимизации функции невязки E .

Необходимым условием минимума $E(x, a_0, a_1, \dots, a_m)$ является выполнение равенства

$$\frac{\partial E(x_i, \vec{a})}{\partial a_j} = 0, \quad j = \overline{0, m}, \quad (3.23)$$

или

$$\begin{cases} 2 \sum_{i=1}^n [F(x_i, a_0, \dots, a_m) - y_i] \frac{\partial F(x_i, a_0, \dots, a_m)}{\partial a_0} = 0; \\ \dots \dots \dots \\ 2 \sum_{i=1}^n [F(x_i, a_0, \dots, a_m) - y_i] \frac{\partial F(x_i, a_0, \dots, a_m)}{\partial a_m} = 0. \end{cases} \quad (3.24)$$

После преобразования

$$\begin{cases} \sum_{i=1}^n \left[F(x_i, a_0, \dots, a_m) \frac{\partial F(x_i, a_0, \dots, a_m)}{\partial a_0} \right] - \sum_{i=1}^n \left[y_i \frac{\partial F(x_i, a_0, \dots, a_m)}{\partial a_0} \right] = 0; \\ \dots \dots \dots \\ \sum_{i=1}^n \left[F(x_i, a_0, \dots, a_m) \frac{\partial F(x_i, a_0, \dots, a_m)}{\partial a_m} \right] - \sum_{i=1}^n \left[y_i \frac{\partial F(x_i, a_0, \dots, a_m)}{\partial a_m} \right] = 0. \end{cases} \quad (3.25)$$

Система уравнений (3.25) содержит столько же уравнений, сколько неизвестных коэффициентов (a_0, a_1, \dots, a_m) , и называется системой нормальных уравнений.

Величина $E \geq 0$ при любых (a_0, a_1, \dots, a_m) , следовательно, у нее обязательно должен существовать хотя бы один минимум. Поэтому, если система нормальных уравнений (3.25) имеет единственное решение, то оно и является минимумом для величины E .

Система уравнений (3.25) в общем случае является системой трансцендентных уравнений относительно искомых параметров \vec{a} . Поэтому для решения такой системы часто обращаются к поисковым методам минимизации.

Применение поисковых методов для минимизации функции невязки открывает большие возможности перед идентификацией. Однако поиска можно избежать, если идентифицируемая модель линейна относительно искомых параметров \vec{a} .

Модели такого рода являются частным случаем функциональных моделей и образуются в результате разложения искомой функции по заданной системе функций.

Рассмотрим вкратце такой случай. Пусть модель $F(x, \vec{a})$ линейна относительно неизвестных параметров и имеет вид

$$F(x, \vec{a}) = a_0 f_0(x) + a_1 f_1(x) + \dots + a_m f_m(x) = \sum_{j=1}^m a_j f_j(x), \quad (3.26)$$

где a_j – искомые коэффициенты модели, $f_j(x)$ – известные непрерывно-дифференцируемые функции входной переменной.

Система функций $f_j(x)$ должна быть линейно независимой, т.е. никакая линейная комбинация ее элементов не может быть равна нулю при любых x .

В результате эксперимента можно получить n наблюдений

$$y_i = F_0(x_i); \quad i = \overline{1, n}. \quad (3.27)$$

Для определения неизвестных коэффициентов \vec{a} воспользуемся методом наименьших квадратов. В этом случае функция невязки (3.22) после подстановки в нее выражение (3.26) принимает вид

$$E = \sum_{i=1}^n \left[\sum_{j=1}^m a_j f_j(x_i) - y_i \right]^2 \rightarrow \min_{\vec{a}}. \quad (3.28)$$

Теперь найдем неизвестные параметры a_j модели из необходимых условий минимума квадратичной функции относительно неизвестных параметров a_j ($j = \overline{1, n}$):

$$\frac{\partial E(x, \vec{a})}{\partial a_j} = \frac{\partial}{\partial a_j} \left\{ \sum_{i=1}^n \left[\sum_{j=1}^m a_j f_j(x_i) - y_i \right]^2 \right\} = 0; \quad j = \overline{0, m}. \quad (3.29)$$

Систему уравнений (3.32) можно сразу разрешить, если функции $f_0(x), f_1(x), K, f_m(x)$ выбрать в виде ортогональных полиномов, обладающих свойством:

$$\begin{cases} \sum_{i=1}^n f_j(x_i)f_l(x_i) = 0; & j, l = \overline{0, m}; \quad j \neq l; \\ \sum_{i=1}^n [f_j(x_i)]^2 \neq 0; & j = \overline{0, m}. \end{cases} \quad (3.33)$$

Последнее означает, что хотя бы в одной из точек $x_i, i = \overline{1, n}$ многочлен $f_j(x) \neq 0$. Такие многочлены называют ортогональными многочленами Чебышева.

При выполнении условий (3.33) в левой части каждого из уравнений системы (3.30) остается только один член (диагональные элементы), и эта система распадается на $m + 1$ уравнений с одним неизвестным в каждом. Поэтому сразу можно написать выражение для искомых коэффициентов

$$a_j = \frac{\sum_{i=1}^n f_j(x_i)y_i}{\sum_{i=1}^n [f_j^2(x_i)]^2}; \quad j = \overline{0, m}. \quad (3.34)$$

Многочлены Чебышева определяются по формулам:

$$\begin{aligned} f_0(x) &= 1; \\ f_1(x) &= x - \frac{n+1}{2}; \\ &\dots\dots\dots \\ f_{m+1}(x) &= f_1(x)f_m(x) - \frac{m^2(n^2 - m^2)}{4(4m^2 - 1)} f_{m-1}. \end{aligned} \quad (3.35)$$

Из формулы (3.35) следует, что полином $f_{m+1}(x)$ зависит только от объема выборки n . Следовательно, вычисленные по формуле (3.34) коэффициенты a_j не зависят от того, каков будет порядок модели (3.26). Зная многочлены Чебышева $f_{m+1}(x)$, при каждом увеличении степени уравнения статики объекта на единицу используют все ранее найденные коэффициенты a_j ($j = \overline{0, m}$), вычисляют только один коэффициент a_{m+1} . Эта особенность делает полезным использование ортогональных полиномов Чебышева при решении задач идентификации объекта путем последовательного улучшения модели.

Несмотря на преимущества использования полиномов Чебышева, для практических целей, как правило, ограничиваются степенным разложением

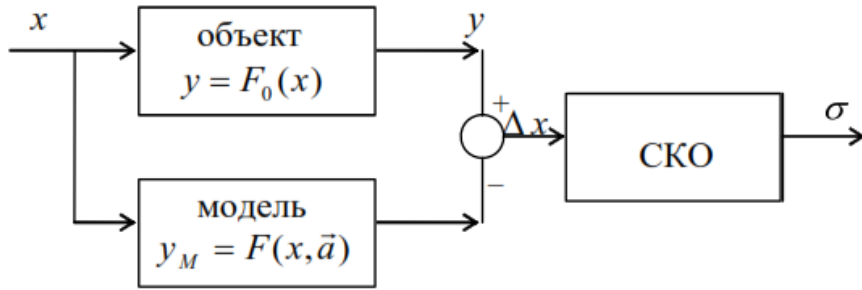


Рис. 3.2

В качестве меры близости выходов используют норму равномерного приближения

$$\Delta = \max|y - y_M| = \max|y - F(x, \bar{a})|; \quad x \in [x_{\min}, x_{\max}] \quad (3.39)$$

или дискретную норму среднеквадратичного приближения

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n [y_i - F(x_i, \bar{a})]^2}. \quad (3.40)$$

Модель считается адекватной, если выполняется условие $\Delta \leq \varepsilon_1$ или $\sigma \leq \varepsilon_2$, где $\varepsilon_1, \varepsilon_2$ – заданные положительные числа.

Уменьшение ошибки модели можно достичь путем увеличения степени полинома (3.36). Однако с некоторого момента при повышении степени полинома (3.36) среднеквадратическая ошибка вместо того, чтобы уменьшаться, может увеличиваться из-за искажения результатов измерения выхода неконтролируемыми помехами.

Поэтому в рассмотренной выше процедуре идентификации нелинейной модели неясным остается вопрос выбора числа m – степени модели (3.36).

Обычно его определяют на основе анализа наблюдений выходов объекта и модели.

Если для оценки адекватности выбирается равномерная норма (3.39), то полином, который доставляет

$$\min \Delta = \min_m \max_{X_{\min} \leq X \leq X_{\max}} \left| y - \sum_{j=1}^m a_j x^j \right|, \quad (3.41)$$

является оптимальным.

В случае, когда выбирается среднеквадратичная норма, то полином, который доставляет

$$\min_m \sigma = \min_m \left(\frac{1}{n} \sum_{i=1}^n \left[y_i - \sum_{j=0}^m a_j x_i^j \right] \right)^{1/2} \quad (3.42)$$

является полиномом наилучшего среднеквадратичного дискретного приближения.

Следует отметить, что, если результаты измерений выхода объекта в процессе идентификации содержат помехи, то выбор оптимальной величины m из условия (3.41) и (3.42) приведет к неоправданному увеличению степени полинома (3.36) и усложнению модели. В этом случае желательно выбрать m с учетом величины ошибки измерения выхода объекта.

Среднеквадратичную ошибку (СКО) измерения y можно определить по результатам повторных измерений y_i^0 ($i = \overline{1, k}$) выхода объекта при фиксированном значении x^0 входа

$$\sigma_y = \sqrt{\frac{1}{k-1} \sum_{i=1}^k (y_i^0 - \bar{y}^0)^2}, \quad (3.43)$$

где $\bar{y}^0 = \frac{1}{k} \sum_{i=1}^k y_i^0$.

Среднеквадратическая ошибка модели

$$\sigma_M = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^m a_j x_i^j \right)^2}, \quad (3.44)$$

при фиксированных значениях x_i, y_i ($i = \overline{1, n}$) и a_j ($j = \overline{1, m}$) зависит от степени полинома (3.36). Естественно потребовать, чтобы СКО модели была соизмерима с СКО измерения выхода объекта, т.е.

$$\sigma_M \leq \sigma_y. \quad (3.45)$$

При выборе m условие (3.45) проверяется в процессе последовательного увеличения степени полинома (3.36), начиная с $m = 1$. При этом для каждого m требуется заново вычислять все коэффициенты a_j ($j = \overline{0, m}$). Если же в качестве базисных функций $f_j(x)$ выбрана ортогональная система полиномов Чебышева, то увеличение числа членов ряда (3.36) на единицу не требует пересчета найденных ранее коэффициентов. Последовательное увеличение m продолжается до тех пор, пока не будет выполнено условие (3.45).

Рассмотрим другой подход к определению наилучшей степени полинома (3.36). Пусть выход объекта $y = F_0(x)$ измерена достаточно точно и является гладкой и низкочастотной функцией. Если $F_0(x)$ есть полином степени m , то

$$\frac{d^m F_0(x)}{dx^m} = const, \quad \frac{d^{m+1} F_0(x)}{dx^{m+1}} = 0.$$

Используем это свойство полинома для оценки числа m . Производные $F_0(x)$ неизвестны, поэтому заменим их конечными разностями выхода объекта y в точках $y_1 = y_j$ ($j = \overline{1, n}$):

$$\begin{aligned} \Delta^1 y_j &= (y_{j+1} - y_j) / \Delta x, & j &= \overline{1, n-1}; \\ \Delta^2 y_j &= (\Delta^1 y_{j+1} - \Delta^1 y_j) / \Delta x, & j &= \overline{1, n-2}; \\ &\dots\dots\dots & &\dots\dots\dots \\ \Delta^k y_j &= (\Delta^{k-1} y_{j+1} - \Delta^{k-1} y_j) / \Delta x, & j &= \overline{1, n-k}; \quad k = \overline{1, n-1}, \end{aligned}$$

где $\Delta^1 y_j$ – конечная разность i -го порядка в $y_1 = y_j$;

Δx – шаг измерения x в эксперимента.

Обычно значения $\Delta^k y_j$ флуктуируют, поэтому следует определить среднее значение их модулей

$$\overline{\Delta^k y} = \frac{1}{n-k} \sum_{j=1}^{n-k} |\Delta^k y_j|, \quad k = \overline{1, n-1}.$$

Если при некотором k все конечные разности порядка k или среднее значение их модулей близко к нулю, то следует принять $m = k - 1$.

3.4. Идентификация объекта моделирования при наличии случайных помех

В приведенных в предыдущих разделах методах идентификации предполагалось, что выход объекта полностью определяется наблюдаемым входным воздействием. В действительности же в подавляющем большинстве случаев выходная величина объекта определяется также ненаблюдаемыми и неуправляемыми входными воздействиями (помехами).

Структурная схема объекта изображена на рис. 3.3.

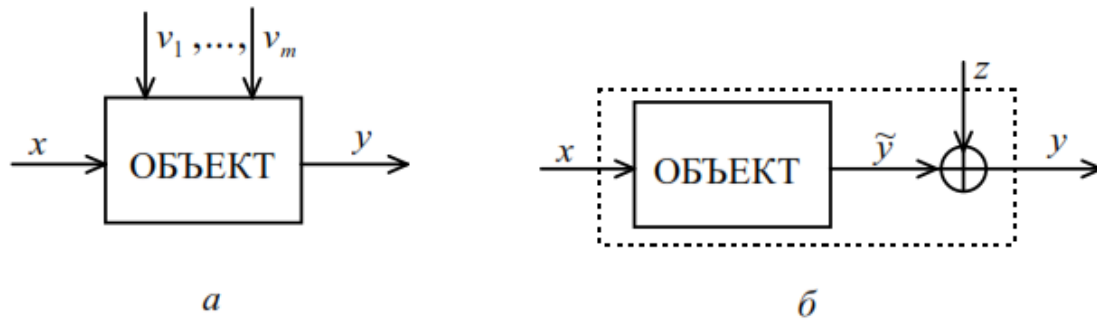


Рис. 3.3

На этой схеме x, y – наблюдаемые входной и выходной сигналы (в общем случае векторы), V_1, \dots, V_m – случайные факторы (помехи), порожденные либо самим объектом, либо средствами сбора и передачи информации о его работе. Как правило, V_1, \dots, V_m – ненаблюдаемые случайные процессы.

Для того, чтобы наиболее простым способом учесть действие всех помех при идентификации объекта, структурную схему объекта целесообразно представить так, как показано на рис. 3.3, б, то есть заменить все помехи одним эквивалентным возмущением $z(t)$, приведенным непосредственно к выходу объекта. Рассмотренная структура имеет только два входных воздействия $x(t)$ и $z(t)$. Предположим, что $x(t)$ является детерминированным сигналом (без помех), наблюдаемым точно. В практических задачах сигналы $z(t)$ и $\tilde{y}(t)$, как правило, являются ненаблюдаемыми, а выходной сигнал $y(t)$ – наблюдаемым с помехой. Обычно $y(t)$ является аддитивной смесью детерминированной $\tilde{y}(t)$ и случайной $z(t)$ составляющих

$$y(t) = \tilde{y}(t) + z(t).$$

Описание проблемы может быть осложнено тем, что шум может быть не только аддитивным, но и мультипликативным (например, дрейф параметров объекта).

В случае, когда помехи малы или совсем отсутствуют, ($z = 0$) выходной сигнал y полностью определяется только наблюдаемым входным сигналом x . Такой объект является детерминированным, в отличие от стохастического ($z \neq 0$), в котором помехи оказывают существенные влияния на выход.

Помехи при идентификации приводят к необходимости сбора большего числа данных, чем это необходимо в детерминированном объекте. Поведение выхода такого объекта может быть представлено в виде функции

$$y = F_0(x, z). \quad (3.46)$$

Конкретизируем вид модели объекта (3.46). Будем рассматривать такие объекты, у которых регулярная $\tilde{y}(t)$ и случайная z составляющие выхода y могут быть разделены, то есть модель (3.46) представлена в виде (случай аддитивной помехи)

$$y = F_0(x) + z. \quad (3.47)$$

Кроме того, будем предполагать, что статистические свойства стохастического сигнала z не зависят от входа x , то есть целиком определяются какой-то определенной плотностью распределения вероятности $P(z)$, которая предполагается неизменной, то есть $z(t)$ является стационарным случайным процессом.

В одномерном случае плотность нормального распределения имеет вид

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z - m_z)^2}{2\sigma^2}\right]. \quad (3.48)$$

Она характеризуется двумя параметрами: средним значением (математическим ожиданием)

$$m_z = M(z)$$

и дисперсией

$$\sigma^2 = M[(z - m_z)^2],$$

где M – знак (оператора) математического ожидания

$$M(\cdot) = \int_{-\infty}^{\infty} (\cdot) P(z) dz.$$

Оценка величины дисперсии эквивалентного возмущения z . Степень стохастичности объекта определяется уровнем действующих помех. Поэтому оценка дисперсии возмущений является весьма полезной при идентификации объекта.

Если случайный процесс $z(t)$ стационарен, его математическое ожидание и дисперсия не меняются во времени, то есть $m_z(t) = m_z = const$; $\sigma_z^2(t) = \sigma_z^2 = const$.

Оценки математического ожидания и дисперсии случайного сигнала $z(t)$ можно определить по его реализации в дискретные моменты времени

$$z(t_i), \quad t_i = i\Delta t, \quad i = \overline{1, K},$$

где K – объем выборки,

Δt – интервал дискретности измерения величины z , по формулам:

$$\tilde{m}_z = \frac{1}{K} \sum_{i=1}^K z(t_i); \quad (3.49)$$

$$\tilde{\sigma}_z^2 = \frac{1}{K-1} \sum_{i=1}^K [z(t_i) - \tilde{m}_z]^2. \quad (3.50)$$

Однако использование этих формул невозможно из-за недоступности измерений действующих на объект помех. Поэтому на рис. 3.3.б все ненаблюдаемые помехи, воздействующие на различные части объекта, приведены к выходу объекта и представлены в виде эквивалентного аддитивного шума $z(t)$.

Важным свойством эквивалентного возмущения $z(t)$ является то, что оно доступно измерению, и, следовательно, его статистические характеристики могут быть оценены экспериментально. Действительно, для того чтобы получить реализацию этого возмущения, достаточно измерять выход объекта $y(t)$ в дискретные моменты времени $t_i = i\Delta t$ $i = \overline{1, k}$ при постоянном уровне входного воздействия $x(t) = const$. По дискретной реализации $y(t_i)$ выхода объект определяем

$$\bar{y} = \frac{1}{K} \sum_{i=1}^K y(t_i); \quad \sigma_y^2 = \frac{1}{K-1} \sum_{i=1}^K [y(t_i) - \bar{y}]^2. \quad (3.51)$$

Учитывая, что $z = y - \tilde{y}$, определяем $\tilde{\sigma}_z^2$ по закону сложения дисперсий

$\tilde{\sigma}_z^2 = \sigma_y^2 + \sigma_{\tilde{y}}^2$, $\sigma_{\tilde{y}}^2 = 0$, так как $\tilde{y}(t) = const$, поскольку $x(t) = const$. Поэтому $\tilde{\sigma}_z^2 = \sigma_y^2$, то есть

$$\tilde{\sigma}_z^2 = \frac{1}{K-1} \sum_{i=1}^K [y(t_i) - \bar{y}]^2 \quad (3.52)$$

И среднеквадратичное отклонение z равно

$$\tilde{\sigma}_z = \sqrt{\frac{1}{K-1} \sum_{i=1}^K [y(t_i) - \bar{y}]^2}. \quad (3.53)$$

Величина σ_z оценивает влияние на выход объекта совокупности помех.

Проверка гипотезы нормальности распределения вероятности. Для методов идентификации стохастических объектов очень важно, чтобы результаты наблюдений случайной величины z подчинялись нормальному закону распределения. Поэтому проверка нормальности распределения составляет основное

содержание предварительной обработки результатов наблюдений при идентификации стохастических объектов.

Для основательной проверки нормальности распределения чаще всего используется один из двух наиболее строгих и надежных критериев согласия: χ^2 – критерий Пирсона и $K - C$ – критерий Колмогорова-Смирнова. Но эти критерии связаны с достаточно трудоемкими вычислениями. Существуют простые рекомендации для быстрой «прикладной» оценки нормальности распределения. Приведем некоторые простые рекомендации.

Пусть имеется ограниченный ряд наблюдений z_1, z_2, \dots, z_K непрерывно распределенной случайной величины z . Среднее значение наблюдений (эмпирическое или выборочное среднее) можно определить по формуле

$$\bar{z} = \frac{1}{K} \sum_{i=1}^K z_i. \quad (3.54)$$

Выборочная дисперсия $\tilde{\sigma}_z^2$ и среднеквадратическое $\tilde{\sigma}_z$ эмпирического распределения соответственно могут быть найдены по формулам:

$$\tilde{\sigma}_z^2 = \frac{1}{K-1} \sum_{i=1}^K (z_i - \bar{z})^2; \quad \tilde{\sigma}_z = \sqrt{\tilde{\sigma}_z^2}. \quad (3.55)$$

Для проверки нормальности распределения при не очень больших выборках ($K < 120$) необходимо вычислить среднее абсолютное отклонение (CAO) по формуле

$$CAO = \frac{1}{K} \sum_{i=1}^K |z_i - \bar{z}|. \quad (3.56)$$

Для выборки, имеющей приближенно нормальный закон распределения, должно быть справедливо выражение

$$\left| \frac{CAO}{\tilde{\sigma}_z} - 0.7979 \right| < \frac{0.4}{\sqrt{K}}. \quad (3.57)$$

Следовательно, гипотеза нормальности распределения выборки принимается, если выполняется условие (3.57).

Как было отмечено выше, при идентификации стохастического объекта исследователь не располагает выборкой z_i , ($i = \overline{1, K}$), так как случайная величина z , как правило, не наблюдаема. Однако необходимые для проверки усло-

вия (3.57) статистические характеристики случайной величины z могут быть определены по наблюдениям $y_i = y(t_i)$, ($i = \overline{1, K}$) выходной величины объекта $y(t)$, которая измеряется в смеси с помехой z (рис. 3.3.б) в дискретные моменты времени t_i при $x(t) = const$.

Оценки $\tilde{\sigma}_z^2$ и $\tilde{\sigma}_z$ вычисляются соответственно по формулам (3.52) и (3.53). Приведем формулы для вычисления \bar{z} и САО по выборке y_i , ($i = \overline{1, K}$).

Из рис. 3.3,б следует, что $y = \tilde{y} + z$, где $\tilde{y} = const$ так как $x(t) = const$. Учитывая, что $z = y - \tilde{y}$, имеем

$$\bar{z} = \frac{1}{K} \sum_{i=1}^K z_i = \frac{1}{K} \sum_{i=1}^K (y_i - \tilde{y}) = \frac{1}{K} \sum_{i=1}^K y_i - \frac{1}{K} \sum_{i=1}^K \tilde{y} = \bar{y} - \tilde{y}; \quad (3.58)$$

$$CAO = \frac{1}{K} \sum_{i=1}^K |z_i - \bar{z}| = \frac{1}{K} \sum_{i=1}^K |y_i - \tilde{y} - (\bar{y} - \tilde{y})| = \frac{1}{K} \sum_{i=1}^K |y_i - \bar{y}|, \quad (3.59)$$

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

Быструю проверку гипотезы нормальности распределения для сравнительно широкого класса выборок $3 < K < 1000$ можно выполнить, используя размах варьирования r :

$$r = z_{\max} - z_{\min} = y_{\max} - \tilde{y} - (y_{\min} - \tilde{y}) = y_{\max} - y_{\min},$$

где y_{\max}, y_{\min} – наибольшее и наименьшее значение y в выборке y_i , ($i = \overline{1, K}$).

Подсчитывают отношение $r/\tilde{\sigma}_z$ и сопоставляют с критическими верхними и нижними границами этого отношения, приведенными в табл. 3.1. Если $r/\tilde{\sigma}_z$ меньше нижней или больше верхней границы, то нормального распределения нет. В противном случае гипотеза нормальности распределения подтверждается. Особенно важно, чтобы это условие соблюдалось при $\alpha = 0.10$ (10 %-ный уровень значимости) или доверительной вероятности $p = 1 - \alpha = 0.9$.

Сглаживание выхода объекта. Выход стохастического объекта при идентификации измеряется с помехой z . Наличие помех присуще реальным объектам, а на имитационной модели на ЭВМ они могут быть специально генерированы.

Чтобы получить действительные характеристики объекта, необходимо исключить из выходного сигнала $y(t)$ составляющие, определяемые помехой

$z(t)$, оставив только реакцию на входное воздействие $x(t)$. Для этого следует отфильтровать помеху $z(t)$ из $y(t)$.

Методы фильтрации зашумленных сигналов подробно излагаются в дисциплине «Математические основы теории систем». Здесь рассмотрим два наиболее простых и распространенных метода сглаживания применительно к идентифицируемому статистическим характеристика.

Таблица 3.1

Критические границы отношения $r/\tilde{\sigma}_z$

Объем выборки	Нижние границы			Верхние границы		
	Вероятность границы (уровень значимости α)					
1	2	3	4	5	6	7
	0.01	0.05	0.1	0.1	0.05	0.01
5	2.02	2.15	2.22	2.712	2.753	2.803
10	2.51	2.67	2.76	3.57	3.685	3.875
1	2	3	4	5	6	7
15	2.80	2.97	3.07	4.02	4.17	4.44
20	2.99	3.18	3.29	4.32	4.49	4.80
25	3.15	3.34	3.45	4.53	4.71	5.06
30	3.27	3.47	3.59	4.7	4.89	5.26
35	3.38	3.58	3.70	4.84	5.04	5.42
40	3.47	3.67	3.79	4.96	5.16	5.56
45	3.55	3.75	3.88	5.06	5.26	5.67
50	3.62	3.83	3.95	5.14	5.35	5.77
60	3.75	3.96	4.08	5.29	5.51	5.94
70	3.85	4.06	4.19	5.41	5.63	6.07
80	3.94	4.16	4.28	5.51	5.73	6.18
90	4.02	4.24	4.36	5.60	5.82	6.27
100	4.10	4.31	4.44	5.68	5.90	6.36
150	4.38	4.59	4.72	5.96	6.18	6.64
200	4.59	4.78	4.90	6.15	6.39	6.84
500	5.13	5.37	5.49	6.72	6.94	7.42
1000	5.57	5.79	5.92	7.11	7.33	7.80

Сглаживание скользящим средним. Среди фильтров с заданной структурой наиболее широкое распространение нашли фильтры, основанные на операции скользящего среднего. Выход такого фильтра связан с его входом соотношением

$$y(t) = \frac{1}{T} \int_{t-T}^t y(\tau) d\tau, \quad (3.60)$$

где T – память фильтра (настраиваемый параметр).

Интеграл (3.60) представим в виде

$$\bar{y}(t) = \frac{1}{T} \int_0^t y(\tau) d\tau - \frac{1}{T} \int_0^{t-T} y(\tau) d\tau.$$

Отсюда следует, что фильтр скользящего среднего представляет собой параллельное соединение двух интегрирующих звеньев, одно из которых последовательно соединено со звеном запаздывания.

3.5. Построение регрессионной модели стохастического объекта

Пусть задан некоторый стохастический объект (рис. 3.5), входная и выходная координаты X и Y которого являются случайными величинами.

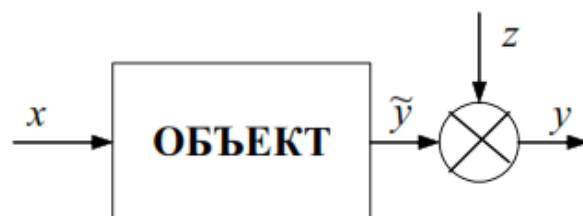


Рис. 3.5

На Y влияет не только входная координата X , но и случайная помеха Z (нестабильность режима работы объекта, стохастические воздействия внешней среды, погрешности измерения Y и т.д.). Поэтому нельзя говорить о функциональной зависимости Y от X . В подобных случаях следует говорить о наличии стохастической (вероятностной) связи между переменными X и Y объекта в статике.

Случайные величины X и Y являются зависимыми, если закон распределения вероятностей одной из них зависит от значений другой, т. е.

$$F\left(\frac{Y}{X=x}\right) = P \{Y < y \text{ при условии } X = x\}.$$

Функция $F\left(\frac{Y}{X=x}\right)$, или просто $F\left(\frac{Y}{X}\right)$, называется условным интегральным законом распределения вероятностей Y при $X = x$. $\frac{dF\left(\frac{Y}{X}\right)}{dx} = f\left(\frac{Y}{X}\right)$ называется условной плотностью распределения вероятностей.

Предположим, что можно установить $X = x_1$, тогда поведение случайной величины Y будет полностью характеризоваться условной плотностью распре-

деления вероятностей $f\left(\frac{Y}{X_1}\right)$. Обозначим условные числовые характеристики Y – математическое ожидание и дисперсию через $M\left[\frac{Y}{X_1}\right]$ и σ_{y/x_1}^2 . Затем зафиксируем $X = x_2$ и найдем $f\left(\frac{Y}{X_2}\right)$, $M\left[\frac{Y}{X_2}\right]$, σ_{y/x_2}^2 , положим $X = x_3$, найдем $f\left(\frac{Y}{X_3}\right)$, $M\left[\frac{Y}{X_3}\right]$, σ_{y/x_3}^2 , и т.д.

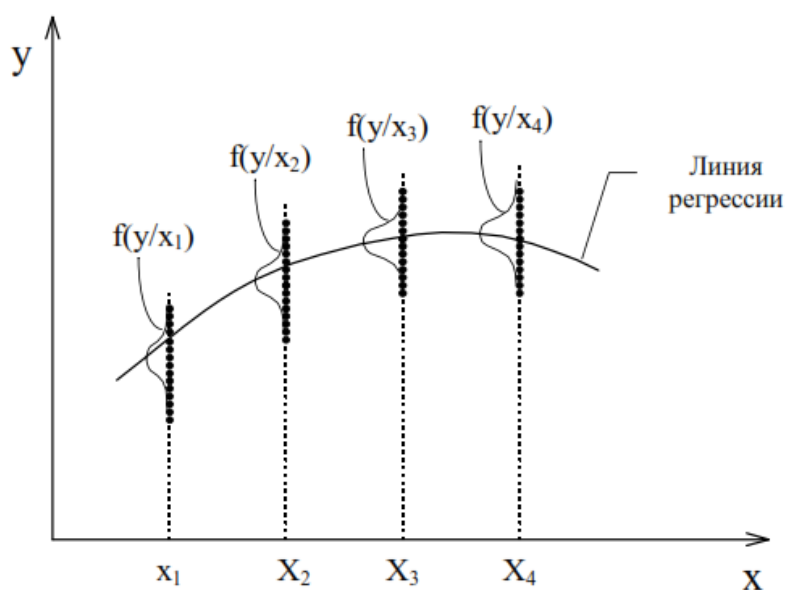


Рис. 3.6

Будем считать, что структура $f\left(\frac{Y}{X}\right)$ не зависит от x , а параметры функции плотности $M\left[\frac{Y}{X}\right]$ и $\sigma_{y/x}^2$ зависят от того, какое значение x принимает случайная величина X .

Зависимость $M\left[\frac{Y}{X}\right]$ от x называется регрессионной. Ниже будем говорить только о регрессионной зависимости

$$M\left[\frac{Y}{X}\right] = f(x). \quad (3.61)$$

Регрессионная зависимость показывает, как изменятся среднее значение Y при изменении X .

Если соединить плавной линией точки $M\left[\frac{Y}{X_1}\right]$, $M\left[\frac{Y}{X_2}\right]$, $M\left[\frac{Y}{X_3}\right]$ и т.д., то получи линию регрессии. Эта линия есть статическая характеристика объекта.

Уравнением регрессии называют функцию $f(x)$, описывающую линию регрессии. Уравнения регрессии классифицируют на линейные (корреляционные) и нелинейные.

При построении регрессионной модели статики (3.61) объекта широко применяется пассивный метод идентификации. Этот метод применяют при изучении статики объектов, уровень помех которых достаточно велик, а также в тех случаях, когда недопустимо нанесение искусственных возмущений на входе объекта.

Пассивный метод идентификации основан в получении статистической информации об объекте по данным его нормальной эксплуатации. Затем реализации входной x и выходной y величин обрабатываются таким образом, чтобы определить регрессионную модель

$$M\left[\frac{Y}{X}\right] = f(x, \vec{a}), \quad (3.62)$$

где \vec{a} – вектор коэффициентов модели.

Определение уравнения регрессии состоит из двух этапов: выбор типа уравнения регрессии и расчет его коэффициентов.

Первый этап обычно осуществляется либо путем эмпирического подбора типа уравнения регрессии по виду корреляционного поля между входными и выходными величинами, либо путем теоретического изучения закономерности физического процесса, отражением которого является стохастическая связь между этими величинами. Иногда оба подхода используются в сочетании друг с другом.

Второй этап – расчет коэффициентов уравнения регрессии – чаще всего выполняется методом наименьших квадратов. При этом минимизация по a выражения

$$E = \sum_{i=1}^n \left(y_i - M\left[\frac{Y}{X_i}\right] \right)^2 \quad (3.63)$$

приводит к оценкам вектора \vec{a} . Очевидно, на данной реализации для принятого уравнения модели объекта (3.62) полученные оценки \vec{a} дают наименьшее СКО реализации, полученной по модели от наблюдаемой реализации y_i , ($i = \overline{1, n}$) объекта.

Следует отметить, что пассивно-статистический метод имеет целый ряд существенных недостатков по сравнению с активными методами:

1. полученная модель объекта справедлива только в пределах использованного экспериментального статистического материала;
2. трудно разделить эффекты от коррелированности части входных величин многомерного объекта;

3. индивидуальные коэффициенты регрессии не имеют какого-либо физического смысла;

4. не извлекается информация об ошибке опытов;

5. требуется получить большой объем экспериментальных данных и производить трудоемкие вычисления.

Указанные недостатки в значительной степени снижают ценность моделей, полученных пассивно-статистическим методом. К этому методу прибегают только в тех случаях, когда другие методы не могут быть использованы.

Предварительный анализ экспериментального статистического материала составляет основную задачу корреляционного анализа при идентификации стохастического объекта. При этом суть корреляционного анализа сводится к оценке силы стохастической связи между случайными величинами X и Y и установлению вида зависимости между ними в виде уравнения регрессии.

Чтобы предварительно определить наличие корреляционной связи между X и Y , наносят экспериментальные точки (x_i, y_i) на график и строят корреляционное поле (рис. 3.7).

По тесноте группирования точек вокруг прямой или кривой линии, по наклону линии можно визуальным образом судить о наличии корреляционной связи.

Корреляционное поле характеризует вид связи между X и Y , т.е. наличие линейной или нелинейной зависимости.

Существует три вида корреляции: линейная, нелинейная и множественная корреляция. При линейной корреляции линия регрессии аппроксимируется уравнением прямой, при нелинейной – уравнением кривой. Множественная корреляция определяет связь между многими величинами, и при этом используется уравнение множественной регрессии.

Наиболее распространенной является линейная корреляция.

Понятие корреляции дает возможность судить о том, насколько тесно ложатся экспериментальные точки на аппроксимирующую кривую линии регрессии. Если регрессия определяет предполагаемое соотношение между переменными, то корреляция показывает, насколько хорошо это соотношение отражает действительность.

Задача стохастического анализа объекта ставится таким образом: по данной выборке объема n оценить силу (тесноту) корреляционной связи между X и Y , найти уравнение приближенной регрессии и оценить допускаемую ошибку.

Оценка тесноты корреляционной связи. В общем виде задача выявления и оценки силы стохастической связи не решена. Существуют показатели, оценивающие те или иные стороны стохастической связи.

Поведение случайной величины Y описывается обычно следующими функциональными и числовыми характеристиками:

– интегральным законом распределения вероятностей

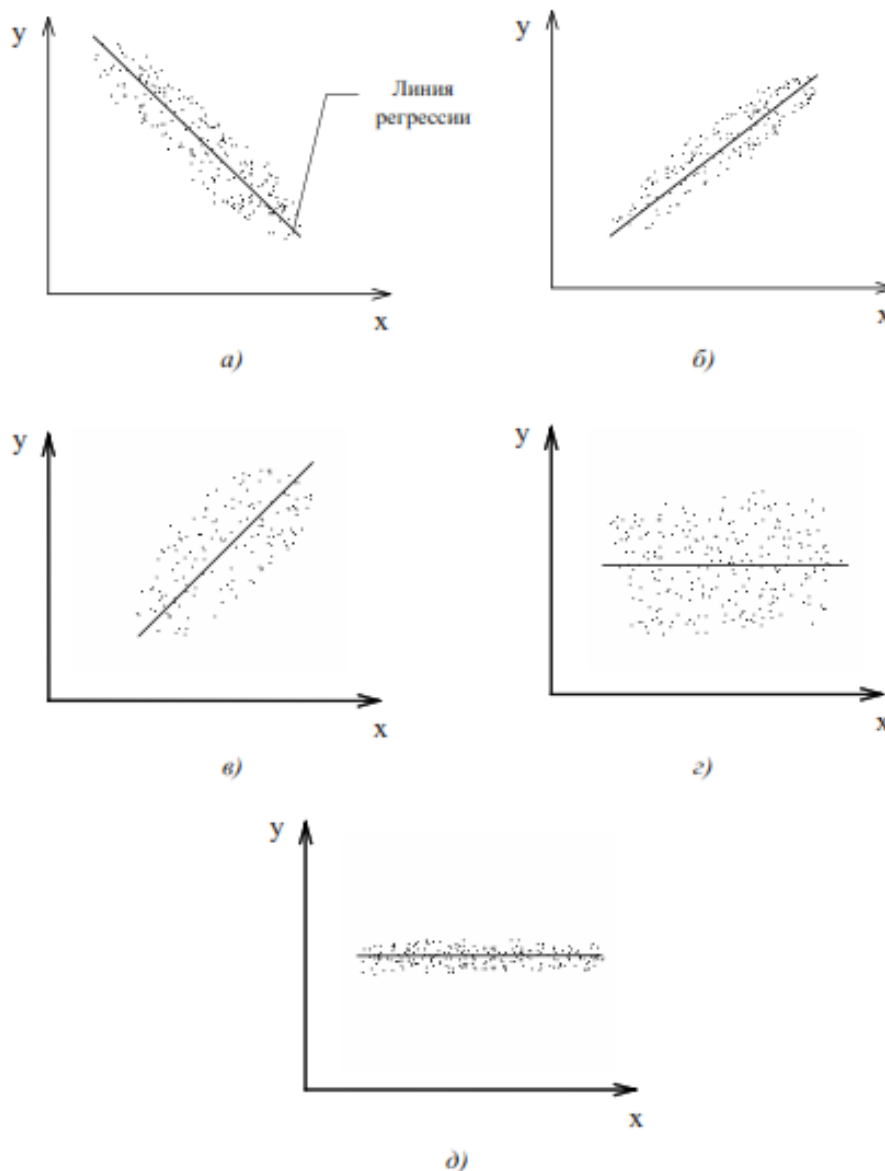
$$F(Y) = P\{Y < y\};$$

– дифференциальным законом распределения вероятностей

$$f(y) = \frac{dF(Y)}{dy} \text{ (или плотность вероятностей);}$$

– математическим ожиданием $M[Y] = \int_{-\infty}^{+\infty} f(y)ydy$;

– дисперсией $\sigma_y^2 = M\{(Y - M[Y])^2\} = \int_{-\infty}^{+\infty} (Y - M[Y])^2 f(y)dy$.



Представлены: а – сильная отрицательная корреляция; б – сильная положительная корреляция; в – слабая положительная корреляция; г, д – отсутствие корреляции.

Рис. 3.7

Отметим, что $F(-\infty) = 0$; $F(\infty) = 1$; $F(Y)$ – возрастающая функция. Если $F(Y)$ непрерывна по Y , то $P\{Y = y\} = 0$, однако, $P\{y < Y \leq y + dy\} = f(y)dy$.

Кроме того, $\int_{-\infty}^{+\infty} f(y)dy = \int_{-\infty}^{+\infty} dF = 1$.

Поведение системы двух случайных величин X и Y , значения (x, y) которых изменяются от $-\infty$ до $+\infty$, полностью описываются интегральным законом распределения вероятностей

$$F(x, y) = P\{X < x, Y < y\}$$

или плотностью распределения вероятностей

$$F(x, y) = \frac{\partial F(x, y)}{\partial x \partial y},$$

т.е. $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy$. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$.

Отметим, что $F(-\infty, -\infty) = 0$, $F(\infty, \infty) = 1$. Функция $F(x, y)$ есть не убывающая функция (x, y) , так как с ростом x и y вероятность выполнения двух неравенств $X < x, Y < y$ не может уменьшаться. Далее, $F(x, \infty) = F(x)$, $F(\infty, y) = F(y)$, где $F(x)$ и $F(y)$ – функции распределения X и Y .

По известным $f(x, y)$ всегда можно найти $f(x)$ и $f(y)$.

$$f(x) = \frac{dF(x)}{dx} = \left. \frac{\partial F(x, y)}{\partial x} \right|_{y=\infty} = \frac{F(x, \infty)}{dx} = \int_{-\infty}^{\infty} f(x, y) dy;$$

$$f(y) = \frac{dF(y)}{dy} = \left. \frac{\partial F(x, y)}{\partial y} \right|_{x=\infty} = \frac{F(\infty, y)}{dy} = \int_{-\infty}^{\infty} f(x, y) dx.$$

Обратная задача – определение $f(x, y)$ по известным $f(x)$ и $f(y)$ в общем случае не имеет решения, т.к. эти две функции не характеризуют взаимную зависимость величин X и Y . Но для независимых случайных величин $f(x, y) = f(x)f(y)$.

Основной числовой характеристикой двух случайных величин X и Y является ковариационный (корреляционный) момент

$$K_{xy} = M[(X - M[X])(Y - M[Y])] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)(X - M[X])(Y - M[Y]) dx dy$$

Назовем случайную величину $\dot{X} = X - M[X]$ центрированной, ее значение \dot{x} есть отклонение x от среднего значения $M[X]$. Тогда момент K_{xy} равен среднему значению произведения двух центрированных случайных величин \dot{X}, \dot{Y} и характеризует тесноту взаимной связи X и Y .

Действительно, пусть X и Y независимые, тогда $f(x, y) = f(x)f(y)$ и

$$K_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\dot{x}\dot{y} dx dy = \int_{-\infty}^{\infty} f(x)\dot{x} dx \int_{-\infty}^{\infty} f(y)\dot{y} dy;$$

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\dot{x} dx &= \int_{-\infty}^{\infty} f(x)(X - M[X]) dx = \int_{-\infty}^{\infty} f(x)x dx - \int_{-\infty}^{\infty} f(x)M[X] dx = \\ &= M[X] - M[X] \int_{-\infty}^{\infty} f(x) dx = M[X] - M[X] = 0; \end{aligned}$$

$$K_{xy} = 0.$$

Обратное утверждение – если $K_{xy} = 0$, то X и Y независимы – в общем случае доказать нельзя. Более того, можно найти пример, в котором $K_{xy} = 0$, а X и Y – зависимые случайные величины.

Следовательно, величина K_{xy} не является исчерпывающей характеристикой силы связи X и Y . Оказывается, что K_{xy} характеризует только линейную зависимость случайных величин X и Y . Если $K_{xy} = 0$, то можно говорить о том, что X и Y линейно независимы (не коррелированы).

Ковариационный момент K_{xy} характеризует разброс случайных величин X и Y относительно их средних. Если X мало, то и K_{xy} также будет малым, хотя X и Y в достаточной мере коррелированы между собой.

Размерность K_{xy} не имеет физического смысла, что делает невозможным сравнение ковариационных моментов для двух пар случайных величин X и Y . Поэтому чаще всего используется нормированная величина K_{xy} , т.е. коэффициент корреляции

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}, \quad (3.64)$$

где σ_x^2, σ_y^2 – дисперсии X и Y ; $\sigma_x = +\sqrt{\sigma_x^2}$ и $\sigma_y = +\sqrt{\sigma_y^2}$.

Величина r_{xy} безразмерная, кроме того она слабее, чем K_{xy} , зависит от рассеяния X и Y , так как здесь K_{xy} нормируется делением на СКО X и Y от $M[X]$ и $M[Y]$.

Величину r_{xy} можно вычислить по выборке случайных величин X и Y . Допустим, что проведено n испытаний и при каждом отмечались значения случайных величин X, Y . Через \bar{X} и \bar{Y} обозначим их среднее значение:

$$\bar{X} = \tilde{M}[X] = \frac{1}{n} \sum_{i=1}^n x_i; \quad \bar{Y} = \tilde{M}[Y] = \frac{1}{n} \sum_{i=1}^n y_i, \quad (3.65)$$

где $\tilde{M}[X]$ и $\tilde{M}[Y]$ – оценки математического ожидания X и Y .

Тогда выборочный коэффициент корреляции определяется по формуле

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)\sigma_x \sigma_y}, \quad (3.66)$$

$$\text{где } \sigma_x^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2; \quad \sigma_x = +\sqrt{\sigma_x^2}; \quad (3.67)$$

$$\sigma_y^2 = \frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2; \quad \sigma_y = +\sqrt{\sigma_y^2}; \quad (3.68)$$

σ_x^2, σ_y^2 – выборочные дисперсии.

Уменьшение на единицу знаменателя в приведенных выше формулах связано с тем, что величины \bar{x} и \bar{y} , относительно которых берутся отклонения, сами зависят от объема выборки.

Для удобства вычислений и повышения точности расчетов преобразуем формулы (3.66), (3.67) и (3.68). Несложные преобразования их приводят к следующим формулам:

$$\sigma_x^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]; \quad \sigma_y^2 = \frac{1}{n-1} \left[\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \right]; \quad (3.69)$$

$$r_{xy} = \frac{1}{(n-1)\sigma_x\sigma_y} \left(\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right). \quad (3.70)$$

Преимущество формул (3.69), (3.70) в том, что в них нет операций вычитания близких чисел, как в формулах (3.66), (3.68), приводящих к потере точности вычислений. В формулах (3.69), (3.70) эта операция применяется только один раз.

Коэффициент корреляции не меняется от прибавления к X и Y каких-либо неслучайных слагаемых, от умножения X и Y на положительные числа. Если же одну из величин, не меняя другой, умножить на (-1) , то на (-1) умножится и коэффициент корреляции.

Для независимых случайных величин $r_{xy} = 0$, но $r_{xy} = 0$ может быть и для некоторых зависимых величин, которые при этом называются некоррелированными.

Коэффициент корреляции характеризует не всякую зависимость, а только линейную. Линейная вероятностная зависимость случайных величин X и Y заключается в том, что при возрастании X случайная величина Y имеет тенденцию возрастать (при $r_{xy} > 0$) или убывать (при $r_{xy} < 0$) по линейному закону. Величина r_{xy} характеризует степень тесноты линейной зависимости ($-1 \leq r_{xy} \leq 1$).

При $r_{xy} = \pm 1$ величины X и Y связаны между собой линейной функциональной зависимостью, а при $(-1 \leq r_{xy} \leq 1)$ связь между этими величинами стохастическая. При $r_{xy} = 0$ между X и Y линейной корреляционной связи не существует, однако может существовать нелинейная регрессия.

Линейная регрессионная модель. Пусть с помощью коэффициента корреляции установлено, что зависимость между X и Y существует и она линейная, т. е.

$$M\left[\frac{Y}{X}\right] = a_0 + a_1 x. \quad (3.71)$$

Надо только определить коэффициенты a_0 и a_1 . Определим их методом наименьших квадратов. Построим для этого функцию

$$E = M\left[(y - a_0 - a_1 x)^2\right] = E(a_0, a_1),$$

где M – символ математического ожидания (усреднение на интервале $[-\infty, +\infty]$).

Для краткости записи обозначим $M[X] = \bar{X}$, $M[Y] = \bar{Y}$. Добавим к функции невязки $(y - a_0 - a_1x)$ слагаемое $\bar{Y} + a_1\bar{X}$ и $-\bar{Y} - a_1\bar{X}$ получим

$$\begin{aligned} E(a_0, a_1) &= M \left[(y - a_0 - a_1x + \bar{Y} + a_1\bar{X} - \bar{Y} - a_1\bar{X})^2 \right] = \\ &= M \left\{ [(Y - \bar{Y}) - a_1(X - \bar{X}) + (\bar{Y} - a_0 - a_1\bar{X})]^2 \right\} = M \left[(Y - \bar{Y})^2 \right] + a_1^2 M \left[(X - \bar{X})^2 \right] + \\ &+ M \left[(\bar{Y} - a_0 - a_1\bar{X})^2 \right] - 2a_1 M \left[(Y - \bar{Y})(X - \bar{X}) \right] + 2M \left[(Y - \bar{Y})(\bar{Y} - a_0 - a_1\bar{X}) \right] - \\ &- 2a_1 M \left[(X - \bar{X})(\bar{Y} - a_0 - a_1\bar{X}) \right] = \sigma_y^2 + a_1^2 \sigma_x^2 + (\bar{Y} - a_0 - a_1\bar{X})^2 - \\ &- 2a_1 K_{xy} + 2(\bar{Y} - a_0 - a_1\bar{X}) M \left[(Y - \bar{Y}) \right] - 2a_1 (\bar{Y} - a_0 - a_1\bar{X}) M \left[(X - \bar{X}) \right]. \end{aligned}$$

Так как $M[X - \bar{X}] = M[X] - M[\bar{X}] = \bar{X} - \bar{X} = 0$; $M[Y - \bar{Y}] = 0$, получим

$$E(a_0, a_1) = \sigma_y^2 + a_1^2 \sigma_x^2 + (\bar{Y} - a_0 - a_1\bar{X})^2 - 2a_1 K_{xy}. \quad (3.72)$$

Из необходимого условия минимума функции (3.72) $\frac{\partial E}{\partial a_0} = 0$; $\frac{\partial E}{\partial a_1} = 0$

получим следующую систему уравнений для определения a_0 и a_1 :

$$\left. \begin{aligned} \bar{Y} - a_1\bar{X} - a_0 &= 0; \\ a_1\sigma_x^2 + \bar{X}\bar{Y} - a_1\bar{X}^2 - a_0\bar{X} - K_{xy} &= 0. \end{aligned} \right\} \quad (3.73)$$

Первое уравнение (3.73) умножим на \bar{X} и вычтем из него второе, получим

$$-a_1\sigma_x^2 + K_{xy} = 0.$$

Отсюда находим

$$a_1 = \frac{K_{xy}}{\sigma_x^2} = r_{xy} \frac{\sigma_y}{\sigma_x}. \quad (3.74)$$

Из первого уравнения системы (3.73) определяем a_0

$$a_0 = \bar{Y} - r_{xy} \frac{\sigma_y}{\sigma_x} \bar{X}. \quad (3.75)$$

Линейная регрессионная модель имеет вид

$$M\left[\frac{Y}{X}\right] = M[Y] - r_{xy} \frac{\sigma_y}{\sigma_x} M[X] + r_{xy} \frac{\sigma_y}{\sigma_x} X. \quad (3.76)$$

Уравнение линейной регрессии, записанное в стандартном виде

$$M\left[\frac{Y}{X}\right] - M[Y] = r_{xy} \frac{\sigma_y}{\sigma_x} (X - M[X]) \quad (3.77)$$

наилучшим образом (в смысле величины квадрата ошибок) аппроксимирует зависимость условного среднего Y от X .

Если подставить a_0 и a_1 в (3.72), то получим наименьшее значение суммы квадратов невязок

$$\min E(a_0, a_1) = \sigma_y^2 (1 - r_{xy}^2). \quad (3.78)$$

Отсюда следует, что чем ближе r_{xy} к 1, тем лучше качество линейной аппроксимации, тем ближе к линиям регрессии расположены точки $M\left[\frac{Y}{X}\right]$.

Уравнение (3.77) позволяет прогнозировать среднее значение выхода Y для заданного значения входа X .

Теперь покажем, что $|r_{xy}| \leq 1$. Если X и Y связаны линейной функциональной связью, то $E(a_0, a_1) = 0$. Следовательно, $\sigma_y^2 (1 - r_{xy}^2) = 0$. Отсюда $|r_{xy}| = 1$.

Учитывая, что всегда $E \geq 0$, то $|r_{xy}| \leq 1$.

Нелинейная регрессионная модель одномерного объекта. Зададим уравнение регрессии полиномом степени m

$$M\left[\frac{Y}{X}\right] = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m. \quad (3.79)$$

Коэффициенты a_j ($j = \overline{0, m}$) можно определить методом наименьших квадратов. При этом, как правило, уравнение (3.79) наилучшим образом (в смысле минимума суммы квадратов невязок) аппроксимирует линию регрессии. Коэффициенты a_j ($j = \overline{0, m}$) не имеют статистической трактовки.

Для оценки силы связи X и Y в общем случае используется корреляционное отношение

$$\eta_{xy} = \frac{\Delta}{\sigma_y}, \quad (3.80)$$

где $\sigma_y^2 = M[(Y - M[Y])^2]$; $\sigma_y = +\sqrt{\sigma_y^2}$;

$$\Delta^2 = M \left[\left(M \left[\frac{Y}{X} \right] - M[Y] \right)^2 \right]; \quad \Delta = +\sqrt{\Delta^2}$$

σ_y^2 – дисперсия Y относительно $M[Y]$;

Δ^2 – дисперсия линии регрессии $M \left[\frac{Y}{X} \right]$, описываемой уравнением (3.79), относительно $M[Y]$. Величина η_{xy} равна отношению СКО $M \left[\frac{Y}{X} \right]$ от $M[Y]$ к СКО Y от $M[Y]$ (рис. 3.8). Причем $0 \leq \eta_{xy} \leq 1$.

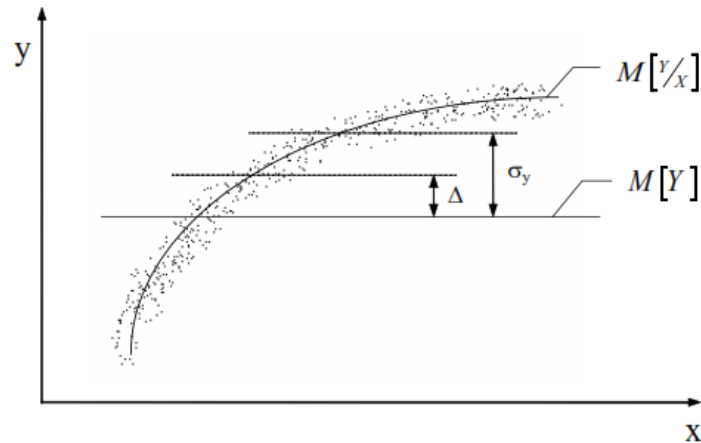


Рис. 3.8

Дисперсию σ_y^2 можно представить суммой

$$\sigma_y^2 = M \left[\left(M \left[\frac{Y}{X} \right] - M[Y] \right)^2 \right] + M \left[\left(Y - M \left[\frac{Y}{X} \right] \right)^2 \right] = \Delta^2 + M \left[\left(Y - M \left[\frac{Y}{X} \right] \right)^2 \right].$$

Последний член характеризует разброс Y относительно $M \left[\frac{Y}{X} \right]$. Отсюда находим

$$\Delta^2 = \sigma_y^2 - M \left[\left(Y - M \left[\frac{Y}{X} \right] \right)^2 \right]. \quad (3.81)$$

Из (3.80) с учетом (3.81) имеем

$$\eta_{xy}^2 = \frac{\Delta^2}{\sigma_y^2} = \frac{\sigma_y^2 - M \left[\left(Y - M \left[\frac{Y}{X} \right] \right)^2 \right]}{\sigma_y^2};$$

$$\eta_{xy}^2 = 1 - \frac{M \left[\left(Y - M \left[\frac{Y}{X} \right] \right)^2 \right]}{M \left[\left(Y - M[Y] \right)^2 \right]}. \quad (3.82)$$

Если все значения Y лежат на линии регрессии, т.е. абсолютно точно удовлетворяют уравнению (3.79), тогда X и Y связаны функциональной связью и

$$M\left[\left(Y - M\left[\frac{Y}{X}\right]\right)^2\right] = 0; \eta_{xy} = 1.$$

Теперь пусть Y не зависит от X , поэтому линия регрессии $M\left[\frac{Y}{X}\right]$ совпадает с $M[Y]$ и

$$\eta_{xy}^2 = 1 - \frac{M\left[\left(Y - M\left[\frac{Y}{X}\right]\right)^2\right]}{M\left[\left(Y - M[Y]\right)^2\right]} = 1 - 1 = 0.$$

Чем меньше величина η_{xy} , тем слабее связь между X и Y . Отметим, что при увеличении порядка полинома m , величина $M\left[\left(Y - M\left[\frac{Y}{X}\right]\right)^2\right]$ будет убывать, а $\eta_{xy} \rightarrow 1$, хотя, на самом деле, при этом сила связи между X и Y остается неизменной. Трудность объективного выбора порядка m усложняет и использование корреляционного отношения η_{xy} для решения практических задач.

Множественная линейная регрессия. Множественная регрессия применяется для описания связи входных величин x_1, x_2, \dots, x_k и выходной случайной величины Y стохастического объекта. Обозначим условное математическое ожидание случайной величины Y при условии, что

$$X_1 = x_1, X_2 = x_2, \dots, X_k = x_k \text{ через } M\left[\frac{Y}{\vec{X} = \vec{x}}\right] = M\left[\frac{Y}{\vec{x}}\right],$$

где $\vec{X} = (X_1, X_2, \dots, X_k)$; $\vec{x} = (x_1, x_2, \dots, x_k)$.

Уравнение линейной множественной регрессии имеет вид

$$M\left[\frac{Y}{X}\right] = a_0 + \sum_{j=1}^k a_j x_j, \quad (3.83)$$

где $a_j (j = \overline{1, k})$ находятся методом наименьших квадратов:

$$a_j = r_{x_j y} \frac{\sigma_y}{\sigma_{x_j}}, \quad j = \overline{1, k}; \quad (3.84)$$

$$a_0 = M[Y] - \sum_{j=1}^k a_j M[X_j]; \quad (3.85)$$

$$\sigma_{x_j}^2 = M[(X_j - M[X_j])^2]; \quad K_{x_j y} = M[(X_j - M[X_j])(Y - M[Y])];$$

$$r_{x_j}^2 = \frac{K_{x_j y}}{\sigma_{x_j} \sigma_y}, \quad j = \overline{1, k}.$$

Оценки приведенных выше числовых характеристик случайных процессов можно определить по экспериментальным данным. Пусть в результате эксперимента имеем матрицу наблюдений

$$\begin{pmatrix} x_{11} & x_{21} & x_{31} & L & x_{k1} & y_1 \\ x_{12} & x_{22} & x_{32} & L & x_{k2} & y_2 \\ L & L & L & L & L & L \\ x_{1n} & x_{2n} & x_{3n} & L & x_{kn} & y_n \end{pmatrix}, \quad (3.86)$$

где n – количество опытов;

x_{ji} – значение j -го входа в i -том опыте.

Тогда имеем следующие оценки характеристик

$$M[X_j] = \bar{X} = \frac{1}{n} \sum_{i=1}^n x_{ji}; \quad M[Y] = \bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad \sigma_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{Y})^2;$$

$$\sigma_{x_j}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ji} - \bar{X}_j)^2; \quad r_{x_j}^2 = \frac{\sum_{i=1}^n (x_{ji} - \bar{X}_j)(y_i - \bar{Y})}{(n-1)\sigma_{x_j} \sigma_y};$$

где всюду $j = \overline{1, k}$.

Коэффициенты a_j, a_0 получены при условии, что все входные воздействия линейно независимы, т.е. $r_{x_j x_l} \approx 0$, $j = \overline{1, k}$, $l = \overline{1, k}$, $j \neq l$, где

$$r_{x_j x_l} = \frac{\sum_{i=1}^n (x_{ji} - \bar{x}_j)(x_{li} - \bar{x}_l)}{(n-1)\sigma_{x_j} \sigma_{x_l}}, \quad j = \overline{1, k}, \quad l = \overline{1, k}, \quad j \neq l.$$

Рассмотрим общий случай, учитывая возможные связи между факторами (входами).

Перейдем от натурального масштаба к новому, проведя нормировку всех значений случайных величин по формулам:

$$\dot{y}_i = \frac{y_i - \bar{y}}{\sigma_y}; \dot{x}_i = \frac{x_{ji} - \bar{x}_j}{\sigma_{x_j}}; j = \overline{1, k}, i = \overline{1, n}. \quad (3.87)$$

При этом имеем исходный статистический материал в безразмерном масштабе:

$$\begin{pmatrix} \dot{x}_{11} & \dot{x}_{21} & \dot{x}_{31} & \dots & \dot{x}_{k1} & \dot{y}_1 \\ \dot{x}_{12} & \dot{x}_{22} & \dot{x}_{32} & \dots & \dot{x}_{k2} & \dot{y}_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dot{x}_{1n} & \dot{x}_{2n} & \dot{x}_{3n} & \dots & \dot{x}_{kn} & \dot{y}_n \end{pmatrix}.$$

В новом масштабе имеем:

$$\bar{\dot{x}}_j = 0; \bar{\dot{y}} = 0; \tilde{\sigma}_{\dot{x}_j} = 1 \quad \tilde{\sigma}_{\dot{y}_j} = 1; j = \overline{1, k}.$$

$$r_{\dot{x}_j \dot{y}} = \frac{\sum_{i=1}^n (\dot{x}_{ji} - \bar{\dot{x}}_j)(\dot{y}_i - \bar{\dot{y}})}{(n-1)\sigma_{\dot{x}_j}\sigma_{\dot{y}}} = \frac{1}{n-1} \sum_{i=1}^n \dot{x}_{ji}\dot{y}_i; j = \overline{1, k},$$

$$r_{\dot{x}_j \dot{y}} = \frac{1}{n-1} \sum_{i=1}^n \dot{x}_j \dot{x}_l; j = \overline{1, k}, j \neq l.$$

$$\sigma_{\dot{x}_j}^2 = \frac{1}{n-1} \sum_{i=1}^n (\dot{x}_{ji} - \bar{\dot{x}}_j)^2 = \frac{1}{n-1} \sum_{i=1}^n \dot{x}_{ji}^2; j = \overline{1, k};$$

$$\sigma_{\dot{y}}^2 = \frac{1}{n-1} \sum_{i=1}^n \dot{y}_i^2.$$

Вычисленные по этим формулам выборочные коэффициенты корреляции равны коэффициентам корреляции между переменными, выраженными в натуральном масштабе, т.е.

$$r_{\dot{x}_j \dot{x}_i} = r_{x_j x_i}; r_{\dot{x}_j \dot{y}} = r_{x_j y}.$$

Решая систему линейных уравнений (3.90), находим A_j , $j = \overline{1, k}$. Если все входные воздействия линейно независимы, то $r_{\dot{x}_k \dot{x}_1} = 0$ и имеем $A_j = r_{\dot{x}_k \dot{x}_j}$, $j = \overline{1, k}$.

Для изучения тесноты связи между функцией отклика Y и факторами (x_1, x_2, \dots, x_k) используют коэффициент множественной корреляции

$$R = \sqrt{A_1 r_{\dot{x}_1 \dot{x}_k} + A_2 r_{\dot{x}_2 \dot{x}_k} + \dots + A_k r_{\dot{x}_k \dot{x}_k}}. \quad (3.91)$$

Очевидно, что в случае линейно независимых входов

$$R = \sqrt{r_{\dot{x}_1 \dot{x}_k}^2 + r_{\dot{x}_2 \dot{x}_k}^2 + \dots + r_{\dot{x}_k \dot{x}_k}^2}. \quad (3.92)$$

R всегда положителен и принимает значение от 0 до 1. Чем больше R , тем лучше качество предсказаний данной моделью опытных данных.

От уравнения (3.88) можно перейти к уравнению (3.83) с переменными в натуральных масштабах, используя формулы (3.87) нормирования переменных входа X_j :

$$\dot{x}_j = \frac{x_j - \bar{x}_j}{\sigma_{x_j}}, \quad j = \overline{1, k}.$$

Нетрудно получить следующие формулы для пересчета коэффициентов модели:

$$a_j = \frac{\sigma_y}{\sigma_{x_j}}, \quad j = \overline{1, k}; \quad a_0 = \bar{Y} - \sum_{j=1}^k a_j \bar{x}_j.$$

Нелинейная регрессионная модель многомерного объекта. Уравнение нелинейной регрессии стохастического многомерного объекта задается обычно полиномом

$$M\left[\frac{Y}{\vec{x}}\right] = a_0 + \sum_{j=1}^k a_j x_j + \sum_{j=1}^k b_j x_j^2 + \sum_{l,j=1, j < l}^k b_{jl} x_j x_l + \dots, \quad (3.93)$$

где x_j – входные переменные объекта, $j = \overline{1, k}$.

Коэффициенты уравнения (3.93), как правило, определяются методом наименьших квадратов. Наибольшие трудности вызывает структурная иденти-

фикация модели, т.е. выбор порядков полинома по каждому из входов, а также вычисление определителя плохо обусловленной матрицы, часто встречающейся при идентификации коэффициентов модели.

Использование метода Брандона для построения уравнения множественной регрессии позволяет избежать эти трудности.

По этому методу уравнение регрессии записывается в виде

$$M\left[\frac{Y}{\vec{x}}\right] = af_1(x_1)f_2(x_2)\dots f_k(x_k), \quad (3.94)$$

где $f_j(x_j)$ – неизвестная функция переменной x_j .

Порядок расположения факторов x_1, x_2, \dots, x_k в уравнении (3.94) не безразличен. Чем сильнее влияние на Y оказывает вход x_j , тем меньше должен быть порядковый номер индекса j . Это необходимо для повышения точности обработки результатов наблюдений.

Вид функции $f_j(x_j)$ выбирается при помощи построения эмпирических линий регрессии. Вначале по точкам выборки (y_i, \vec{x}_i) , $\vec{x}_i = (x_{1i}, x_{2i}, \dots, x_{ki})$, $i = \overline{1, n}$, строятся поле корреляции и эмпирическая линия регрессии $M\left[\frac{Y}{x_1}\right]$ (рис. 3.9)

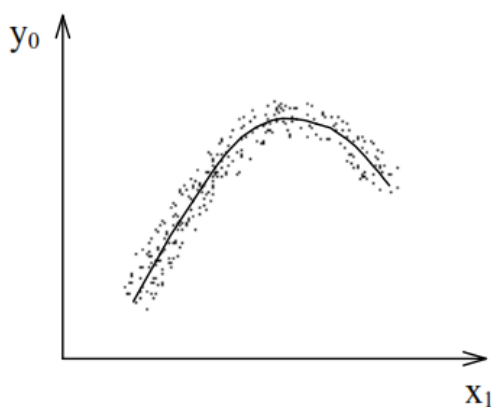


Рис. 3.9

По эмпирической линии регрессии определяется вид зависимости $M\left[\frac{Y}{x_1}\right] = f_1(x_1)$ и методом наименьших квадратов рассчитываются коэффициенты этого уравнения регрессии. Затем составляется выборка новой величины (y_{1i}, \vec{x}_i) , где

$$Y_{1i} = \frac{Y_i}{f_1(x_{1i})}, \quad i = \overline{1, n}. \quad (3.95)$$

Полученная величина Y_1 не зависит уже от x_1 , а определяется только параметрами x_1, x_2, \dots, x_k . Поэтому можно записать

$$M\left[\frac{Y_1}{\vec{x}}\right] = af_2(x_2)f_3(x_3)\dots f_k(x_k). \quad (3.96)$$

По точкам новой выборки величин Y_1 и x_2 вновь строятся корреляционное поле и эмпирическая линия регрессии, характеризующая зависимость Y_1 от x_2 (рис. 3.10):

$$M\left[\frac{Y}{x_2}\right] = f_2(x_2). \quad (3.97)$$

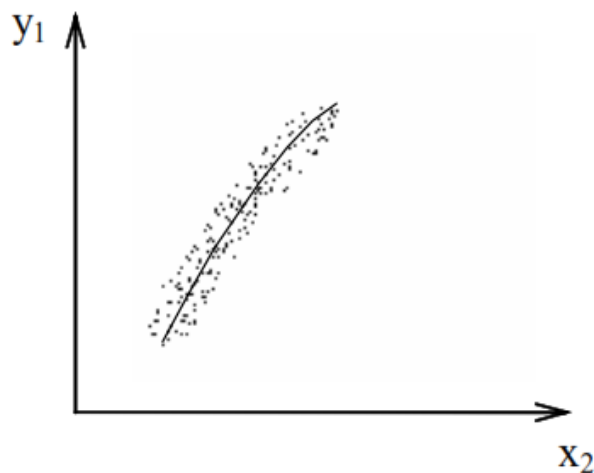


Рис. 3.10

Рассчитываются коэффициенты уравнения (3.97) и вновь составляются выборки новой величины (y_{2i}, \vec{x}_i) , где

$$Y_{2i} = \frac{Y_{1i}}{f_2(x_{2i})} = \frac{Y_i}{f_1(x_{1i})f_2(x_{2i})}, \quad i = \overline{1, n}. \quad (3.98)$$

Полученная величина Y_2 не зависит уже от двух факторов x_1 и x_2 и может быть определена из следующего уравнения регрессии:

$$M\left[\frac{Y_2}{\vec{x}}\right] = af_3(x_3)f_4(x_4)\dots f_k(x_k). \quad (3.99)$$

Такая процедура определения функций $f_3(x_3), f_4(x_4)$ и т.д. продолжается до получения выборки величины Y_k :

$$Y_{ki} = \frac{Y_{k-1,i}}{f_k(x_k)} = \frac{Y_{0i}}{f_1(x_{1i})f_2(x_{2i})\dots f_k(x_{ki})}, \quad i = \overline{1, n}. \quad (3.100)$$

Полученная величина Y_k не зависит от факторов x_1, x_2, \dots, x_k и определяется коэффициентом исходного уравнения

$$M\left[\frac{Y_k}{\vec{x}}\right] = a = \frac{1}{n} \sum_{i=1}^n Y_{ki}, \quad (3.101)$$

где n – объем выборки.

Таким образом, окончательно получаем зависимость условного среднего выхода объекта Y от всех факторов (входов) x_1, x_2, \dots, x_k подставляя в уравнение (3.94) численное значение коэффициента a и найденные выражения функций $f_j(x_j)$, $j = \overline{1, k}$. Таким образом, одновременно решаются задачи структурной и параметрической идентификации.

Оценка значимости коэффициентов и адекватности регрессионной модели. После того как уравнение регрессии найдено, необходимо провести статический анализ результатов. Он заключается в проверке значимости коэффициентов модели в сравнении с ошибкой воспроизводимости и оценке адекватности модели эксперименту. Такое исследование называется регрессионным анализом.

Оценку значимости коэффициентов модели выполняют по критерию Стьюдента. Для каждого проверяемого коэффициента a_j вычисляют

$$t_j = \frac{|a_j|}{\sigma_{aj}}, \quad (3.102)$$

где σ_{aj} – СКО j -го коэффициента.

Как правило, для получения дисперсии воспроизводимости проводится отдельная серия из m повторных опытов при фиксированных значениях входов. Пусть при этом получено y_u^0 , $u = \overline{1, m}$. Тогда дисперсия воспроизводимости определяется по формуле

$$\sigma_{\%00}^2 = \frac{1}{m-1} \sum_{u=1}^m (y_u^0 - \bar{y}^0)^2; \quad \bar{y}^0 = \frac{1}{m} \sum_{u=1}^m y_u^0. \quad (3.103)$$

Тогда

$$\sigma_{a_j} = \sqrt{\sum_{i=1}^n \left(\frac{\partial a_j}{\partial y_j} \right)^2} \sigma_{\text{воспр}}^2. \quad (3.104)$$

Значения частных производных находят из выражения коэффициентов, найденных, например, методом наименьших квадратов.

Если t_j больше табличного $t(\alpha, f)$ (находят из таблицы распределения Стьюдента) для выбранного уровня значимости $\alpha = 1 - P$ (P – доверительная вероятность) и числа степеней свободы $f = f_{\text{воспр}} = m - 1$, то коэффициент a_j значительно отличается от нуля.

Незначимые коэффициенты исключаются из уравнения регрессии. Оставшиеся коэффициенты пересчитываются заново, поскольку коэффициенты закоррелированы друг с другом.

Адекватность модели проверяется по F – критерию Фишера. Для этого вычисляют дисперсию адекватности

$$\sigma_{\text{ад}}^2 = \frac{1}{n-l} \sum_{i=1}^n \left(Y_i - M \left[\frac{Y}{x_i} \right] \right)^2, \quad (3.105)$$

где l – число коэффициентов модели, найденных по выборке объема n .

Определяют значение F – критерия ($F \geq 1$) по формуле

$$F = \begin{cases} \sigma_{\text{ад}}^2 / \sigma_{\text{воспр}}^2, & \sigma_{\text{ад}}^2 > \sigma_{\text{воспр}}^2; \\ \sigma_{\text{воспр}}^2 / \sigma_{\text{ад}}^2, & \sigma_{\text{ад}}^2 \leq \sigma_{\text{воспр}}^2. \end{cases} \quad (3.106)$$

Если F окажется меньше табличного значения $F_m(\alpha, f_{\text{ад}}, f_{\text{воспр}})$ (находятся из таблицы распределения Фишера) для уровня значимости α и степеней свободы $f_{\text{ад}} = n - 1$, $f_{\text{воспр}} = m - 1$, модель адекватна объекту (эксперименту).

3.6. Идентификация динамических объектов

Оператор F_0 таких объектов имеет память. В этом случае выход Y в момент времени t отражает не только, точнее не столько состояние входа X в этот момент, сколько его значения в предыдущие моменты времени. Объект как бы запоминает возмущение $X(t)$ и реагирует на него лишь через определенный промежуток времени. Поэтому динамические объекты часто называют инерционными.

Здесь будем рассматривать идентификацию непрерывной модели линейного детерминированного стационарного объекта. Эта модель для одномерного случая представляет собой обыкновенное дифференциальное уравнение вида

$$\sum_{i=0}^p a_i y^{(i)}(t) = \sum_{j=0}^l b_j x^{(j)}(t), \quad (3.107)$$

где $y^{(j)}(t), x^{(j)}(t)$ – j -я производная функций y и x , т.е. $y^{(j)} = \frac{d^j y}{dt^j}$,

$$x^{(j)} = \frac{d^j x}{dt^j}, \quad a_p = 1, \quad l \leq p.$$

Модель определяется $(p + l + 1)$ параметрами

$$C = (a_0, a_1, \dots, a_{p-1}, b_0, b_1, \dots, b_l). \quad (3.108)$$

Структурными параметрами модели (3.107) являются числа p и l , которые должны быть выбраны в процессе структурной идентификации. Модель (3.107) часто удобно записать в виде системы дифференциальных уравнений первого порядка (в форме Коши), обозначив $y_1(t) = y(t)$:

$$\left. \begin{aligned} \dot{y}_1(t) &= y_2(t); \\ \dot{y}_2(t) &= y_3(t); \\ &\dots\dots\dots \\ \dot{y}_{p-1}(t) &= y_p(t); \\ \dot{y}_p(t) &= -\sum_{i=0}^{p-1} a_i y_{i+1}(t) + \sum_{j=0}^l b_j x^{(j)}(t). \end{aligned} \right\} \quad (3.109)$$

Естественно обобщить эту модель, приведя ее к виду:

$$\left. \begin{aligned} \dot{y}_1(t) &= -\sum_{i=1}^p a_{1i} y_i(t) + \sum_{j=0}^l b_{1j} x^{(j)}(t); \\ &\dots\dots\dots \\ \dot{y}_p(t) &= -\sum_{i=1}^p a_{pi} y_i(t) + \sum_{j=0}^l b_{pj} x^{(j)}(t). \end{aligned} \right\} \quad (3.110)$$

Если модель (3.107) сводится к виду (3.110), то последняя в общем виде не сводится к виду (3.107). Поэтому модель (3.110) следует считать наиболее общей формой непрерывной параметрической модели динамического детерми-

нированного линейного стационарного объекта. Эту модель удобно записывать в векторной форме

$$\vec{Y}'(t) = A\vec{Y}(t) + B\vec{X}'(t), \quad (3.111)$$

где $\vec{Y}(t) = (y_1(t), y_2(t), \dots, y_p(t))$ – вектор состояния,

$\vec{X}'(t) = (x(t), x^{(1)}(t), \dots, x^{(l)}(t))$ – вектор возмущения,

A, B – матрицы коэффициентов модели, причем $A = (a_{qi})$ – квадратная матрица ($p \times p$), $B = (a_{qj})$ – прямоугольная матрица ($p \times (l+1)$).

Таким образом, идентифицируемыми параметрами модели (3.111) являются элементы матрицы A и B , т.е. $p(p+l+1)$ параметров модели, которые и образуют две матрицы неизвестных параметров $C = (A, B)$.

Исходной информацией для идентификации в этом случае являются реализации $x(t), y(t)$ в промежутке $0 \leq t \leq T$, характеризующие состояния входа и выхода идентифицируемого объекта

$$\langle x(t), y(t) \rangle \quad (0 \leq t \leq T). \quad (3.112)$$

Распространим сказанное на многомерный случай, когда число входов $n > 1$, а число выходов $m = 1$, т.е. вектор входа имеет вид:

$$\vec{X}(t) = (x_1(t), \dots, x_n(t)).$$

В силу линейности вектор состояния \vec{Y} образуется как сумма векторов:

$$\vec{Y} = \sum_{i=1}^n \vec{Y}'_i, \quad (3.113)$$

где вектор

$$\vec{Y}'_i(t) = A_i y_i(t) + B_i \vec{X}'_i(t), \quad (3.114)$$

отражает влияние i -го входа.

Здесь $\vec{X}'_i(t) = (x(t), x^{(1)}(t), \dots, x^{(l)}(t))$ – вектор возмущения i -го входа; A_i, B_i – матрицы влияния i -го входа.

Теперь, подставляя (3.113) и (3.114) в (3.111), можно записать уравнение для линейной динамической модели со многими входами

$$\vec{Y}'(t) = \sum_{i=1}^n A_i y_i(t) + \sum_{i=1}^n B_i \vec{X}'_i(t). \quad (3.115)$$

Эта модель имеет $np(p+l+1)$ идентифицируемых параметров. При этом информация о работе объекта, необходимая для идентификации, имеет вид:

$$\langle \vec{X}(t), y(t) \rangle (0 \leq t \leq T).$$

Случай $n > 1$ и $m > 1$ сводится к m -кратному повторению предыдущего.

Идентификация параметров модели. Исходной информацией для построения процедуры идентификации является вид идентифицируемой модели (3.107) и наблюдения (3.112). Необходимо определить $(p+l+1)$ параметров

$$a_i, (i = \overline{0, p-1}) \text{ и } b_j, (j = \overline{0, l}).$$

В общем случае, т.е. при произвольных значениях параметров a_i и b_j , при подстановке реализаций (3.112) в уравнение модели (3.107) равенство в (3.107) выполняться не будет. Естественно подобрать параметры a_i и b_j такими, чтобы равенство (3.107) было восстановлено или хотя бы правая и левая части выражения (3.107) отличались друг от друга наименьшим образом. Чтобы выполнять это условие, необходимо построить функцию невязки правой и левой частей этого уравнения и минимизировать ее с помощью вариации параметров a_i и b_j . Сформулируем требования, которым должна удовлетворять эта функция невязки:

- она не должна принимать отрицательных значений;
- ее минимум должен соответствовать решению поставленной задачи;
- этот минимум должен быть равен нулю в случае совпадения структуры модели и объекта.

Этим условиям удовлетворяет невязка в виде среднего квадрата разности правой и левой части уравнения модели (3.107) при подстановке туда функций $x(t)$, $y(t)$ – наблюдений объекта, т.е.

$$E(C) = \int_0^T \left[\sum_{i=0}^p a_i y^{(i)}(t) - \sum_{j=0}^l b_j x^{(j)}(t) \right]^2 dt. \quad (3.116)$$

Минимизируя это выражение по параметрам a_i и b_j , которые образуют вектор (3.108), определим искомые параметры. Такая задача минимизации формулируется в виде

$$E(C) = \min_C \Rightarrow C^* = (a_0^*, \dots, a_{p-1}^*, b_0^*, \dots, b_l^*). \quad (3.117)$$

Результат минимизации C^* и дает значения идентифицируемых параметров, для которых при совпадении структуры объекта и модели $E(C^*)=0$.

Для решения задачи (3.117) воспользуемся тем, что функция $E(C)=0$ имеет простую структуру относительно идентифицируемых параметров — это квадратичная форма.

Задача минимизации гладких функций (а квадратичная форма – гладкая функция) сводится к решению системы уравнений, которая образуется в результате приравнивания нулю частных производных:

$$\frac{\partial E(C)}{\partial a_i} = 0, \quad (i = \overline{0, p-1});$$

$$\frac{\partial E(C)}{\partial b_j} = 0, \quad (j = \overline{0, l}).$$

После очевидных преобразований получим следующую систему линейных алгебраических уравнений:

$$\left. \begin{aligned} \sum_{i=0}^p a_i \phi_{ik} - \sum_{j=0}^l b_j \lambda_{jk}, \quad (k = \overline{0, p-1}); \\ \sum_{i=0}^p a_i \lambda_{iq} - \sum_{j=0}^l b_j v_{qj}, \quad (q = \overline{0, l}), \end{aligned} \right\} \quad (3.118)$$

где, напомним, $ap = 1$ и

$$\left. \begin{aligned} \phi_{ik} &= \int_0^T y^{(i)}(t) y^{(k)}(t) dt; \\ \lambda_{jk} &= \int_0^T x^{(j)}(t) y^{(k)}(t) dt; \\ v_{iq} &= \int_0^T x^{(i)}(t) x^{(q)}(t) dt, \end{aligned} \right\} \quad (3.119)$$

где $\phi_{ik}, \lambda_{jk}, v_{iq}$ – коэффициенты системы уравнений (3.118).

Таким образом, задача идентификации линейной динамической стационарной системы сведена к решению системы линейных алгебраических уравнений, что осуществляется стандартными вычислительными методами.

Однако для вычисления выражений (3.119) необходимо знать производные входного $x(t)$ и выходного $y(t)$ сигналов объекта. Их можно получить двумя способами:

1) численное дифференцирование – данный способ связан с локальной (точечной) оценкой производных:

$$\left. \begin{aligned} \dot{z}(t_j) &= \frac{1}{\Delta t} [z(t_j) - z(t_j - \Delta t)]; \\ \ddot{z}(t_j) &= \frac{1}{\Delta t} [\dot{z}(t_j) - \dot{z}(t_j - \Delta t)]; \\ &\dots\dots\dots \\ z^{(i)}(t_j) &= \frac{1}{\Delta t} [z^{(i-1)}(t_j) - z^{(i-1)}(t_j - \Delta t)]; \end{aligned} \right\} \quad (3.120)$$

где Δt – интервал, характеризующий базу оценок.

Но этот способ обладает двумя недостатками.

Во-первых, точность оценки производных очень быстро падает с ростом порядка производной. Практически удовлетворительно не удастся оценить производные выше второго порядка.

Во-вторых, для оценки i -й производной при $t = 0$ необходимо иметь значения функции при $t < 0$, т.е. $z(-\Delta t), z(-2\Delta t), \dots, z(-i\Delta t)$. Так как этих значений нет в измерениях, то приходится для (3.119) брать пределы интегрирования не $[0, T]$, а $[i\Delta t, T]$, где i – наибольший порядок производной подынтегральной функции.

2) в ряде случаев целесообразно использовать разложение $z(t)$ в ряд по системе заданных в аналитическом виде функций. Тогда полученную аналитическую зависимость, аппроксимирующую $z(t)$, можно дифференцировать нужное число раз.

Все вышеизложенное относилось к простейшей модели (3.107). Если же объект описывается обобщенной моделью (3.111), то функция невязки будет не одна. Количество их будет равно P по числу уравнений (3.110). Для k -ой невязки получаем из k -го уравнения модели (3.110):

$$E_k(a_{k1}, \dots, a_{kp}, b_{k0}, \dots, b_{kl}) = \int_0^T \left[\dot{y}_k(t) - \sum_{i=0}^p a_{ki} y_i(t) - \sum_{j=0}^l b_{kj} x^{(j)}(t) \right]^2 dt, \quad (3.121)$$

где $k = \overline{0, p}$.

Дифференцируя эти невязки по коэффициентам a и b и приравнивая производные нулю, получаем систему линейных уравнений, откуда определяются идентифицируемые параметры объекта:

$$\frac{\partial E_k}{\partial a_{ki}} = 0, \quad \frac{\partial E_k}{\partial b_{kj}} = 0, \quad (i = \overline{1, p}); (j = \overline{0, l}), (k = \overline{1, p}). \quad (3.122)$$

Теперь рассмотрим условия применимости рассмотренного метода идентификации. Как было показано, он сводится к решению системы линейных алгебраических уравнений или аналогичной ей системы, получающейся из (3.122). Однозначность решения задачи идентификации требует однозначности решения системы (3.118). А это выполняется только в случае, если ее определить (3.123)

$$\Delta = \begin{vmatrix} \phi_{01} \dots \phi_{p1} & \lambda_{11} \dots \lambda_{l1} \\ \dots & \dots \\ \phi_{0p-1} \dots \phi_{pp-1} & \lambda_{1p-1} \dots \lambda_{lp-1} \\ \lambda_{10} \dots \lambda_{1p} & v_{11} \dots v_{l1} \\ \dots & \dots \\ \lambda_{l0} \dots \lambda_{lp} & v_{l1} \dots v_{ll} \end{vmatrix} \neq 0. \quad (3.123)$$

Это и есть условие однозначности решения задачи идентификации. Если оно не выполняется, то следует либо взять другую реализацию входа и выхода объекта $\langle x(t), y(t) \rangle$, либо уменьшить количество идентифицируемых параметров, т.е. уменьшить числа p и l , тем самым понизить порядок модели.

Обе из указанных мер могут привести к выполнению условия (3.123). Если же, несмотря на предпринятые меры, это условие по-прежнему не выполняется, то следует снова обратиться к одной из них и т.д. до тех пор, пока не выполнится условие (3.123). При этом следует помнить, что если первая мера (новая реализация) не всегда приводит к цели, то вторая (понижение порядка модели) при последовательном применении всегда приводит к выполнению условия (3.123).

Аналогично формулируется условие однозначности для обобщенной модели. Здесь определитель системы (3.122) будет более высокого порядка.

3.7. Идентификация закона распределения случайной величины

Случайная величина. В инженерной практике необходимо исследовать не только детерминированные, но и стохастические процессы. Практически все процессы в технологических и технических объектах выполняются в непрерыв-

но меняющихся непредвиденным образом условиях. В связи с этим приходится анализировать случайные величины в этих объектах.

Под случайной величиной понимают величину, принимающую в результате испытания значение, которое принципиально нельзя предсказать исходя из условий опыта. Случайная величина обладает целым набором допустимых значений, но в результате каждого отдельного опыта принимает лишь какое-то одно из них. Она может принимать различные значения даже при неизменном комплексе основных факторов.

Различают дискретные и непрерывные случайные величины. Возможные значения дискретных случайных величин можно заранее перечислить. Значения непрерывной случайной величины не могут быть заранее перечислены, они непрерывно заполняют некоторый промежуток. Чтобы полностью охарактеризовать случайную величину, необходимо не только указывать, какие значения она может принимать, но и как часто.

Законы распределения случайной величины. При решении многих прикладных задач необходимые вероятностные характеристики соответствующих случайных величин неизвестны исследователю и должны определяться по экспериментальным данным. Такое статистическое описание результатов наблюдений, построение и проверка различных математических моделей, использующих понятие вероятности, составляет основное содержание математической статистики.

Фундаментальными понятиями статистической теории являются понятия вероятности и распределения вероятности случайной величины.

Пусть дискретная случайная величина Z может принимать в результате опыта значения z_1, z_2, \dots, z_n . Отношение числа опытов n_i , в результате которых случайная величина Z приняла значение z_i , к общему числу производимых опытов n называется частотой проявления события $Z = z_i$ (частотное определение вероятности). Частота $\frac{n_i}{n}$ сама является величиной и меняется в зависимости от количества произведенных опытов. Но при большом числе опытов она имеет тенденцию стабилизироваться около некоторого значения P_i называемого вероятностью события $Z = z_i$ (статистическое определение вероятности):

$$P_i = P\{Z = z_i\} = \lim_{n \rightarrow \infty} \frac{n_i}{n}. \quad (3.124)$$

Сумма вероятностей всех возможных значений случайной величины равна единице

$$\sum_{i=1}^n P_i = 1, \quad (3.125)$$

так как тот факт, что случайная величина примет в результате опыта одно из своих значений, есть достоверное событие. Эта суммарная вероятность распределена определенным образом между отдельными значениями.

Дискретную случайную величину можно полностью задать вероятностным рядом, указав вероятность P_i для каждого значения z_i :

z_i	z_1	z_2	z_3	...	z_n
P_i	P_1	P_2	P_3	...	P_n

Всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями, называется *законом распределения*. Вероятностный ряд является одним из видов законов распределения случайной величины.

Распределение непрерывной случайной величины нельзя задавать при помощи вероятностей отдельных значений. Число значений так велико, что для большинства из них вероятность принять эти значения равна нулю, т.е. событие может произойти, а вероятность его равна нулю. Для непрерывных случайных величин изучается вероятность того, что в результате опыта значение случайной величины попадет в некоторую заданную область.

Аналитическими выражениями законов распределения случайной величины являются функции распределения вероятностей - интегральная и дифференциальная.

Интегральная функция распределения $F(z)$ случайной величины Z показывает вероятность того, что случайная величина не превышает некоторого заданного или текущего значения z , т.е.

$$F(z) = P\{Z \leq z\}. \quad (3.126)$$

Следовательно, вероятность того, что значение случайной величины Z заключено между z_1 и z_2 , равна разности значений функции распределения, вычисленных в двух этих точках:

$$P\{z_1 < Z \leq z_2\} = F(z_2) - F(z_1). \quad (3.127)$$

Аналогично,

$$P\{Z > z\} = 1 - F(z). \quad (3.128)$$

Интегральная функция распределения случайной величины обладает следующими свойствами:

- 1) $\lim_{z \rightarrow -\infty} F(z) = F(-\infty) = 0$;
- 2) $\lim_{z \rightarrow \infty} F(z) = F(\infty) = 1$;
- 3) $F(z) \geq 0$ для всех z ;
- 4) $F(z_2) \geq F(z_1)$ если $z_2 > z_1$.

В виде функции распределения можно задать распределение как непрерывной, так и дискретной случайной величины. Как видно из определения, $F(z)$ есть неубывающая функция z . Ордината кривой $F(z)$, соответствующая точке z_1 , представляет собой вероятность того, что случайная величина Z при испытании окажется меньше z_1 .

Интегральная функция распределения дискретной случайной величины всегда есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям этих значений. Сумма всех скачков равна единице.

Возможные виды интегральных функций распределения изображены на рис. 3.11 для непрерывной (а) и дискретной (б) случайной величины:

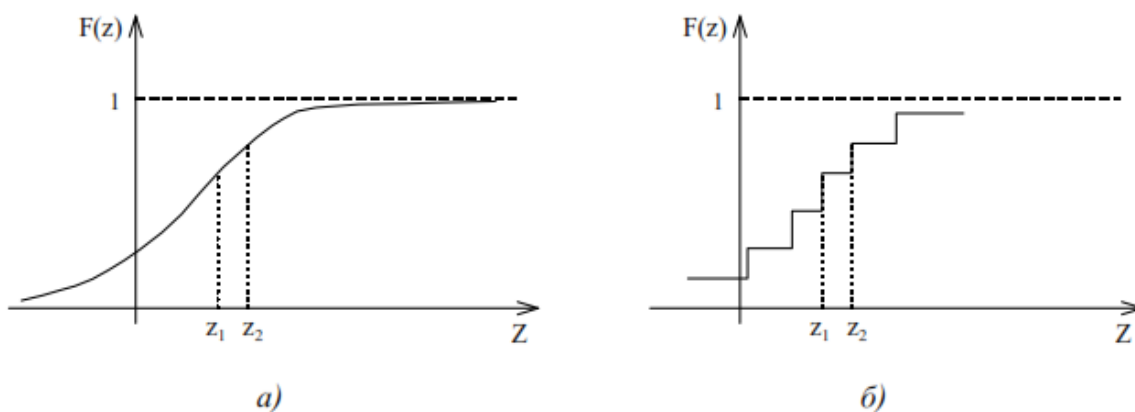


Рис. 3.11

Если функция $F(z)$ дифференцируема для всех значений случайной величины Z , то закон распределения вероятностей может быть выражен в аналитической форме также с помощью дифференциальной функции распределения вероятностей:

$$f(z) = \frac{dF(z)}{dz} = \lim_{\Delta z \rightarrow 0} \frac{P\{z < Z \leq z + \Delta z\}}{\Delta z}, \quad (\Delta z > 0). \quad (3.129)$$

Таким образом, значение функции $f(z)$ приближено равно отношению вероятности попадания случайной величины в интервал $(z, z + \Delta z)$ к длине Δz этого интервала, когда Δz – бесконечно малая величина. Поэтому функцию $f(z)$ называют также функцией плотности распределения вероятностей (или короче – функцией плотности вероятности).

Отметим основные свойства функции $f(z)$:

- 1) $f(z) \geq 0$;
- 2) $\int_{-\infty}^{\infty} f(z) dz = 1$;

$$3) \int_{-\infty}^z f(x)dx = F(z);$$

$$4) \lim_{|z| \rightarrow \infty} F(z) = 0, \text{ (} x \text{ – переменная интегрирования).}$$

С помощью дифференциальной функции распределения вычисляется вероятность нахождения случайной величины в любой области из множества ее возможных значений. Например,

$$P\{Z \leq z_1\} = \int_{-\infty}^{z_1} f(z)dz; \quad (3.130)$$

$$P\{Z > z_2\} = \int_{z_2}^{-\infty} f(z)dz; \quad (3.131)$$

$$P\{z_1 < Z \leq z_2\} = \int_{z_1}^{z_2} f(z)dz. \quad (3.132)$$

Для непрерывной случайной величины вероятность можно определить как относительную долю площади под кривой плотности распределения вероятностей $f(z)$. Так, например, вероятность того, что случайная величина Z примет значение, меньшее z_1 , равна относительной доле площади под кривой $f(z)$ слева от точки z_1 (рис. 3.12,а); вероятность того, что эта величина Z примет значение, большее z_2 , равна относительной доле площади под кривой $f(z)$ справа от точки z_2 (рис. 3.12,б); вероятность того, что она примет значение, заключенное между значениями z_1 и z_2 , равна относительной доле площади под кривой $f(z)$ между точками z_1 и z_2 (рис. 3.12,в).

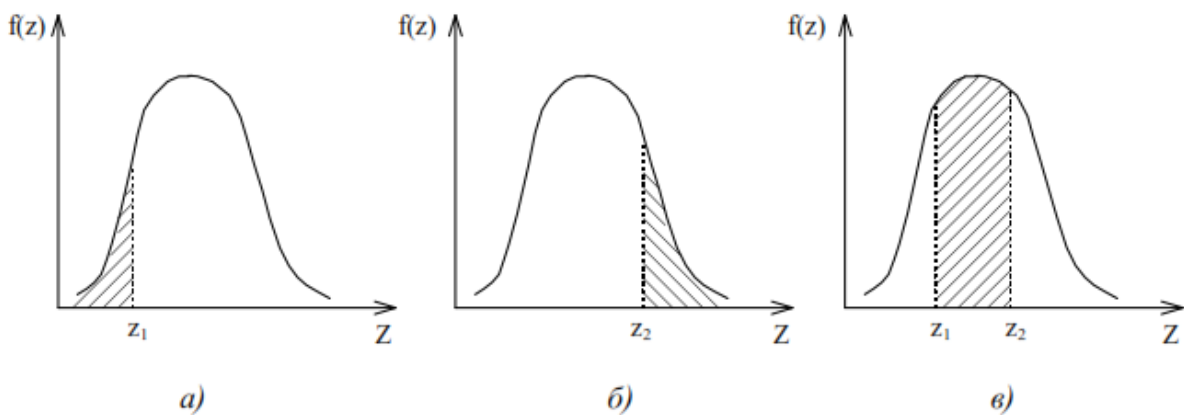


Рис. 3.12

Теоретическое и эмпирическое распределение. Распределение вероятностей называется теоретическим. Оно может быть установлено для соответствующей случайной величины из анализа генеральной совокупности.

Под генеральной совокупностью понимают все допустимые значения случайной величины. Содержательный смысл этого понятия состоит в том, что предполагается существование некоторых вполне определенных свойств, неслучайных закономерностей присущих данной совокупности, тех свойств, которые и должны быть выяснены исследователем. Фактически эти свойства являются объективным отображением вероятностных свойств изучаемого объекта, которые могут быть охарактеризованы с помощью соответствующих законов распределения вероятностей или связанных с ними числовых параметров. Считается, что указанные свойства не изменяются во времени и присущие генеральной совокупности неслучайные закономерности сохраняют постоянным свой характер, т.е. являются устойчивыми.

На практике исследователь всегда располагает лишь ограниченным числом значений случайной величины, представляющим собой некоторую выборку из генеральной совокупности.

Выборка – это конечный набор значений случайной величины, полученный в результате наблюдений. Число элементов выборки называется ее объемом. Если, например, z_1, z_2, \dots, z_n – наблюдаемые значения случайной величины Z (возможно, и совпадающие), то объем данной выборки равен n .

Выборка называется репрезентативной (представительной), если она достаточно полно характеризует генеральную совокупность. Для обеспечения репрезентативности выборки чаще всего используется случайный выбор элементов (рандомизация). При таком выборе предполагается, что каждая возможная выборка фиксированного объема имеет одну и ту же вероятность выбора, а последовательные наблюдения взаимно независимы.

Смысл статистических методов заключается в том, что по выборке ограниченного объема n , т. е. по некоторой части генеральной совокупности, высказывать обоснованное суждение о ее свойствах в целом. Подобное суждение может быть получено путем построения эмпирических (выборочных) аналогов вероятностных характеристик исследуемой величины, иначе говоря, путем оценивания параметров (характеристик) генеральной совокупности с помощью некоторых подходящих функций от результатов наблюдений – оценок.

При многократном извлечении выборок одного и того же объема и последующем нахождении множества оценок одного и того же параметра получают различные числовые значения этих оценок, изменяющиеся от одной выборки к другой случайным образом. Иными словами, любая оценка произвольного параметра α есть случайная величина. В этом ее принципиальное отличие от самого оцениваемого параметра α , являющегося неслучайным.

Оценка $\tilde{\alpha}$ параметра α называется состоятельной, если при неограниченном увеличении объема выборки n значение $\tilde{\alpha}$ с полной мерой достоверности (с вероятностью единица) стремится к своему теоретическому значению α .

Состоятельность оценки гарантирует исследователю увеличение точности оценивания с ростом n и то, что хотя бы в пределе при $n \rightarrow \infty$ он может получить точное значение α .

Оценка $\tilde{\alpha}$ называется несмещенной, если $m_{\tilde{\alpha}} = \alpha$ для любого n ($m_{\tilde{\alpha}}$ – математическое ожидание $\tilde{\alpha}$). Несмещенность означает отсутствие систематической погрешности при оценивании параметра α .

Предположим, что в результате эксперимента получена выборка z_1, z_2, \dots, z_n значений случайной величины Z . Пусть z – некоторая точка оси z . Обозначим через n_z число (частота) выборочных значений, расположенных левее z на этой же оси. Отношение n_z/n представляет собой относительную частоту наблюдаемых в выборке значений случайной величины Z , меньших z . Эта частота есть функция от z . Обозначим ее $F^*(z)$:

$$F^*(z) = \frac{n_z}{n}. \quad (3.133)$$

Функция распределения $F^*(z)$, получаемая по выборке, называется эмпирической или выборочной функцией распределения (в отличие от распределения генеральной совокупности, или теоретического распределения). Она устанавливает распределение относительных частот, т.е. определяет для каждого значения z относительную частоту события $Z < z$.

Для каждой выборки эмпирическая функция распределения будет своей, но все эмпирические функции распределения одной и той же случайной величины будут иметь нечто общее, что является информацией о функции распределения этой случайной величины.

Доказано, что с вероятностью 1 при $n \rightarrow \infty$ максимальная разность между функциями распределения случайной величины $F^*(z)$ и $F(z)$ стремится к нулю (теорема Гливенко):

$$P\{\sup |F^*(z) - F(z)| \rightarrow 0\} = 1. \quad (3.134)$$

Практически это означает, что при достаточно большой выборке функцию распределения генеральной совокупности приблизительно можно заменить функцией распределения выборки.

Эмпирическая функция распределения обладает следующими свойствами:

- 1) $F^*(z) \in [0,1] \forall z$, т.е. значения эмпирической функции принадлежат отрезку $[0,1]$;
- 2) $F^*(z_2) \geq F^*(z_1)$, если $z_2 > z_1$, т.е. $F^*(z)$ неубывающая функция;
- 3) если z_1 – наименьшее, а z_2 – наибольшее значение случайной величины Z , то $F^*(z) = 0$ при $Z \leq z_1$ и $F^*(z) = 1$ при $Z > z_2$.

График эмпирической функции распределения имеет вид, приведенный на рис. 3.11,б.

Полигоном эмпирического распределения называют ломаную, отрезки которой соединяют точки $(z_1 \cdot n_1), (z_2 \cdot n_2), \dots, (z_n \cdot n_n)$, где z_i – элементы выборки, а n_i – соответствующие им частоты, или точки $(z_1 \cdot P_1^*), (z_2 \cdot P_2^*), \dots, (z_n \cdot P_n^*)$, где $P_i^* = \frac{n_i}{n}$ – относительные частоты соответствующих z_i .

При обработке выборок больших объемов весь отрезок, в котором заключены все наблюдаемые значения случайной величины Z , разбивают на ряд интервалов длины h и находят n_i – число элементов выборки, попавших в i -ый интервал.

Гистограммой эмпирического распределения называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат интервалы длины h , а высоты равны отношению $\frac{n_i}{h}$ (плотность частоты). Площадь i -го прямоугольника равна $h \frac{n_i}{h} = n_i$. Площадь гистограммы частот равна сумме всех частот, т.е. объему выборки n .

Если учитываются не абсолютные частоты n_i , а относительные $P_i^* = \frac{n_i}{n}$ тогда гистограмма эмпирического распределения представляет собой ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат интервалы длины h , а высоты равны отношению $\frac{P_i^*}{h}$ (плотность относительной частоты). Площадь i -го прямоугольника равна $h \frac{P_i^*}{h} = P_i^*$. Площадь гистограммы относительных частот равна сумме всех относительных частот, т.е. единице.

На рис. 3.13 приведен возможный вид гистограммы эмпирического распределения случайной величины.

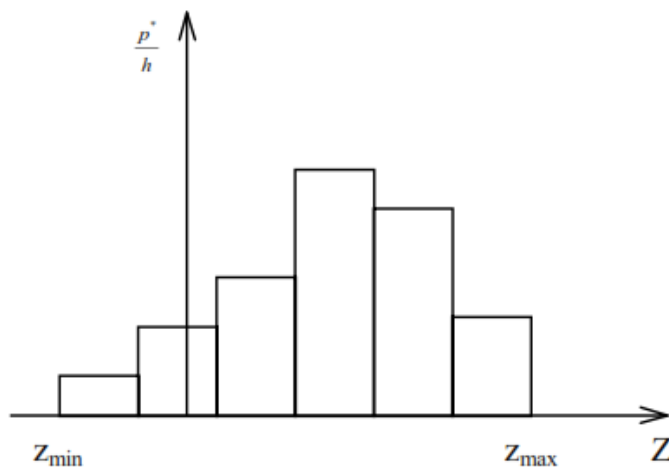


Рис. 3.13

Числовые характеристики случайной величины. Интегральная и дифференциальная функции распределения являются исчерпывающими вероятностными характеристиками случайной величины. Однако в прикладных задачах некоторые основные свойства случайных величин могут быть определены более просто с помощью определенных числовых параметров. Наибольшую роль среди них на практике играют два параметра, характеризующие центр рассеяния (центр распределения) случайной величины и степень ее рассеяния вокруг этого центра. Наиболее распространенной характеристикой центра распределения является математическое ожидание m_z случайной величины Z , часто называемое также генеральным средним значением или начальным моментом первого порядка.

Для дискретных случайных величин

$$m_z = M|Z| = \sum_{i=1}^N z_i P_i, \quad (3.135)$$

где $z_i (i = \overline{1, N})$ – элементы генеральной совокупности; P_i – соответствующие им вероятности.

Для непрерывных случайных величин математическое ожидание выражается интегралом:

$$m_z = M|z| = \int_{-\infty}^{\infty} z f(z) dz. \quad (3.136)$$

Степень рассеяния случайной величины z относительно m_z может быть охарактеризована с помощью генеральной дисперсии σ_z^2 . σ_z^2 есть математическое ожидание квадрата отклонения случайной величины от ее математического ожидания, т.е.

$$\sigma_z^2 = M[(Z - m_z)^2]. \quad (3.137)$$

Для дискретной случайной величины

$$\sigma_z^2 = \sum_{i=1}^N (Z - m_z)^2 P_i, \quad (3.138)$$

для непрерывной

$$\sigma_z^2 = \int_{-\infty}^{\infty} (Z - m_z)^2 f(z) dz. \quad (3.139)$$

Дисперсия σ_Z^2 называется также вторым центральным моментом случайной величины z (первый центральный момент всегда равен 0). Другие обозначения для дисперсии $D_z, D[z], s^2$.

Если $f(z)$ все с большей степени концентрируются вблизи m_z , то значения σ_Z^2 уменьшаются. Если же имеются весьма удаленные от m_z значения случайной величины z и для них $f(z)$ не слишком мала, то дисперсия σ_Z^2 увеличивается.

Квадратный корень из дисперсии σ_Z^2 называется средним квадратическим отклонением (или стандартом) σ_z :

$$\sigma_z = +\sqrt{\sigma_Z^2}. \quad (3.140)$$

Определение оценок математического ожидания \bar{Z} , дисперсии S_Z^2 и среднего квадратического отклонения S_z по выборке объема n производят по формулам:

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n z_i; \quad (3.141)$$

$$S_Z^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{Z})^2; \quad (3.142)$$

$$S_z = +\sqrt{S_Z^2}. \quad (3.143)$$

Отметим, что множитель $\frac{1}{n-1}$ (вместо $\frac{1}{n}$) в формуле (3.142) вводится для получения несмещенной оценки дисперсии.

Оценка результатов наблюдений случайной величины. Рассмотрим некоторые методы статистических оценок. Важнейшая задача математической статистики, решение которой позволило бы, в принципе, решить и все остальные задачи – это нахождение функции распределения наблюдаемой случайной величины.

Пусть имеется экспериментальная выборка большого объема n случайной величины z . Проверяется гипотеза о том, что функция распределения случайной величины есть $P = f(Z)$. Надо установить вид этого теоретического распределения и его адекватность.

Решение поставленной задачи включает следующие этапы:

1) построение эмпирической кривой (гистограммы) распределения случайной величины по заданной выборке и выбор типа теоретического распределения (выдвижение рабочей гипотезы);

2) определение параметров теоретического распределения;

3) проверка адекватности теоретического распределения (подтверждение рабочей гипотезы);

4) уточнение теоретического закона распределения, если первоначальная рабочая гипотеза отвергается.

Построение эмпирической кривой (гистограммы) распределения. Гистограмма $\tilde{f}(z)$ является эмпирическим аналогом функции плотности распределения $f(z)$.

При обработке выборок больших объемов используют метод «сгруппированных данных». Выборка объема n преобразуется в статистический ряд. Для этого весь диапазон изменения случайной величины в выборке $z_{\min} - z_{\max}$ делится на k интервалов. Число интервалов можно выбирать по полуэмпирической формуле (правило Штюргеса):

$$k = 1 + 3.21 \lg n, \quad (3.144)$$

где найденное значение округляется до ближайшего целого числа.

Длина интервала h равна

$$h = \frac{z_{\max} - z_{\min}}{k}. \quad (3.145)$$

Подсчитывается число элементов выборки $n_j (j = \overline{1, k})$, попавших в каждый интервал. Величина, равная

$$P_j^* = \frac{n_j}{n} \quad (3.146)$$

определяет относительную частоту попадания случайной величины в j -ый интервал.

Середина j -го интервала z_j^* определяется формулой

$$z_j^* = \frac{z_{j-1} + z_j}{2}. \quad (3.147)$$

Полученный статистический ряд можно записать в виде табл. 3.2.

Статистический ряд

Номер интервала	Интервал	z_j^*	n_j	P_j^*
1	$(z_{\min} = z_0, z_1)$	z_1^*	n_1	P_1^*
2	(z_1, z_2)	z_2^*	n_2	P_2^*
.		.	.	.
.		.	.	.
.		.	.	.
k	$(z_{k-1}, z_{\max} = z_k)$	z_k^*	n_k	P_k^*
Σ			n	1

По данным табл. 3.2. строят гистограмму эмпирического распределения, представляющую собой ступенчатую кривую, значение которой на j -ом интервале (z_{j-1}, z_j) ($j = 1, 2, \dots, k$) ($z_j = z_{j-1} + h$) постоянно и равно n_j или P_j^* , или с учетом условия $\int_{-\infty}^{\infty} f(z) dz = 1$ равно $\frac{P_j^*}{h}$ (рис. 3.13).

Кривую эмпирического распределения можно аппроксимировать различной теоретической кривой распределения – законом Пуассона, Вейбула, биномиальным, показательным, логарифмическим, нормальным и др. По виду гистограммы предполагают, что случайная величина Z распределена по тому или иному закону (выдвигают рабочую гипотезу).

Определение параметров теоретического распределения. Окончательное уточнение неизвестной функции распределения сводится к определению некоторых числовых параметров распределения. По выборке могут быть рассчитаны выборочные статистические характеристики (выборочное среднее, дисперсия и т.д.), которые являются оценками соответствующих генеральных параметров. Оценки, получаемые по выборке, сами являются случайными величинами. К оценке обычно предъявляются требования состоятельности и несмещенности.

Для получения оценок используют различные методы. Широко применяется метод максимального правдоподобия. Оценки, получаемые при помощи этого метода, отвечают изложенным требованиям.

Сущность метода максимального правдоподобия заключается в нахождении таких оценок неизвестных параметров, для которых функция правдоподобия при случайной выборке объема n будет иметь максимальное значение. Пусть известен общий вид плотности вероятности $f(z, a)$ теоретического распределения; a – неизвестный параметр, входящий в выражение закона распределения. Экспериментально получена выборка значений случайной величины z_1, z_2, \dots, z_n .

Функция

$$L(z, \tilde{a}) = \sum_{i=1}^n \ln f(z_i, \tilde{a}) \quad (3.148)$$

называется функцией правдоподобия, где \tilde{a} – оценка параметра a .

В качестве оценки для a следует взять то значение \tilde{a} , для которого функция $L(z, \tilde{a})$ имеет наибольшее возможное значение. Необходимое условие экстремума этой функции дает уравнение

$$\frac{dL(z, \tilde{a})}{d\tilde{a}} = 0, \quad (3.149)$$

из которого определяют искомое значение \tilde{a} . Если максимумов несколько, необходимо выбрать среди них наибольший.

Достаточным условием максимума является выполнение неравенства

$$\frac{d^2L(z, \tilde{a})}{d\tilde{a}^2} \leq 0. \quad (3.150)$$

В общем случае требуется оценить одновременно несколько параметров распределения. В этом случае формулировка принципа максимального правдоподобия сохраняется: надо найти такую совокупность допустимых значений параметров $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_m$ функции распределения $f(z, a_1, a_2, \dots, a_m)$, которая обращает функцию правдоподобия

$$L(z, \tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_m) = \sum_{i=1}^n \ln f(z_i, \tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_m) \quad (3.151)$$

в максимум. Необходимое условие экстремума L функции дает система уравнений

$$\frac{\partial L(z, \tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_m)}{\partial \tilde{a}_j} = 0, \quad j = \overline{1, m}, \quad (3.152)$$

а неотрицательная определенность матрицы

$$\left(\frac{\partial^2 L(z, \tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_m)}{\partial \tilde{a}_i \partial \tilde{a}_j} \right), \quad i, j = \overline{1, 2, \dots, m}, \quad (3.153)$$

в стационарной точке является достаточным условием того, чтобы этот локальный экстремум был максимумом функции правдоподобия.

Из системы уравнений (3.152) находят оценки \tilde{a}_j ($j = \overline{1, m}$).

Найдем методом максимального правдоподобия оценку для параметра λ показательного распределения с плотностью

$$f(z) = \lambda e^{-\lambda z} \quad (3.154)$$

по выборке z_1, z_2, \dots, z_n .

Для этого распределения функция правдоподобия имеет вид

$$L = \sum_{i=1}^n \ln(\tilde{\lambda} e^{-\tilde{\lambda} z_i}) = n \ln \tilde{\lambda} - \tilde{\lambda} \sum_{i=1}^n z_i. \quad (3.155)$$

$$\frac{\partial L}{\partial \tilde{\lambda}} = \frac{n}{\tilde{\lambda}} - \sum_{i=1}^n z_i = 0;$$

$$\tilde{\lambda} = \frac{n}{\sum_{i=1}^n z_i} = \frac{1}{\bar{z}}, \quad (3.156)$$

где \bar{z} – среднее выборки.

Пусть распределение случайной величины Z подчинено нормальному закону

$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z-m)^2}{2\sigma^2}\right], \quad (3.157)$$

где m, σ^2 – параметры распределения.

Функция правдоподобия

$$L(z, \tilde{m}, \tilde{\sigma}^2) = \sum_{i=1}^n \ln\left[\frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left(-\frac{(z_i - \tilde{m})^2}{2\tilde{\sigma}^2}\right)\right]. \quad (3.158)$$

Необходимые условия экстремума L :

$$\left. \begin{aligned} \frac{\partial L(z, \tilde{m}, \tilde{\sigma}^2)}{\partial \tilde{m}} &= 0; \\ \frac{\partial L(z, \tilde{m}, \tilde{\sigma}^2)}{\partial \tilde{\sigma}^2} &= 0. \end{aligned} \right\} \quad (3.159)$$

Опуская промежуточные математические операции, из системы (3.159) находим:

$$\tilde{m} = \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i, \quad (3.160)$$

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2. \quad (3.161)$$

Оценка $\tilde{\sigma}^2$ получается несколько смещенной. Для получения несмещенной оценки умножаем $\tilde{\sigma}^2$ на $n/(n-1)$. Несмещенная оценка дисперсии при этом имеет вид

$$\tilde{L}^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2. \quad (3.162)$$

Уменьшение знаменателя в (3.162) на единицу непосредственно связано с тем, что величина \bar{z} относительно которой берутся отклонения, сама зависит от элементов выборки. Каждая величина, зависящая от элементов выборки и входящая в формулу выборочной дисперсии, называется связью. Знаменатель выборочной дисперсии всегда равен разности между объемом выборки и числом связей ℓ , наложенных на эту выборку. Эта разность

$$f = n - \ell \quad (3.163)$$

называется числом степеней свободы выборки.

В практических вычислениях для дисперсии $\tilde{\sigma}^2$ удобна формула

$$\tilde{\sigma}^2 = \frac{1}{n-1} \left[\sum_{i=1}^n (z_i)^2 - \frac{1}{n} \left(\sum_{i=1}^n (z_i) \right)^2 \right], \quad (3.164)$$

легко вытекающая из формулы (3.162). Преимущество формулы (3.164) в том, что в ней нет операций вычитания близких чисел, как в формуле (3.162), что приводит к потере точности.

Как правило, число наблюдений n велико. Поэтому для упрощения расчетов удобно использовать не соотношения (3.160) и (3.162), а менее трудоемкие формулы, позволяющие оценить m и σ^2 по сгруппированным данным:

$$\tilde{m} = \bar{z} = \frac{1}{n} \sum_{j=1}^k z_j^* n_j = \sum_{j=1}^k z_j^* P_j^*, \quad (3.165)$$

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{j=1}^k z_j^* n_j - \bar{z}^2 = \sum_{j=1}^k z_j^* P_j^* - \bar{z}^2. \quad (3.166)$$

Проверка адекватности теоретического распределения. Проверку рабочей гипотезы о законе распределения случайной величины по выборке проводят при помощи критериев согласия. Критерием согласия называется критерий гипотезы о том, что генеральная совокупность имеет распределение предполагаемого типа (например, нормальное распределение). Критерии согласия позволяют определить вероятность того, что при гипотетическом законе распределения наблюдающееся в рассматриваемой выборке отклонение вызывается случайными причинами, а не ошибкой в гипотезе. Если эта вероятность ве-

лика, то отклонение от гипотетического закона распределения следует признать случайным и считать, что гипотеза о предполагаемом законе распределения не опровергается.

Вероятностный характер критериев не позволяет однозначно принять или опровергнуть проверяемую гипотезу. Критерий позволяет утверждать, что гипотеза не противоречит опытным данным, если вероятность наблюдаемого отклонения от гипотетического закона велика, или что гипотеза не согласуется с опытными данными, если эта вероятность мала.

Среди различных критериев согласия наиболее употребителен универсальный критерий согласия χ^2 (Пирсона).

Проверку гипотезы о виде функции распределения с помощью этого критерия производят следующим образом:

1) По выборке строят гистограмму. Если в каком-либо j -м интервале число наблюдений n_j окажется меньше пяти, то его объединяют с соседним интервалом (или интервалами) так, чтобы число наблюдений в таком объединенном интервале оказалось большим или равном пяти (должно быть $n_j \geq 5 \forall j$).

Пусть K_0 – окончательное число интервалов, группированное после объединения, тогда очевидно, что

$$\sum_{j=1}^{K_0} n_j = n, \quad n_j \geq 5. \quad (3.167)$$

2) Задаются видом гипотетической функции распределения и для каждого из параметров m этого распределения находят оценки, причем эти оценки можно определить как по исходным, так и по сгруппированным данным.

3) Определяют теоретическую вероятность $P_j^m (j = \overline{1, k})$ в каждый из K_0 интервалов случайной величины с заданным распределением, параметры которого или известны, или оценены в п.2. P_j^m вычисляют по формуле

$$P_j^m = \int_{z_j^* - h/2}^{z_j^* + h/2} f(z) dz = P \left\{ z_j^* - \frac{h}{2} < z \leq z_j^* + \frac{h}{2} \right\}; \quad (3.168)$$

$$n_j^m = n P_j^m, \quad (3.169)$$

где $f(z)$ – выбранный теоретический закон распределения;

n_j^m – абсолютная частота попадания величины Z в j -й интервал по теоретическому распределению.

В случае нормального закона распределения

$$P_j^m = P \left\{ z_j^* - \frac{h}{2} < z \leq z_j^* + \frac{h}{2} \right\} = \frac{1}{\tilde{\sigma}\sqrt{2\pi}} \int_{z_j^* - h/2}^{z_j^* + h/2} \exp \left[-\frac{(z - \tilde{m})^2}{2\tilde{\sigma}^2} \right] dz. \quad (3.170)$$

Этот интеграл можно вычислить с помощью таблицы значений функции Лапласа $\Phi(x)$:

$$P_j^m = \Phi \left[\frac{\left(z_j^* + \frac{h}{2} \right) - \tilde{m}}{\tilde{\sigma}} \right] - \Phi \left[\frac{\left(z_j^* - \frac{h}{2} \right) - \tilde{m}}{\tilde{\sigma}} \right] = \Phi(\alpha) - \Phi(\beta). \quad (3.171)$$

Функция Лапласа $\Phi(x)$ – нечетная функция, т.е. $\Phi(-x) = -\Phi(x)$, поэтому таблица значений $\Phi(x)$ составлена лишь для $x \geq 0$.

4) Вычисляют значение χ^2 критерия Пирсона по формуле

$$\chi^2 = n \sum_{j=1}^{K_0} \frac{(P_j^* - P_j^m)^2}{P_j^m} = \sum_{j=1}^{K_0} \frac{(n_j - nP_j^m)^2}{nP_j^m}. \quad (3.172)$$

Известно, что для данного критерия согласия случайная величина χ^2 при больших n имеет χ^2 – распределение (при $n_j P_j^m \geq 5 \forall j$) с числом степеней свободы

$$f = K_0 - r - 1, \quad (3.173)$$

где r – число определяемых в п.2. неизвестных параметров гипотетического распределения, а уменьшение числа степеней свободы еще на единицу объясняется наличием линейного соотношения (3.167) между эмпирическими величинами n_j и n , входящими в расчетную формулу (3.172).

5) Задавшись уровнем значимости q (или доверительной вероятностью P , $q = 1 - P$) по таблице χ^2 -распределения Пирсона находят критическое значение $\chi_{кр}^2$ при заданном числе степеней свободы f .

6) Сравнивают значения χ^2 и $\chi_{кр}^2$ и выносят решение о принятии (если $\chi^2 - \chi_{кр}^2$) или отклонении (если $\chi^2 > \chi_{кр}^2$) рассматриваемой гипотезы о виде функции распределения. Если гипотеза не согласуется с выбранным эмпирическим распределением, выдвигается новая рабочая гипотеза с последующей ее проверкой на адекватность.

После того, как определен закон распределения случайной величины, можно говорить о том, что имеется модель изменения этой величины.

Контрольные вопросы

1. Опишите постановку задачи идентификации объектов.
2. Опишите идентификацию статики линейного детерминированного объекта. Основные подходы, методы, способы решения.
3. Опишите идентификацию статики нелинейного детерминированного объекта. Основные подходы, методы, способы решения.
4. Опишите идентификацию объекта при наличии случайных помех. Основные подходы, методы, способы решения.
5. Опишите построение регрессионной модели стохастического объекта. Основные подходы, методы, способы решения.
6. Опишите идентификацию динамических объектов. Основные подходы, методы, способы решения.
7. Опишите идентификацию закона распределения случайной величины. Основные подходы, методы, способы решения.

4. ПОСТРОЕНИЕ МОДЕЛЕЙ СТАТИКИ СЛОЖНЫХ МНОГОМЕРНЫХ ОБЪЕКТОВ

4.1. Выбор существенных переменных модели объекта моделирования

Для определения структуры модели сложного объекта необходимо выяснить, какие входы объекта будут включены в его модель. Для этого выявляют все возможные входы и из них выделяют наиболее существенные. Существенным является тот вход, который наибольшим образом оказывает влияние на выход объекта.

Для этого следует выполнить ранжирование выбираемых входов (факторов). В этом случае выбор наиболее существенных факторов очевиден – они должны быть расположены в начале ранжированного ряда.

Поэтому первым этапом задачи идентификации объекта следует считать задачу ранжирования его входов. Так как модели объекта еще нет, то ранжирование можно сделать, например, методом экспертных оценок.

Метод экспертных оценок. Для этого определяются все входы объекта. При их выборе принимают во внимание следующее. Входы должны непосредственно воздействовать на объект, не зависеть от других переменных, измеряться и управляться. Исходя из этих соображений, составляется ряд входов. Широкому кругу специалистов предлагается расположить эти факторы в порядке убывания степени их влияния на выбранную выходную величину объекта, т.е. проранжировать факторы.

При обработке результатов экспертного опроса весьма эффективным является метод ранговой корреляции. Сущность этого метода заключается в следующем.

Пусть m экспертов ранжируют n факторов. Каждому фактору каждый эксперт присваивает ранг – целое число от 1 до n , приписывая ранг 1 наиболее важному фактору, ранг 2 – следующему по важности и т.д. На основании этих данных составляется матрица рангов

$$A = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix}, \quad (4.1)$$

где a_{ij} – ранг i -го фактора ($i = \overline{1, n}$), определенного j -м экспертом ($j = \overline{1, m}$), номера строк соответствуют номерам экспертов, а номера столбцов – номерам ранжируемых факторов. Это означает, что j -я строка представляет собой мнение j -го эксперта, а i -й столбец – мнение всех экспертов по поводу i -го фактора.

При назначении рангов экспертами нужно соблюдать следующие условия:

1. Сумма рангов, назначенных всем факторам, должны быть одинакова для каждого эксперта и равна

$$\sum_{i=1}^n i = \frac{n \cdot (n+1)}{2}.$$

Это означает, что сумма элементов любой строки матрицы (4.1) должна быть равна

$$\sum_{i=1}^n a_{ij} = \frac{n \cdot (n+1)}{2}, \quad (j = \overline{1, m}).$$

2. Если эксперт какие-то q факторов считает одинаковыми, то приписывает им один ранг. Этот ранг равен среднему из q целых рангов, которые получены при условии, что эксперту удалось их проранжировать.

Например, равноценность двух факторов x_3, x_8 ($q = 2$), стоящих на шестом месте в ранжированном ряду, приводит к тому, что их ранги равны: $(6+7)/2=6.5$. Как видно, в этом случае ранги могут быть дробными. Отметим, что дробные ранги кратны $1/2$.

На базе данных матрицы (4.1) вычисляем:

$$\sum_i i = \sum_{j=1}^m a_{ij}, \quad i = \overline{1, n}; \quad (4.2)$$

$$\bar{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\sum_i i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m a_{ij} = \frac{1}{2} m(n+1); \quad (4.3)$$

$$S = \sum_{i=1}^n \left[\sum_{j=1}^m a_{ij} - \frac{1}{n} m(n+1) \right]^2 = \sum_{i=1}^n (\sum_i - \bar{\Sigma})^2, \quad (4.4)$$

где \sum_i – сумма рангов i -го фактора ($i = \overline{1, n}$);

$\bar{\Sigma}$ – среднее значение всех рангов;

S – сумма квадратов отклонений частных рангов от среднего.

Очевидно, максимальное значение S принимает тогда, когда эксперты одинаково проранжируют все факторы и отклонение суммарных рангов от среднего значения будет $m(n-1)/2, m(n-3)/2, \dots$. Сумма квадратов членов этого ряда определяется выражением

$$S_m = \frac{1}{12} m^2 n (n^2 - 1). \quad (4.5)$$

Степень согласованности мнений экспертов характеризует коэффициент конкордации (согласованности) W

$$W = S / S_m. \quad (4.6)$$

Коэффициент конкордации меняется от нуля (отсутствие какого-либо согласия в мнениях экспертов) до единицы (полное согласие). Статистическую значимость коэффициента W определяют в зависимости от числа факторов. При $n > 7$ используют χ^2 – критерий Пирсона:

$$\chi_P^2 = m(n-1)W. \quad (4.7)$$

Расчетные значение χ^2 сравнивают с табличным при выбранном уровне значимости α и числе степеней свободы $f = n - 1$. Гипотеза о наличии согласия между специалистами принимается в том случае если $\chi_P^2 \geq \chi_T^2 \cdot \chi_T^2$ определяется из таблицы распределения Пирсона.

В случае $n > 7$ пользуются F -критерием Фишера, расчетное значение которого определяют по формуле

$$F_P = \frac{0.5 \cdot \ln(m-1)W}{1-W} \quad (4.8)$$

с числом степеней свободы $f_1 = n - 1 - \frac{2}{m}$ и $f_2 = \frac{m-1}{f_1}$.

Мнение специалистов считается согласованным, если расчетное значение критерия Фишера больше или равно табличному $(F_p \geq F_T)F_T$ находится из таблицы распределения Фишера.

Далее оценивают степень влияния каждого фактора на выбранную выходную величину. Для этого используют ряд показателей, но наиболее часто - сумму рангов данного фактора. Чем меньше эта величина, тем сильнее влияет фактор на выходную величину.

Расположим теперь входы в порядке возрастания суммы их рангов: $x_{l_1}, x_{l_2}, \dots, x_{l_n}$, где индекс l_i равен номеру фактора с соответствующей суммой рангов (x_{l_1} – фактор с минимальной суммой рангов). Здесь на первом месте стоит самый существенный вход (его номер l_1), а далее следуют остальные в порядке уменьшения их влияния на выход.

Для анализа результатов ранжирования используют диаграмму рангов (рис. 4.1). Наиболее благоприятным видом диаграммы является возрастание рангов по параболе. В этом случае легко разделить факторы на группы и по К-критерию Линка-Уэллеса отсеять несущественные.

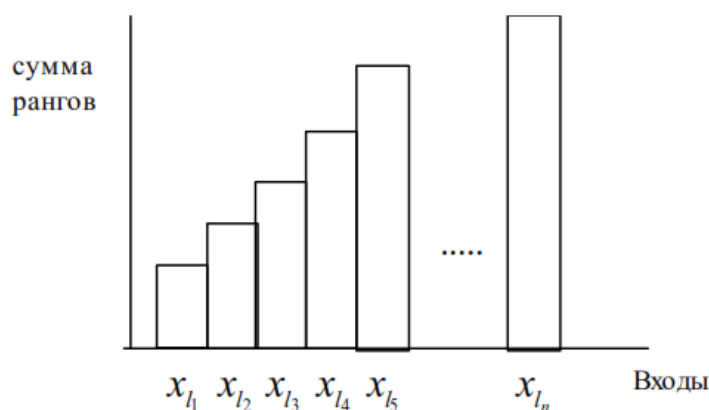


Рис. 4.1

Расчетная величина К-критерия определяется по формуле

$$K_p = \frac{k(\bar{a}_{\max} - \bar{a}_{\min})}{\sum_{i=1}^k a_{l_i}}, \quad (4.9)$$

где K – число факторов, включаемых в группу;

$\bar{a}, \bar{a}_{\max}, \bar{a}_{\min}$ – среднее значение суммы рангов фактора (средние ранги), максимальное и минимальное значения \bar{a}_i в группе;

$$\bar{a}_i = m \sum_{l_i}.$$

Если значение K_p меньше табличного (табл. 4.1) при выбранном уровне значимости α , числе сравниваемых средних K и числа специалистов m , то сравниваем средние ранги a_{i_j} ($i = \overline{1, k}$) не различаются между собой.

Можно рекомендовать следующий порядок анализа результатов ранжирования. Ориентируясь на вид диаграммы рангов факторы, разбивают на отдельные группы. Вычисляют K_p для первой (наиболее значимой) группы факторов. С помощью K -критерия проверяют, различаются ли между собой средние ранги факторов, включенных в эту группу, т.е. образуют ли они в действительности единую группу. Если это подтверждается, тогда объединяют факторы первой и второй (соседней) групп и с помощью K -критерия проверяют эти две группы как единую. Если эти две группы не образуют единую группу, то в первую группу включают не все факторы второй группы, а только один (или несколько) соседний с первой группой фактор и т.д.

Таблица 4.1

Критические значения K -критерия при $\alpha = 0.05$

Число экспертов m	Число факторов в группе K										
	2	3	4	5	6	7	8	9	10	11	12
2	3.43	2.35	1.74	1.39	1.15	0.39	0.87	0.77	0.70	0.63	0.58
3	1.90	1.44	1.14	0.94	0.80	0.70	0.62	0.56	0.51	0.47	0.43
4	1.62	1.25	1.01	0.84	0.72	0.63	0.57	0.51	0.47	0.43	0.40
5	1.53	1.19	0.96	0.81	0.70	0.61	0.55	0.50	0.45	0.42	0.39
6	1.50	1.17	0.95	0.80	0.69	0.61	0.55	0.49	0.45	0.42	0.39
7	1.49	1.17	0.95	0.80	0.69	0.61	0.55	0.50	0.45	0.42	0.39
8	1.49	1.18	0.96	0.81	0.70	0.62	0.55	0.50	0.46	0.42	0.39
9	1.50	1.19	0.97	0.82	0.71	0.62	0.56	0.51	0.47	0.43	0.40
10	1.52	1.20	0.98	0.83	0.72	0.63	0.57	0.52	0.47	0.44	0.41
11	1.54	1.22	0.99	0.84	0.73	0.64	0.58	0.52	0.48	0.44	0.41
12	1.56	1.23	1.01	0.85	0.74	0.65	0.58	0.53	0.49	0.45	0.42
13	1.58	1.25	1.02	0.86	0.75	0.66	0.59	0.54	0.49	0.46	0.42
14	1.60	1.26	1.03	0.87	0.76	0.67	0.60	0.55	0.50	0.46	0.43
15	1.62	1.28	1.05	0.89	0.77	0.68	0.61	0.55	0.51	0.47	0.44
20	1.72	1.36	1.12	0.95	0.82	0.73	0.65	0.59	0.54	0.50	0.47
50	2.23	1.77	1.45	1.22	1.06	0.94	0.85	0.77	0.71	0.65	0.61

Факторы, образующие единую группу, включаются в программу исследований. Остальные факторы могут быть отброшены, как не оказывающие существенного влияния на выходную величину объекта.

Метод ранжирования факторов удобно применять при разработке модели объекта, которого еще не существует, т. е. на стадии создания. Если же имеют-

ся предварительные экспериментальные данные, полученные на объекте, то отсеивающий эксперимент можно выполнить методом корреляционного анализа.

Корреляционный анализ. Этот метод используют для установления статистической связи между переменными объекта. Суть его заключается в определении коэффициентов парной корреляции $r_{x_i y}$ между каждым входом x_i и выходом y объекта по имеющимся экспериментальным данным.

Коэффициент парной корреляции является мерой тесноты линейной статистической связи между двумя случайными величинами. В общем случае его величина изменяется от 0 до ± 1 . Если $r_{x_i y} = 0$, то связь между x_i и y отсутствует или она нелинейна. Если $|r_{x_i y}| = 1$, то связь является линейной, функциональной. При нелинейной зависимости между случайными величинами для анализа меры тесноты их связи используют индекс корреляции.

В процессе отсеивающего эксперимента из рассмотрения исключаются факторы, у которых с выбранной выходной величиной объекта коэффициент корреляции близок к нулю.

Коэффициент парной корреляции между входом x_i и выходом y объекта определяется по формуле

$$r_{x_i y} = \frac{\sum_{i=1}^N (x_{ji} - \bar{x}_j)(y_i - \bar{y})}{(N-1)\sigma_{x_j}\sigma_y}, \quad j = \overline{1, n}, \quad (4.10)$$

где N – число измерений;

i – номер опыта,

x_{ji} – значение j -го входа в i -м опыте;

$$\sigma_{x_j} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_{ji} - \bar{x}_j)^2}; \quad \sigma_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2};$$

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ji}; \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i; \quad j = \overline{1, n}.$$

При наличии высокой корреляции у двух входов x_j и x_l

$$r_{x_j x_l} = \frac{\sum_{i=1}^N (x_{ji} - \bar{x}_j)(x_{li} - \bar{x}_l)}{(N-1)\sigma_{x_j}\sigma_{x_l}}, \quad j, l = \overline{1, n}, \quad j \neq l, \quad (4.11)$$

любой из них также можно исключить из рассмотрения, так как они однозначно влияют на выходную величину объекта. Исключается, как правило, тот фактор, который труднее определить экспериментально.

После расчета коэффициентов парной корреляции необходимо установить их статистическую значимость, т.е. проверить гипотезу об отличии вычисленного значения коэффициента от нуля. Для этого из табл. 4.2 при выбранном уровне значимости α и числе степеней свободы $f = N - 2$ находят критическое значение коэффициента корреляции r_K . Линейная связь считается статистически значимой, если $|r_p| \geq r_K$.

Таблица 4.2

Критические значения коэффициента корреляции r_K при $\alpha = 0.05$

f	r_K	f	r_K	f	r_K	f	r_K
1	0.997	9	0.602	17	0.456	45	0.287
2	0.950	10	0.576	18	0.444	50	0.273
3	0.878	11	0.553	19	0.433	60	0.250
4	0.811	12	0.532	20	0.423	70	0.232
5	0.754	13	0.514	25	0.381	80	0.217
6	0.707	14	0.497	30	0.349	90	0.205
7	0.666	15	0.482	35	0.325	100	0.195
8	0.632	16	0.468	40	0.304		

4.2. Оптимальное планирование эксперимента

Главная цель планирования эксперимента заключается в достижении его максимальной точности и минимальной стоимости. Однако достижение максимальной точности требует увеличения числа опытов, что приводит к повышению стоимости эксперимента. Поэтому возникает задача оптимизации в постановке эксперимента.

Методически опыты могут быть поставлены двояко. Традиционный метод постановки опытов состоит в изменении одного какого-либо фактора при сохранении всех других факторов, влияющих на процесс, постоянными. Этот метод постановки опытов известен под названием **метода однофакторного эксперимента**. При такой методике взаимное влияние факторов учесть невозможно.

Оптимальное планирование эксперимента предполагает одновременное изменение всех факторов, влияющих на процесс, что позволяет сразу установить силу взаимодействия факторов и сократить общее число опытов. Такой метод постановки опытов называется **методом многофакторного планирования эксперимента**. В результате исследований получают не сечения статических характеристик объекта, как это имеет место в классическом однофакторном методе, а функциональную зависимость выхода объекта от всех факторов $y = f(x_1, x_2, \dots, x_k)$, что очень существенно.

Теория планирования эксперимента и ее приложения в настоящее время развились в самостоятельное научное направление. Попытаемся здесь весьма кратко изложить сущность метода планированного факторного эксперимента применительно к построению модели статистики объекта исследования.

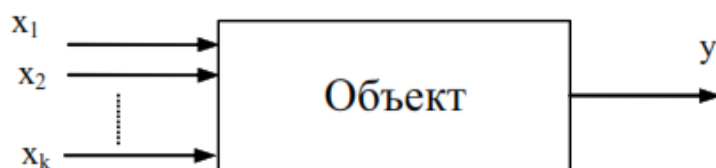


Рис. 4.2

Сущность метода заключается в следующем. Объект исследования рассматривается как «черный ящик» (рис. 4.2). Его входы называют **факторами**, а выходы — **откликами**. Целью исследования является построение модели статистики объекта, то есть функцию отклика

$$y = f(x_1, x_2, \dots, x_k). \quad (4.12)$$

Координатное пространство с координатами x_1, x_2, \dots, x_k называют **факторным пространством**, а геометрическое изображение функции отклика в факторном пространстве — **поверхностью отклика**.

В процессе эксперимента каждый фактор может принимать одно из нескольких фиксированных значений. Эти значения называются **уровнями**. Сочетание определенных уровней всех факторов определяет одно из возможных состояний объекта. Множество возможных сочетаний уровней факторов определяет множество состояний данного объекта и, следовательно, число возможных различных опытов, равное

$$N = p^k, \quad (4.13)$$

где p – число уровней факторов;

k – число факторов.

Для уменьшения размерности факторного пространства и упрощения модели объекта уменьшают число факторов путем отсеивания малосущественных.

Факторы должны быть управляемыми, независимыми и совместимыми. Это означает, что факторы не должны быть функциями других факторов, должна существовать возможность установления фактора на выбранных уровнях независимо от уровня других факторов, а все комбинации уровней факторов должны давать осуществимые и безопасные для объекта режимы.

Перед началом эксперимента необходимо выбрать его план, то есть определить какие сочетания уровней факторов следует реализовать и в каком порядке. Это очень важный этап в рассматриваемом методе, поскольку этим, по сути, ограничивается класс регрессионных моделей, среди которых отыскивается модель

объекта. Следовательно, требуется задать общий вид отыскиваемой модели: линейная, нелинейная с эффектами взаимодействия, квадратичная и т.п.

Математическая модель чаще всего представляется в виде полинома — отрезка ряда Тейлора, в который разлагается неизвестная функция (4.12):

$$y = \alpha_0 + \sum_{j=1}^k \alpha_j x_j + \sum_{u,j=1}^k \alpha_{u,j} x_u x_j + \sum_{j=1}^k \alpha_{jj} x_j^2 + \dots, \quad (4.14)$$

где $\alpha_j = \left. \frac{\partial F}{\partial x_j} \right|_0$; $\alpha_{uj} = \left. \frac{\partial^2 F}{\partial x_u \partial x_j} \right|_0$; $\alpha_{jj} = \left. \frac{\partial^2 F}{\partial x_j^2} \right|_0$; $\alpha_0 = F(\bar{x}^0)$; индекс 0 означает, что

производные вычисляются в точке $x^r = x^{r0}$.

Поскольку в реальном объекте всегда существуют неуправляемые и неконтролируемые переменные, изменение величины y носит случайный характер. Поэтому при обработке экспериментальных данных получают так называемые выборочные коэффициенты регрессии $a_0, a_j, a_{uj}, a_{jj}, \dots$, являющиеся оценками теоретических коэффициентов $\alpha_0, \alpha_j, \alpha_{uj}, \alpha_{jj}, \dots$. Уравнение регрессии, полученное на основании опыта, запишется следующим образом:

$$y = a_0 + \sum_{j=1}^k a_j x_j + \sum_{u,j=1}^k a_{u,j} x_u x_j + \sum_{j=1}^k a_{jj} x_j^2 + \dots \quad (4.15)$$

Коэффициент a_0 называют свободным членом уравнения регрессии, коэффициенты a_j — линейными эффектами, a_{uj} — эффектами парного взаимодействия, a_{jj} — квадратичными эффектами.

4.2.1. Нормирование переменных модели

В процессе эксперимента осуществляются различные комбинации уровней факторов. Эти факторы, как правило, имеют различную физическую природу и размерность. Для упрощения записи и обработки результатов уровни факторов нормируются.

Середине области определения выходной переменной объекта соответствует некоторая комбинация уровней факторов. Эти уровни факторов принимаются как основные (исходные, нулевые, начальные).

Построение плана эксперимента сводится к выбору интервала варьирования факторами Δx_j , то есть к выбору экспериментальных точек, симметричных относительно основного уровня.

Чаще всего эксперименты проводятся только на двух уровнях факторов. Такие эксперименты называются экспериментами типа 2^k . При таком эксперименте уровни факторов представляют собой границы исследуемой области по данному фактору $(x_{j \min}, x_{j \max})$. Тогда для любого фактора имеем

$$x_j^0 = \frac{x_{j \min} + x_{j \max}}{2}, \quad j = \overline{1, k}; \quad (4.16)$$

$$\Delta x_j = \frac{x_{j \max} - x_{j \min}}{2}, \quad j = \overline{1, k}. \quad (4.17)$$

Точка с координатами $(x_1^0, x_2^0, \dots, x_k^0)$ называется *центром плана* (основным уровнем); Δx_j – интервал варьирования по оси x_j . Величина Δx_j своя для каждого фактора. Прибавление Δx_j к основному уровню дает верхний, а вычитание – нижний уровень фактора.

От системы координат x_1, x_2, \dots, x_k перейдем к новой безразмерной системе координат $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k$ путем следующего линейного преобразования координат

$$\hat{x}_j = \frac{x_j - x_j^0}{\Delta x_j}, \quad j = \overline{1, k}, \quad (4.18)$$

где \hat{x}_j – нормированное значение j -го фактора; x_j – натуральное значение j -го фактора $x_j = x_j^0 \pm \Delta x_j$; x_j^0 – натуральное значение основного уровня.

В нормированном виде для всех факторов верхний уровень равен +1, нижний уровень равен -1 ($-1 \leq \hat{x}_j \leq 1$), координаты центра плана равны нулю и совпадают с началом координат.

Нормирование переменных существенно упрощает планирование эксперимента, обработку его результатов, построение модели объекта с безразмерными переменными, упрощение модели и проверку ее адекватности. После построения модели с нормированными факторами $y = F(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k)$ и оценки ее адекватности от этой модели переходят к модели с натуральными факторами $y = F(x_1, x_2, \dots, x_k)$, используя нормирующие соотношения (4.18).

4.2.2. Полный факторный эксперимент

При планировании по схеме полного факторного эксперимента (ПФЭ) реализуются все возможные сочетания уровней факторов. Необходимое количество опытов N при ПФЭ определяется по формуле (4.13).

Если эксперименты проводятся только на двух уровнях факторов и при этом в процессе эксперимента осуществляется все возможные комбинации из k факторов, то постановка опытов по такому плану называется ПФЭ типа 2^k .

ПФЭ типа 2^k широко используется благодаря таким их положительным особенностям, как симметричность относительно центра эксперимента, независимость численных значений коэффициентов уравнения регрессии, полученного экспериментально, от изменения порядка регрессионного полинома, принятого для описания объекта (ортогональность), одинаковая точность модели на равных расстояниях от центра в различных направлениях (ротатабельность). ПФЭ обычно используется при небольших числах факторов ($k \leq 5$).

С помощью ПФЭ типа 2^k может быть получено адекватная модель объекта только в том случае, если в нее не входят степени факторов, то есть этот эксперимент позволяет определить оценки коэффициентов уравнения регрессии только при линейных членах и взаимодействиях:

$$y = a_0 + \sum_{j=1}^k a_j x_j + \sum_{\substack{u,j=1 \\ u < j}}^k a_{u,j} x_u x_j + \sum_{\substack{u,j,l=1 \\ (u < j < l)}}^k a_{jll} x_u x_j x_l + \dots \quad (4.19)$$

Взаимодействие факторов имеет место тогда, когда эффект одного фактора зависит от уровня, на котором находится другой фактор, то есть статическая характеристика объекта нелинейна.

Условия эксперимента описывают с помощью специальной таблицы — матрицы планирования, в которой каждая строка соответствует условиям одного опыта. Для проведения опытов исходные данные по всем факторам желательно свести в таблицу (табл. 4.3). Рассмотрим план ПФЭ 2^2 . Число опытов $N = 4$.

Матрица плана ПФЭ типа 2^2 приведена в табл. 4.4, в которой введен столбец фиктивной переменной $\hat{x}_0 = +1$ для вычисления свободного члена a_0 . Для сокращения записи обычно опускают единицы в матрице планирования. В первом опыте оба фактора находятся на верхнем уровне, во втором фактор x_1 — на нижнем, x_2 — на верхнем и т.д.

В случае $k > 2$ при построении матрицы плана ПФЭ типа 2^k руководствуются следующим. При любом числе факторов k необходимо дважды повторить матрицу планирования для случая $k-1$ сначала при значении k -го фактора на верхнем уровне, а затем на нижнем. В табл. 4.5 показано последовательное достраивание матрицы ПФЭ при увеличении k от 2 до 4. Для простоты записи плана единицы опущены, а оставлены только знаки.

Таблица 4.3
Исходные данные

Значения факторов	X ₁	X ₂	X _k
x _j ⁰				
Δx _j				
x _{j min}				
x _{j max}				

Таблица 4.4
Матрица плана ПФЭ 2²

Номер опыта	\hat{x}_0	\hat{x}_1	\hat{x}_2
1	+1	+1	+1
2	+1	-1	+1
3	+1	+1	-1
4	+1	-1	-1

Матрица плана ПФЭ типа 2^k обладает следующими свойствами:

$$\sum_{i=1}^N \hat{x}_{ui} \hat{x}_{ji} = 0, u \neq j, u, j = \overline{0, k}; \quad (4.20)$$

$$\sum_{i=1}^N \hat{x}_{ji} = 0, j = \overline{1, k}, j \neq 0; \quad (4.21)$$

$$\sum_{i=1}^N \hat{x}_{ji}^2 = N, j = \overline{0, k}, \quad (4.22)$$

где \hat{x}_{ji} – нормированное значение j -го фактора в i -ом опыте.

Таблица 4.5

Матрица плана ПФЭ 2², 2³, 2⁴

План		№ опыта	\hat{x}_0	\hat{x}_1	\hat{x}_2	\hat{x}_3	\hat{x}_4
2 ²	2 ²	1	+	+	+	+	+
		2	+	—	+	+	+
		3	+	+	—	+	+
		4	+	—	—	+	+
2 ³	2 ³	5	+	+	+	—	+
		6	+	—	+	—	+
		7	+	+	—	—	+
		8	+	—	—	—	+
2 ⁴	2 ⁴	9	+	+	+	+	—
		10	+	—	+	+	—
		11	+	+	—	+	—
		12	+	—	—	+	—
		13	+	+	+	—	—
		14	+	—	+	—	—
		15	+	+	—	—	—
		16	+	—	—	—	—

Свойство (4.20) называется свойством ортогональности матрицы планирования. Это свойство резко уменьшает трудности, связанные с расчетом коэффициентов модели.

4.2.3. Вычисление коэффициентов модели

Приведем метод вычисления коэффициентов линейной модели с нормированными факторами

$$y = a_0 + \sum_{j=1}^k a_j \hat{x}_j \quad (4.23)$$

по результатам ПФЭ типа 2^k .

Используем матричный метод регрессионного анализа, который очень удобен для решения задач на ЭВМ. Для этого введем следующие матрицы: матрицу независимых переменных

$$X = \begin{pmatrix} \hat{x}_{01} & \hat{x}_{11} & \hat{x}_{21} & \dots & \hat{x}_{k1} \\ \hat{x}_{02} & \hat{x}_{12} & \hat{x}_{22} & \dots & \hat{x}_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ \hat{x}_{0N} & \hat{x}_{1N} & \hat{x}_{2N} & \dots & \hat{x}_{kN} \end{pmatrix}, \quad (4.24)$$

матрицу – столбец откликов (вектор наблюдений) и матрицу – столбец коэффициентов

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}, \quad (4.25)$$

и матрицу, транспонированную к X

$$X^T = \begin{pmatrix} \hat{x}_{01} & \hat{x}_{02} & \hat{x}_{03} & \dots & \hat{x}_{0N} \\ \hat{x}_{11} & \hat{x}_{12} & \hat{x}_{13} & \dots & \hat{x}_{1N} \\ \dots & \dots & \dots & \dots & \dots \\ \hat{x}_{k1} & \hat{x}_{k2} & \hat{x}_{k3} & \dots & \hat{x}_{kN} \end{pmatrix}. \quad (4.26)$$

В матричной форме система нормированных уравнений для определения коэффициентов a_j ($j = \overline{0, k}$) методом наименьших квадратов запишется следующим образом:

$$X^T X A = X^T Y. \quad (4.27)$$

Отсюда

$$A = (X^T X)^{-1} X^T Y \quad (4.28)$$

Матрица моментов (информационная матрица Фишера) $X^T X$ имеет вид

$$X^T X = \begin{pmatrix} \sum \hat{x}_{0i}^2 & \sum \hat{x}_{0i} \hat{x}_{1i} & \dots & \sum \hat{x}_{0i} \hat{x}_{ki} \\ \sum \hat{x}_{1i} \hat{x}_{0i} & \sum \hat{x}_{1i}^2 & \dots & \sum \hat{x}_{1i} \hat{x}_{ki} \\ \vdots & \vdots & \vdots & \vdots \\ \sum \hat{x}_{ki} \hat{x}_{0i} & \sum \hat{x}_{ki} \hat{x}_{1i} & \dots & \sum \hat{x}_{ki}^2 \end{pmatrix}, \quad (4.29)$$

где суммирование всюду осуществляется по i от 1 до N .

Учитывая свойства ортогональности (4.20) матрица коэффициентов нормальных уравнений $X^T X$ становится диагональной и ее диагональные элементы равны N

$$X^T X = \begin{pmatrix} N & 0 & 0 & \dots & 0 \\ 0 & N & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & N \end{pmatrix}. \quad (4.30)$$

Матрица, обратная матрице моментов $(X^T X)^{-1}$, получается равной

$$(X^T X)^{-1} = \begin{pmatrix} 1/N & 0 & 0 & \dots & 0 \\ 0 & 1/N & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1/N \end{pmatrix}. \quad (4.31)$$

Далее имеем:

$$X^T Y = \begin{pmatrix} \sum_{i=1}^N \hat{x}_{0i} y_i \\ \sum_{i=1}^N \hat{x}_{1i} y_i \\ \vdots \\ \sum_{i=1}^N \hat{x}_{ki} y_i \end{pmatrix}; \quad (4.32)$$

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} 1/N & 0 & 0 & \dots & 0 \\ 0 & 1/N & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1/N \end{pmatrix} \times \begin{pmatrix} \sum_{i=1}^N \hat{x}_{0i} y_i \\ \sum_{i=1}^N \hat{x}_{1i} y_i \\ \vdots \\ \sum_{i=1}^N \hat{x}_{ki} y_i \end{pmatrix} = \begin{pmatrix} \frac{1}{N} \sum_{i=1}^N \hat{x}_{0i} y_i \\ \frac{1}{N} \sum_{i=1}^N \hat{x}_{1i} y_i \\ \vdots \\ \frac{1}{N} \sum_{i=1}^N \hat{x}_{ki} y_i \end{pmatrix} \quad (4.33)$$

где суммирование осуществляется по i от 1 до N .

Следовательно, любой коэффициент a_j определяется по формуле

$$a_j = \frac{1}{N} \sum_{i=1}^N \hat{x}_{ji} y_i, \quad j = \overline{0, k}. \quad (4.34)$$

Если в рассмотрение ввести нелинейную модель с коэффициентами взаимодействия

$$y = a_0 + \sum_{j=1}^k a_j x_j + \sum_{\substack{u, j=1 \\ u < j}}^k a_{uj} x_u x_j + \sum_{\substack{u, j, l=1 \\ (u < j < l)}}^k a_{jil} x_u x_j x_l + \dots, \quad (4.35)$$

то эффекты взаимодействия определяются аналогично линейным эффектам:

$$a_{uj} = \frac{1}{N} \sum_{i=1}^N \hat{x}_{ui} \hat{x}_{ji} y_i, \quad u, j = \overline{1, k}, \quad u < j; \quad (4.36)$$

$$a_{ujl} = \frac{1}{N} \sum_{i=1}^N \hat{x}_{ui} \hat{x}_{ji} \hat{x}_{li} y_i, \quad u, j, l = \overline{1, k}, \quad u < j < l \quad (4.37)$$

и т.д.

В связи с тем, что матрица $(X^T X)^{-1}$ для спланированного эксперимента – матрица диагональная коэффициенты уравнения регрессии некоррелированы между собой. При этом выборочные коэффициенты a_j оказываются несмешанными оценками для соответствующих генеральных коэффициентов α_j ($a_j \rightarrow \alpha_j$), то есть величины коэффициентов уравнения регрессии характеризуют вклад каждого фактора в величину y .

Если выход объекта зависит еще и от квадрата факторов

$$y = a_0 + \sum_{j=1}^k a_j x_j + \dots + \sum_{j=1}^k a_{jj} x_j^2, \quad (4.38)$$

то эксперимент типа 2^k не позволяет определить квадратичные эффекты a_{jj} . Действительно, значения \hat{x}_{jj}^2 ($j = \overline{1, k}$) не отличаются друг от друга и от вектор столбца \hat{x}_0 . Поэтому величина a_0 представляет собой смешанную оценку значения свободного члена и вкладов квадратичных членов, то есть

$$a_0 \rightarrow \alpha_0 + \sum_{j=1}^k \alpha_{jj},$$

где a_0 – вычисленный выше коэффициент;

α_0, α_{jj} – неизвестные истинные значения a_0, a_{jj} .

Для получения модели типа (4.38), включающей степени факторов, необходимо проводить эксперимент при $p > 2$.

4.2.4. Оценка значимости коэффициентов модели

Для упрощения полученной модели проверяют значимость ее коэффициентов. Ее выполняют для каждого коэффициента отдельно. Оценка значимости коэффициентов производится по критерию Стьюдента, либо по доверительному интервалу, который при использовании факторного эксперимента одинаков для всех коэффициентов.

Критерий Стьюдента

$$t_j = \frac{|a_j|}{\sigma_{aj}}, \quad j = 0, 1, 2, \dots, \quad (4.39)$$

где a_j – j -й коэффициент уравнения регрессии;

σ_{aj} – среднее квадратичное отклонение коэффициента a_j .

Диагональные элементы ковариационной матрицы $(X^T X)^{-1}$ равны между собой, поэтому все коэффициенты уравнения регрессии определяются с одинаковой точностью:

$$\sigma_{aj} = +\sqrt{\sigma_{aj}^2} = +\sqrt{\frac{\sigma_{\text{воспр}}^2}{N}}, \quad (4.40)$$

где $\sigma_{\text{воспр}}^2$ – дисперсия воспроизводимости.

Дисперсия воспроизводимости, как правило, определяется с помощью повторных опытов в нулевой точке (центр эксперимента):

$$\sigma_{\text{воспр}}^2 = \frac{S_{\text{воспр}}}{f_{\text{воспр}}} = \frac{1}{m-1} \sum_{i=1}^m (y_i^0 - \bar{y}^0)^2, \quad (4.41)$$

где m – число повторных опытов;

$f_{\text{воспр}} = m - 1$ – число степеней свободы дисперсии воспроизводимости;

$S_{\text{воспр}}$ – остаточная сумма квадратов воспроизводимости;

$$\bar{y}^0 = \frac{1}{m} \sum_{i=1}^m y_i^0. \quad (4.42)$$

Таким образом, вычисляют t_i для каждого i -го коэффициента. Для t -критерия имеются специальные таблицы распределения Стьюдента, рассчитанные для различных $f_{\text{воспр}}$, m и доверительных вероятностей P (уровней значимости q).

Коэффициент a_j считается значимым (значимо отличается от нуля) с некоторой доверительной вероятностью, если вычисленный t_i больше табличного значения t_m при такой же доверительной вероятности и соответствующем числе степеней свободы.

Для оценки значимости коэффициентов по доверительному интервалу вычисляют доверительный интервал для j -го коэффициента по формуле

$$\Delta a_j = \pm t_m \sigma_{aj} = \pm t_m \sqrt{\frac{\sigma_{\text{воспр}}^2}{N}}. \quad (4.43)$$

Как было отмечено выше, доверительный интервал Δa_j одинаков для всех коэффициентов. Определение значимости коэффициентов уравнения регрессии с помощью соотношения (4.43) удобно тем, что позволяет применить правило: коэффициент значим, если его абсолютная величина больше доверительного интервала.

Незначимые коэффициенты исключаются из модели. При этом если коэффициенты модели некоррелированы между собой (матрица моментов диагональная), то исключение незначимых коэффициентов не скажется на остальных коэффициентах. В обратном случае оставшиеся коэффициенты пересчитываются заново, поскольку коэффициенты закоррелированы друг с другом.

4.2.5. Проверка адекватности модели

Полученную с помощью факторного планированного эксперимента модель объекта необходимо проверить на адекватность. Это осуществляется с помощью критерия Фишера (при наличии степеней свободы):

$$F = \frac{\sigma_1^2}{\sigma_2^2}, \quad (4.44)$$

где σ_1^2 , σ_2^2 – дисперсии двух независимых выборок.

При расчете F -критерия в числитель ставят большую из двух эмпирических дисперсий, поэтому $F \geq 1$. Критерий Фишера показывает, значимо ли отличаются друг от друга эти дисперсии.

F -критерий формируют следующим образом:

$$F = \begin{cases} \frac{\sigma_{\text{ЕЛБ}}^2}{\sigma_{\%0''-p}^2}, & \text{H}^0: \sigma_{\text{ЕЛБ}}^2 \geq \sigma_{\%0''-p}^2; \\ \frac{\sigma_{\%0''-p}^2}{\sigma_{\text{ЕЛБ}}^2}, & \text{H}^0: \sigma_{\text{ЕЛБ}}^2 < \sigma_{\%0''-p}^2, \end{cases} \quad (4.45)$$

где $\sigma_{\text{ад}}^2$ – дисперсия адекватности (остаточная дисперсия).

Дисперсия воспроизводимости вычисляется по формуле (4.41), а дисперсия адекватности по формуле

$$\sigma_{\text{ад}}^2 = \frac{1}{N-r} \sum_{i=1}^N (y_i - y_i^M)^2, \quad (4.46)$$

где r – число значимых коэффициентов модели; y_i, y_i^M – экспериментальные отклики и отклики, рассчитанные по модели при тех же значениях факторов.

Число степеней свободы для дисперсии адекватности

$$f_{\text{ад}} = N - r. \quad (4.47)$$

Для F -критерия имеются специальные таблицы распределения Фишера, рассчитанные для различных $f_{\text{ад}} = f_1$, $f_{\text{воспр}} = f_2$ и доверительных вероятностей P (уровней значимости q). Если

$$F \leq F_m(f_1, f_2, p), \quad (4.48)$$

то с соответствующей доверительной вероятностью P можно принять, что $\sigma_{\text{ад}}^2 = \sigma_{\text{воспр}}^2$ и считать модель в исследованной области адекватной объекту.

Таким образом, с помощью F -критерия проверяется лишь тот факт, что дисперсия адекватности не отличается значимо от дисперсии воспроизводимости, то есть ошибка модели и ошибка связанная с точностью эксперимента, на основе которого построена модель, близки друг другу. Этого достаточно для утверждения, что полученная модель адекватно описывает результаты эксперимента.

Если $F > F_m(f_1, f_2, p)$, то для адекватного описания эксперимента необходимо увеличить порядок модели (например, ввести в модель квадраты факторов). В этом случае эксперимент проводится при числе уровней факторов больше чем 2.

Для случаев, когда повторные опыты ставятся во всех (или в нескольких) точках плана одинаковое или неодинаковое число раз, формулы (4.41) и (4.46) несколько изменяются.

4.2.6. Дробный факторный эксперимент

С увеличением числа факторов количество опытов при ПФЭ растет по показательной функции. В то же время часто заранее известно, что число определяемых коэффициентов в модели объекта намного меньше числа опытов в соответствующем ПФЭ (например, если можно ограничиться линейным приближением модели). В этом случае используя дробный факторный эксперимент (ДФЭ) и получая дробные реплики можно резко сократить число опытов.

ДФЭ позволяет минимизировать число опытов, но при этом оценки коэффициентов получаются смешанными.

Эффективность ДФЭ при определении модели объекта зависит от выбранной системы смешивания. Необходимо выбрать такие дробные реплики, при которых интересующие экспериментатора линейные эффекты смешиваются с теми эффектами взаимодействия, о которых с наибольшей вероятностью можно предположить, что они близки к нулю. Как правило, такое предположение чаще всего справедливо относительно взаимодействий самого высокого порядка.

Для определения системы смешивания оценок используется так называемый определяющий контраст — символическое произведение всех столбцов матрицы планирования равное +1 или -1. Чтобы узнать, какие взаимодействия

смешаны с данным линейным эффектом (то есть, в каких столбцах плана имеются одинаковые элементы), необходимо умножить определяющий контраст на столбец, соответствующий данному эффекту.

Положим, что контраст в случае трех факторов равен

$$\hat{x}_1 \hat{x}_2 \hat{x}_3 = 1. \quad (4.49)$$

Умножим по очереди определяющий контраст на $\hat{x}_1 \hat{x}_2 \hat{x}_3$, получим

$$\left. \begin{aligned} \hat{x}_1 &= \hat{x}_1^2 \hat{x}_2 \hat{x}_3 = \hat{x}_2 \hat{x}_3 \\ \hat{x}_2 &= \hat{x}_2^2 \hat{x}_1 \hat{x}_3 = \hat{x}_1 \hat{x}_3 \\ \hat{x}_3 &= \hat{x}_3^2 \hat{x}_1 \hat{x}_2 = \hat{x}_1 \hat{x}_2 \end{aligned} \right\} \quad (4.50)$$

поскольку $\hat{x}_i^2 = 1$.

Полученным соотношениям (4.50) соответствует система смешанных оценок:

$$\left. \begin{aligned} a_1 &= \alpha_1 + \alpha_{23}; \\ a_2 &= \alpha_2 + \alpha_{13}; \\ a_3 &= \alpha_3 + \alpha_{12}. \end{aligned} \right\} \quad (4.51)$$

При увеличении числа факторов целесообразность использования дробных реплик для получения линейной модели объекта резко возрастает.

Рассмотрим пример ДФЭ. Положим, что интересующая нас модель объекта является функцией трех переменных (факторов), причем заранее известно, что в рабочей области эта модель близка к линейной, то есть

$$y = a_0 + a_1 \hat{x}_1 + a_2 \hat{x}_2 + a_3 \hat{x}_3. \quad (4.52)$$

При ПФЭ и двух уровнях факторов мы должны осуществить $N = 2^2 = 8$ опытов. Следовательно, для определения четырех коэффициентов $a_j (j = \overline{0,3})$ модели проводится вдвое больше число опытов.

Используя ДФЭ для решения поставленной задачи можно уменьшить число опытов. Чтобы дробная реплика представляла собой ортогональный план, в качестве реплики следует брать ПФЭ для меньшего числа факторов. Число опытов при этом должно быть больше (или равно) числа неизвестных коэффициентов модели. Следовательно, для рассматриваемого примера можно ограничиться четырьмя опытами. При этом ортогональность плана достигается, если использовать план ПФЭ типа 2^2 , а в качестве плана для \hat{x}_3 использовать столбец $\hat{x}_1 \hat{x}_2$. Матрица такого эксперимента приведена в таблице 4.6.

Полуреплика от ПФЭ типа 2^3

№ опыта	\hat{x}_0	\hat{x}_1	\hat{x}_2	$\hat{x}_3 = \hat{x}_1 \cdot \hat{x}_2$	y
1	+	+	+	+	y_1
2	+	—	+	—	y_2
3	+	+	—	—	y_3
4	+	—	—	+	y_4

Такой сокращенный план (половина ПФЭ) называется **полурепликой**. Пользуясь таким планированием, можно оценить свободный член и три коэффициента модели при линейных членах. Но если модель объекта не строго линейна и коэффициенты при парных взаимодействиях не равны нулю, то найденные коэффициенты будут смешанными в соответствии с системой (4.50), поскольку элементы столбцов $\hat{x}_1\hat{x}_2$, $\hat{x}_1\hat{x}_3$ и $\hat{x}_2\hat{x}_3$ (эти столбцы в табл. 4.6 не указаны) совпадают с элементами столбцов \hat{x}_3 , \hat{x}_2 , \hat{x}_1 соответственно. Если модель объекта действительно близка к линейной, то со смешанностью коэффициентов вполне можно примириться.

По указанному в табл. 4.6 плану проводится половина ПФЭ типа 2^3 и получается полуреплика от 2^3 , что принято обозначать как 2^{3-1} . Если знаки одного из элементов столбцов ($\hat{x}_1\hat{x}_2$ или \hat{x}_3) изменить на противоположные, то получим вторую половину ПФЭ (еще 4 опыта) и, следовательно, вторую полуреплику от 2^3 . При осуществлении обоих полуреplik мы провели бы ПФЭ 2^3 (8 опытов) и получили бы отдельные оценки коэффициентов как для линейных эффектов, так и для эффектов взаимодействий.

Планирование эксперимента, когда некоторые факторы приравнивают к произведению нескольких факторов, называется планированием со смешиванием. Его обозначают ДФЭ 2^{k-l} , где k – число факторов, l – число факторов, приравниваемых к произведениям

Пусть $k = 7$, $l = 4$, тогда план ДФЭ 2^{7-4} включает матрицу плана ПФЭ 2^3 для трех факторов $\hat{x}_1\hat{x}_2$ и \hat{x}_3 , а столбцы плана для остальных факторов представляют собой некоторые произведения факторов $\hat{x}_1\hat{x}_2$ и \hat{x}_3 , например, $\hat{x}_4 = \hat{x}_1\hat{x}_2$, $\hat{x}_5 = \hat{x}_1\hat{x}_3$, $\hat{x}_6 = \hat{x}_2\hat{x}_3$, $\hat{x}_7 = \hat{x}_1\hat{x}_2\hat{x}_3$. Такие равенства называют **генерируемыми соотношениями**, их выбор произволен, однако повторение не допускается. При таком планировании все линейные эффекты смешаны с эффектами взаимодействия.

Число опытов в дробной реплике должно удовлетворять неравенству $k+1 \leq N < 2^k$, где k – число факторов, для получения несмешанных оценок линейных эффектов.

Если $N = k+1$, то есть число опытов равно числу определяемых коэффициентов линейной модели, дробная реплика представляет собой **насыщенный линейный ортогональный план**.

Поскольку число опытов в насыщенных планах равно числу определяемых коэффициентов, число степеней свободы дисперсии адекватности $f_{ад} = 0$. Для проверки адекватности линейной модели, полученной по насыщенному плану, необходим дополнительный эксперимент.

Таким образом, оптимальные двухуровневые планы 2^k и 2^{k-1} имеют следующие преимущества:

- планы ортогональны, и поэтому все вычисления просты, все коэффициенты определяются независимо друг от друга;
- каждый коэффициент определяется по результатам всех N опытов;
- для данного числа опытов N эти планы имеют минимальный определитель ковариационной матрицы $(X^T X)^{-1}$; в связи с этим все коэффициенты определяются с одинаковой и минимальной дисперсией, то есть планы обладают свойством Д-оптимальности;
- линейные планы 2^k и 2^{k-1} обладают свойством ротатабельности.

Покажем справедливость свойства ротатабельности в случае линейной модели

$$y = a_0 + \sum_{j=1}^k a_j \hat{x}_j. \quad (4.53)$$

Вследствие отсутствия корреляции между коэффициентами по закону сложения дисперсий для линейной модели (4.53) имеем:

$$\sigma_y^2 = \sigma_{a_0}^2 + \hat{x}_1^2 \sigma_{a_1}^2 + \dots + \hat{x}_k^2 \sigma_{a_k}^2.$$

Так как $\sigma_{aj}^2 = \frac{\sigma_{\text{воспр}}^2}{N}$, $j = \overline{1, k}$, получим

$$\sigma_y^2 = \frac{\sigma_{\text{воспр}}^2}{N} (1 + \hat{x}_1^2 + \hat{x}_2^2 + \dots + \hat{x}_k^2) = \frac{\sigma_{\text{воспр}}^2}{N} (1 + \rho^2);$$

$$\rho^2 = \sum_{j=1}^k \hat{x}_j^2,$$

где ρ – радиус сферы в K -мерном пространстве.

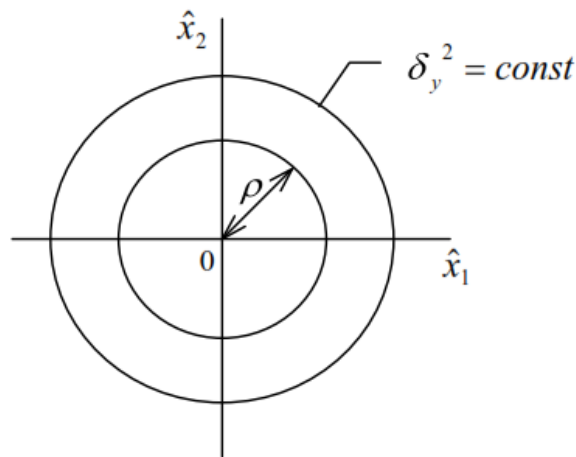


Рис. 4.3

Дисперсия выхода y растет пропорционально квадрату радиуса сферы ρ^2 и одинаково для всех эквидистантных точек (рис. 4.3).

Планирование, обладающее таким свойством, называется **ротатабельным планированием**.

4.2.7. Учет влияния помех

При постановке многофакторного эксперимента на реальном объекте необходимо спланировать его так, чтобы иметь возможность отфильтровать помехи. С этой целью каждый опыт в матрице планирования повторяется в m раз (табл. 4.7). Обычно $m = 3 \dots 5$.

В каждой строчке матрицы планирования определяются средние значение и дисперсия выхода объекта по m параллельным опытам:

$$\bar{y}_i = \frac{1}{m} \sum_{j=1}^m y_{ij}; \quad \sigma_i^2 = \frac{1}{m-1} \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2, \quad i = \overline{1, N}, \quad (4.54)$$

где y_{ij} – j -ое значение выхода объекта в i -ом опыте.

Проверяется однородность выборочных дисперсий (воспроизводимость эксперимента) по критерию Кохрена. Для этого определяется отношение максимальной дисперсии к сумме всех дисперсий:

$$G_{\text{рас}} = \frac{\max \sigma_i^2}{\sum_{i=1}^N \sigma_i^2}. \quad (4.55)$$

ПФЭ типа 2^k

№ опыта	\hat{x}_0	\hat{x}_1	\hat{x}_2	...	\hat{x}_k	y	\bar{y}	σ^2
1	+	+	+		+	$y_{11}, y_{12}, \dots, y_{1m}$	\bar{y}_1	σ^2_1
2	+	—	+		+	$y_{21}, y_{22}, \dots, y_{2m}$	\bar{y}_2	σ^2_2
3	+	+	—		+	$y_{31}, y_{32}, \dots, y_{3m}$	\bar{y}_3	σ^2_3
·	·	·	·		·	·	·	·
·	·	·	·		·	·	·	·
·	·	·	·		·	·	·	·
N	+	—	—		—	$y_{N1}, y_{N2}, \dots, y_{Nm}$	\bar{y}_N	σ^2_N

Полученное отношение сравнивается с табличным значением (определяется из таблицы распределения Кохрена) G_T при заданной доверительной вероятности P и числе степеней свободы $f_1 = m - 1$, $f_2 = N$.

Если $G_{\text{рас}} < G_m$ дисперсии σ_i^2 однородны (опыты воспроизводимы), в противном случае добиваются достижения однородности дисперсий σ_i^2 , как правило, путем увеличения числа повторных опытов m .

Тогда в качестве оценки для дисперсии воспроизводимости

$$\sigma_{\text{воспр}}^2 = \sigma_i^2 = \frac{1}{N(m-1)} \sum_{i=1}^N \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2$$

можно взять среднюю дисперсию

$$\sigma_{\text{воспр}}^2 = \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \quad (4.56)$$

с числом степеней свободы $f_{\text{воспр}} = N(m-1)$.

Коэффициенты модели определяются изложенным в разделе 4.2.3 способом. При этом в формулах (4.34), (4.36), (4.37) для вычисления коэффициентов следует вместо y_i принимать значение \bar{y}_i . Например, коэффициенты линейных эффектов определяются по формуле

$$a_j = \frac{1}{N} \sum_{i=1}^N \hat{x}_{ji} \bar{y}_i, \quad j = \overline{0, k}.$$

Упрощение модели путем отсеивания ее незначимых коэффициентов осуществляется по критерию Стьюдента (раздел 4.2.4), а проверка адекватности

модели — по критерию Фишера (раздел 4.2.5). Однако, в данном случае необходимо учитывать следующее. Дисперсия воспроизводимости определяется по формуле (4.56). Учитывая, что дисперсия среднего \bar{y} , полученного по выборке объема m , в m раз меньше дисперсии единичного измерения \bar{y} , то есть

$$\sigma_{\bar{y}}^2 = \frac{1}{m} \sigma_y^2 = \frac{1}{m} \sigma_{\text{воспр}}^2, \quad (4.57)$$

в рассматриваемом случае дисперсия коэффициентов σ_{aj}^2 определяется следующим образом:

$$\sigma_{aj}^2 = \frac{1}{Nm} \sigma_{\text{воспр}}^2 \quad \forall j \quad (4.58)$$

В матрице планирования каждый опыт повторялся в m раз. Поэтому дисперсию адекватности нужно умножить на m , то есть

$$\sigma_{\text{ад}}^2 = \frac{m}{N-r} \sum_{i=1}^N (\bar{y}_i - y_i^M)^2. \quad (4.59)$$

Незначимые коэффициенты отсеиваются из уравнения регрессии. При этом ввиду ортогональности матрицы планирования остальные коэффициенты не приходится пересчитывать. В случае же не подтверждения адекватности модели необходимо увеличить порядок аппроксимирующего полинома.

Необходимость повышения порядка модели может определяться также исходя из следующего соображения. На практике часто оказывается, что линейная модель, адекватно описывающая опытные данные, поставленные в точках плана, неудовлетворительно характеризует внутреннюю часть изученной области факторного пространства. Поэтому имеет смысл поставить опыт в центре плана ($\hat{x}_j = \hat{x}_j^0, j = \overline{0, k}$). Среднее значение выхода объекта в центре плана должно быть соизмеримо (в пределах ошибки воспроизводимости) с коэффициентом a_0 модели, то есть

$$|a_0 - \bar{y}^0| < \delta, \quad (4.60)$$

где $\bar{y}^0 = \frac{1}{m} \sum_{i=1}^{m_1} y_i^0$;

m_1 — число повторных опытов в центре плана;

δ — малое число.

В случае нарушения условия (4.60) для математического описания рассматриваемой области факторного пространства потребуется уравнение более высокого порядка.

4.2.8. Построение модели с квадратичными эффектами

Для построения модели, содержащей степени факторов, требуется иметь такую систему планирования, в которой каждый фактор будет принимать хотя бы три разных уровня. ПФЭ типа 3^k содержит слишком большое число опытов. При $k > 5$ ПФЭ 3^k становится совершенно неприемлемым для практики.

Сократить число опытов можно, если воспользоваться **композиционными или последовательными планами**, предложенными Боксом и Уилсоном. Ядро таких планов составляет ПФЭ 2^k при $k > 5$ или полуреплика от него ДФЭ 2^{k-1} при $k > 5$. Возможность использования в качестве ядра плана 2^{k-1} при $k > 5$ обусловлена тем, что уже полуреплика обеспечивает получение несмешанных оценок для линейных эффектов и эффектов парного взаимодействия.

Если линейная модель (а также модель с эффектами взаимодействия) оказалась неадекватной, необходимо:

1) добавить $2k$ звездных точек, расположенных на координатных осях факторного пространства. Координаты звездных точек: $(\pm \alpha, 0, 0, \dots, 0)$; $(0, \pm \alpha, 0, \dots, 0)$; ...; $(0, 0, 0, \dots, \pm \alpha)$, где α – расстояние от центра плана до звездной точки (звездное плечо);

2) увеличить число опытов n_0 в центре плана.

Рассмотрим построение композиционного плана при $k = 2$ (рис. 4.4). Точки 1 – 4 образуют ПФЭ 2^2 , точки 5 – 8 – звездные точки с координатами $(\pm \alpha, 0)$ и $(0, \pm \alpha)$, координаты n_0 опытов в центре плана нулевые – $(0, 0)$.

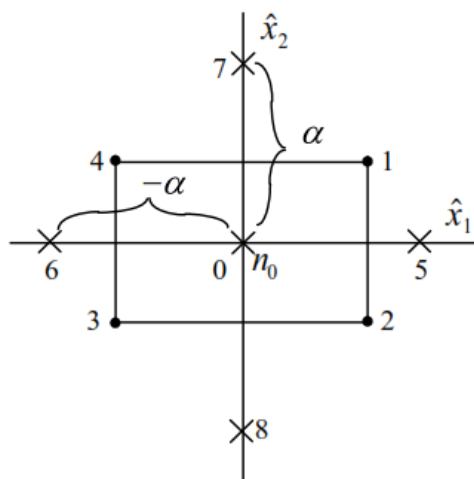


Рис. 4.4

Общее число опытов в матрице композиционного плана второго порядка при k факторах:

$$N = 2^k + 2k + n_0 \text{ при } k < 5;$$

$$N = 2^{k-1} + 2k + n_0 \text{ при } k \geq 5.$$

Матрица плана приведена в табл. 4.8. Композиционные планы легко приводятся к ортогональным выбором соответствующего звездного плеча α . На количество опытов в центре плана n_0 при этом не накладывается никаких ограничений (обычно $n_0 = 1$). При этом достаточно определить α из условия:

$$\alpha^4 + 2^k \alpha^2 - 2^{k-1}(k + 0.5n_0) = 0 \text{ при } k < 5;$$

$$\alpha^4 + 2^{k-1} \alpha^2 - 2^{k-2}(k + 0.5n_0) = 0 \text{ при } k \geq 5.$$

Благодаря ортогональности матрицы планирования все коэффициенты модели определяются независимо друг от друга по формулам:

$$a'_0 = \frac{1}{N} \sum y_i;$$

$$a_j = \frac{\sum \hat{x}_{ji} y_i}{\sum \hat{x}_{ji}^2}, \quad j = \overline{1, k};$$

$$a_{jl} = \frac{\sum \hat{x}_{ji} \hat{x}_{li} y_i}{\sum \hat{x}_{ji}^2 \hat{x}_{li}^2}, \quad j, l = \overline{1, k}, \quad j < l;$$

$$a_{jj} = \frac{\sum \left(\hat{x}_{ji}^2 - \frac{1}{N} \sum \hat{x}_{ji}^2 \right) y_i}{\sum \left(\hat{x}_{ji}^2 - \frac{1}{N} \sum \hat{x}_{ji}^2 \right)^2}, \quad j = \overline{1, k},$$

где суммирование всюду осуществляется по i от 1 до N .

Таблица 4.8

Матрица композиционного плана

№ опыта	\hat{x}_1	\hat{x}_2	y
1	+1	+1	y_1
2	-1	+1	y_2
3	+1	-1	y_3
4	-1	-1	y_4
5	$+\alpha$	0	y_5
6	$-\alpha$	0	y_6
7	0	$+\alpha$	y_7
8	0	$-\alpha$	y_8
9	0	0	y_9
.	.	.	.
.	.	.	.
N	0	0	y_N

В результате расчетов по ортогональной матрице с преобразованными столбцами для квадратичных эффектов получим уравнение вида

$$y = a'_0 + \sum_{j=1}^k a_j \hat{x}_j + \sum_{\substack{j,l=1 \\ (j<l)}}^k a_{jl} \hat{x}_j \hat{x}_l + \sum_{j=1}^k a_{jj} (\hat{x}_j^2 - \overline{\hat{x}_j^2}); \quad \overline{\hat{x}_j^2} = \frac{1}{N} \sum_{i=1}^N \hat{x}_{ji}^2.$$

Чтобы перейти к обычной записи коэффициент a_0 определяют по формуле

$$a_0 = \frac{1}{N} \sum_{i=1}^N y_i - \frac{a_{11}}{N} \sum_{i=1}^N \hat{x}_{1i}^2 - \frac{a_{22}}{N} \sum_{i=1}^N \hat{x}_{2i}^2 - \dots - \frac{a_{kk}}{N} \sum_{i=1}^N \hat{x}_{ki}^2.$$

В результате получим уравнение вида

$$y = a_0 + \sum_{j=1}^k a_j \hat{x}_j + \sum_{\substack{j,l=1 \\ (j<l)}}^k a_{jl} \hat{x}_j \hat{x}_l + \sum_{j=1}^k a_{jj} \hat{x}_j^2. \quad (4.61)$$

Коэффициенты модели, получаемые при помощи ортогональных планов второго порядка, определяются с разной точностью. Поэтому для расчета оценки дисперсий коэффициентов a_j используют следующие выражения:

$$\sigma_{a_j}^2 = \frac{\sigma_{\text{воспр}}^2}{\sum_{i=1}^N \hat{x}_{ji}^2}, \quad j = \overline{1, k};$$

$$\sigma_{a_{jl}}^2 = \frac{\sigma_{\text{воспр}}^2}{\sum_{i=1}^N (\hat{x}_{ji} \hat{x}_{li})^2}, \quad j, l = \overline{1, k}, \quad l > j;$$

$$\sigma_{a_{jj}}^2 = \frac{\sigma_{\text{воспр}}^2}{\sum_{i=1}^N (\hat{x}_{ji}^2 - \overline{\hat{x}_{ji}^2})^2}, \quad j = \overline{1, k};$$

$$\sigma_{a_0}^2 = \frac{1}{N} \sigma_{\text{воспр}}^2 + \sigma_{a_{11}}^2 (\overline{\hat{x}_1^2})^2 + \sigma_{a_{22}}^2 (\overline{\hat{x}_2^2})^2 + \dots + \sigma_{a_{kk}}^2 (\overline{\hat{x}_k^2})^2.$$

Дисперсия $\sigma_{a_0}^2$ получена по закону сложения дисперсий. После вычисления коэффициентов проверяют значимость коэффициентов по критерию Стьюдента и адекватность модели по критерию Фишера.

Следует отметить, что ортогональные планы второго порядка не обладают свойством ротатабельности, что является их недостатком.

Однако простота всех вычислений оставляет за ортогональными планами право на существование.

4.3. Метод группового учета аргументов

При построении математических моделей сложных объектов необходимо учитывать следующие их особенности:

- высокая размерность вектора входных координат $\vec{x} = (x_1, x_2, \dots, x_k)$;
- наличие большого числа внутренних источников случайных помех, статистические характеристики которых, как правило, неизвестны;
- неизученность механизмов действия различных экономических, социальных и физико-химических явлений и процессов в объектах;
- трудность постановки экспериментов для получения большого числа сигналов \vec{x}^3, y^3 (y – одномерный выход объекта).

Указанные особенности затрудняют построение неформальных ММ сложных объектов и вынуждают описывать статические режимы их функционирования уравнениями вида

$$y = f(x) = a_0 + \sum_{j=1}^k a_j x_j + \sum_{j=1}^k a_{jj} x_j^2 + \sum_{j=1}^k a_{j\gamma} x_j x_\gamma + \sum_{j=1}^k a_{j\gamma\rho} x_j x_\gamma x_\rho + \sum_{j=1}^k a_{jjj} x_j^3 + \dots, \quad (4.62)$$

где $j < \gamma < \rho < k$;

\vec{a} – вектор неизвестных коэффициентов размерности q ;

$$\vec{a} = (a_0, a_1, \dots, a_k, \dots) \in E^q.$$

Если порядок h полинома по всем переменным x_j равен числу k входных координат x_j , то выражение (4.62) называют полным полиномом Колмогорова-Габбора.

При увеличении числа входных координат x_j – размерность вектора \vec{a} полного полинома возрастает исключительно быстро. Так, например, при $k = 2$ размерность вектора \vec{a} равна 6, при $k = 4$ она возрастает до 70.

Для определения вектора \vec{a} математической модели сложного объекта по экспериментальным данным $y_i^3; \vec{x}_i^3 = (x_{1i}^3, x_{2i}^3, \dots, x_{ki}^3)$ при $i = \overline{1, n}$ часто применяют МНК.

При этом сталкиваются с двумя затруднениями:

1. Вследствие слабой изученности сложных объектов многие входные координаты, включенные в ММ, оказываются сильно коррелированными, что приводит к плохой обусловленности системы нормальных уравнений МНК. Высокая размерность вектора \vec{a} делает невозможным какой-либо предварительный анализ степени обусловленности матрицы системы уравнений. Поэтому исследователь вынужден убеждаться в неустойчивости задачи определения \vec{a} лишь в результате ее многократного решения на ЭВМ.

2. Задача определения вектора \vec{a} заранее не заданной размерности q полинома типа (4.62) является, как правило, неустойчивой. Обоснованный же вы-

бор числа q структуры полинома (4.62) фактически невозможен из-за большой размерности таблицы экспериментальных данных:

$$[x_{ji}^{\ominus}, y_i^{\ominus}], i = \overline{1, n}, j = \overline{1, n},$$

и высокого уровня помех, налагаемых на выходную координату объекта. Вследствие этого при построении ММ размерность вектора \vec{a} и структура полинома (4.62) задаются в какой-то мере наугад; затем с помощью МНК определяются значения a ; вычисляется СКО σ_a аппроксимации полиномом (4.62) табличных данных $[x_{ji}^{\ominus}, y_i^{\ominus}]$, которые не были использованы для нахождения \vec{a} ; величина σ_a сравнивается с допустимой ошибкой Δ ММ. Если $\sigma_a > \Delta$, размерность \vec{a} увеличивают и (или) видоизменяют структуру ММ (4.62), определяют новые значения a (напомним, что ряд (4.62) не является ортогональным). Затем снова вычисляют величину σ_a , сравнивают ее с Δ и повторяют описанную процедуру до тех пор, пока при некоторой структуре полинома (4.62) и размерности \vec{a} не выполнится условие $\sigma_a > \Delta$. Подобный процесс построения ММ связан с исключительно большим объемом вычислений, так как структура ММ (4.62) и размерность \vec{a} заранее не заданы и должны быть определены в процессе решения задачи.

Указанные трудности вычислительного характера могут быть устранены или уменьшены, если для построения ММ сложного объекта применить метод группового учета аргументов (МГУА).

Согласно этому методу решение трудоемкой исходной задачи (как правило, неустойчивой) аппроксимации таблично заданной функции k переменных $[x_{ji}^{\ominus}, y_i^{\ominus}]$, $i = \overline{1, n}$, $j = \overline{1, k}$ полиномом (4.62), структура которого и размерность $q < nk$ заранее не заданы, заменяется многостадийным процессом решения большого числа сравнительно простых задач аппроксимации табличных данных функциями двух переменных – полиномами заданной структуры с невысокой размерностью вектора \vec{a} .

На каждой из стадий этого процесса производится отбор «наилучших», в некотором смысле, полиномов, которые используются на следующей стадии в качестве фиктивных аргументов новых полиномов. Подобная процедура построения сложной функции (полинома от полиномов) продолжается до тех пор, пока на какой-либо стадии не будет достигнута приемлемая точность описания свойств объекта (табличных данных) некоторой математической зависимостью $y = f(\vec{x}, \vec{a})$, быть может и меньшего, чем k , числа аргументов x_j .

Рассмотрим подробно первые две стадии этой процедуры.

Пусть задана совокупность (таблица) экспериментальных данных $[\vec{x}_i, y_i^{\ominus}]$ при $i = \overline{1, n}$, которую требуется аппроксимировать с допустимой средней квад-

ратической погрешностью Δ некоторой функцией, число переменных x_j которой меньше или равно k .

Разобьем эту совокупность данных на три множества:

обучающее $M_0 = [y_i^{\circ}, \bar{x}_i^{\circ}], i = \overline{1, n_1}; n_1 > 6$;

проверочное $M_n = [y_i^{\circ}, \bar{x}_i^{\circ}], i = \overline{n_1 + 1, n_2}$;

контрольное $M_k = [y_i^{\circ}, \bar{x}_i^{\circ}], i = \overline{n_2 + 1, n}$.

На первой стадии аппроксимации построим $C_k^2 = \frac{k(k-1)}{2}$ полиномов

$y_{j\lambda} = f(x_j, x_\lambda)$ двух аргументов (x_j, x_λ) типа

$$y_{j\lambda} = a_{0j\lambda} + a_{1j\lambda}x_j + a_{2j\lambda}x_\lambda + a_{3j\lambda}x_jx_\lambda + a_{4jj}x_j^2 + a_{5\lambda\lambda}x_\lambda^2, \quad (4.63)$$

где $j, \lambda = \overline{1, k}, j \neq \lambda, j < \lambda$;

Структура частной модели (4.63) может быть выбрана иная, например, при $a_{1j\lambda} = 0$ или $a_{4jj} = a_{5\lambda\lambda} = 0$ и т.п.

Такие полиномы будем называть частными.

Вектор коэффициентов $\vec{a}_{j\lambda}$ каждого частного полинома определяется МНК по экспериментальным данным $[y_i^{\circ}, x_{ji}, x_{\lambda i}] \in M_0, i = \overline{1, n_1}, n_1 > 6$.

Далее вычислим ординаты каждого частного полином $y_{j\lambda}$ с известными коэффициентами во всех точках $[x_{ji}, x_{\lambda i}] \in M_n, i = \overline{n_1 + 1, n_2}$ проверочного множества M_n . Обозначим эти данные через $y_{j\lambda}(i)$. Средняя квадратическая ошибка $\sigma_{j\lambda}$ аппроксимации экспериментальных данных $y_i^{\circ} \in M_n$ частным полиномом $y_{j\lambda}$ равна:

$$\sigma_{j\lambda} = \sqrt{\frac{1}{n_2 - n_1} \sum_{i=n_1+1}^{n_2} (y^{\circ}(i) - y_{j\lambda}(i))^2}, j, \lambda = \overline{1, k}, j \neq \lambda, (j < \lambda). \quad (4.64)$$

Среди всех величин $\sigma_{j\lambda}$ найдем наименьшую и обозначим ее через σ_1^* , $\sigma_1^* = \inf_{i, \lambda} \sigma_{j\lambda}$.

В соответствии с некоторым правилом отбора, в котором используются величины $\sigma_{j\lambda}$ из всех C_n^2 частных полиномов выделим несколько, например, m_1 «наилучших» полиномов и переобозначим их через $y_1, y_2, \dots, y_\ell, \dots, y_{m_1}$.

Теперь вычислим значения $y_1(i)$ отобранных полиномов во всех точках множества M_0 и M_k , а также определим СКО δ_{-1} для всех y_1 на множестве M_k

$$\delta_1 = \sqrt{\frac{1}{n - n_2} \sum_{i=n_2+1}^n (y_i^3 - y_1(i))^2}, \quad 1 = \overline{1, m_1}. \quad (4.65)$$

Наконец найдем величину

$$\delta_1^* = \inf_1 \delta_1,$$

которая понадобится на следующих стадиях аппроксимации.

На второй стадии отобранные полиномы $y_1, y_2, \dots, y_\ell, \dots, y_{m_1}$ рассматриваются как аргументы новых частных полиномов

$$Z_{\ell\gamma} = \tilde{a}_{0\ell\gamma} + \tilde{a}_{1\ell\gamma} y_\gamma + \tilde{a}_{2\ell\gamma} y_\gamma^2 + \tilde{a}_{3\ell\gamma} y_\ell y_\gamma + \tilde{a}_{4\ell\ell} y_\ell^2 + \tilde{a}_{5\gamma\gamma} y_\gamma^2, \quad (4.66)$$

где $1, \gamma = \overline{1, m_1}, 1 \neq \gamma (1 < \gamma)$.

Количество этих частных полиномов равно числу сочетаний из m_1 по два

$$C_{m_1}^2 = \frac{m_1(m_1 - 1)}{2}.$$

Коэффициенты $\tilde{a}_{0\ell\gamma}, \tilde{a}_{1\ell\gamma}, \dots, \tilde{a}_{5\gamma\gamma}$ каждого из этих полиномов определяются методом наименьших квадратов по множеству значений $[y_i^3, y_1(i), y_\gamma(i)] \in M_0$ при $1, \gamma = \overline{1, m_1}, 1 \neq \gamma, i = \overline{1, n_1}$, где значения $y_1(i), y_\gamma(i)$ вычисляются во всех точках множества M_0 по модели (4.63).

Определим СКО $\sigma_{1\gamma}$ аппроксимации каждым частным полиномом $z_{1\gamma}$ экспериментальных данных $y_i^3 \in M_n$. Для этого подставим полиномы y_1, y_2, \dots, y_{m_1} в выражение (4.66), вычислим значение $z_{1\gamma}(i)$ в каждой точке $\vec{x}_i^3 \in M_n, i = \overline{n_1 + 1, n_2}$ и, аналогично формуле (4.64) найдем $\sigma_{1\gamma}$:

$$\sigma_{1\gamma} = \sqrt{\frac{1}{n_2 - n_1} \sum_{i=n_1+1}^{n_2} [y_i^3 - z_{1\gamma}(i)]^2}, \quad 1, \gamma = \overline{1, m_1}, 1 \neq \gamma (\gamma > 1). \quad (4.67)$$

Среди всех величин $\sigma_{1\gamma}$ найдем наименьшую и обозначим ее через σ_2^* :

$$\sigma_2^* = \inf_{1, \gamma} \sigma_{1, \gamma}$$

Далее, в соответствии с принятым правилом отбора, из всех полиномов $z_{1\gamma}$ выделим множество «наилучших» полиномов и переобозначим их через $z_1, z_2, \dots, z_\mu, \dots, z_{m_2}$, где $m_2 < C_{m_1}^2$.

Каждый полином z_μ может быть функцией четырех независимых переменных x_j , поэтому вероятность аппроксимации им с приемлемой точностью экспериментальные данные $[\bar{x}_i^3, y_i^3]$ выше, чем на первой стадии. Вследствие этого необходимо выяснить целесообразность перехода к третьей стадии аппроксимации.

Процедура последовательной аппроксимации оканчивается, когда точность приближения независимых экспериментальных данных $y_i^3 \in M_n$ каким-либо «наилучшим» полиномом удовлетворить нашим требованиям. Заметим, что значения \bar{x}_i^3, y_i^3 из M_0 уже использовались для определения коэффициентов полиномов, а данные из M_n – для отбора «наилучших» полиномов, как на первой, так и на второй стадии.

Для проверки точности приближения вычислим сначала значения отобранных полиномов $z_\mu(i)$, $\mu = \overline{1, m_2}$, $i = \overline{n_2, n}$ во всех точках M_k по формуле

$$\delta_\mu = \sqrt{\frac{1}{n - n_2} \sum_{i=n_2+1}^n [y_i^3 - z_\mu(i)]^2}, \quad \mu = \overline{1, m_2}. \quad (4.68)$$

Найдем величину

$$\delta_2^* = \inf_{\mu} \delta_\mu.$$

Если окажется, что $\delta_2^* \leq \Delta$, то соответствующий полином z_μ типа (4.66) может рассматриваться как математическая модель сложного объекта. Если же $\delta_2^* > \Delta$, то необходима третья стадия аппроксимации, на которой в качестве фиктивных аргументов будут выступать z_1, z_2, \dots, z_{m_2} . Значения этих аргументов предварительно вычисляются во всех точках множества M_0 и сводятся в таблицу $[z_\mu(i), y_i]$, $i = \overline{1, n_1}$, $\mu = \overline{1, m_2}$.

Во многих случаях допустимая СКО (Δ) ММ бывает известна весьма неточно и фактически не может быть использована для сравнения с δ_2^* . В таких случаях для окончания процесса аппроксимации можно применять величину

$$E = \frac{|\delta_1^* - \delta_2^*|}{\delta_2^*}.$$

Если E относительно невелика ($E \leq 0.1 \div 0.15$), то дальнейшая аппроксимация малоэффективна и в качестве ММ объекта можно использовать полином Z_j , для которого величина δ_j будет наименьшей. Если же E относительно

велика, то необходимо третья стадия аппроксимации, вся последовательность расчетов на которой такая же, как и на второй стадии.

Приведем еще одно условие окончания процесса аппроксимации, рекомендованное в трудах Ивахненко А.Г.

Практика применения МГУА показывает, что в большинстве случаев с ростом β величина δ_β^* вначале убывает, а затем начинает быстро возрастать. В таких ситуациях необходимо оканчивать процесс последовательной аппроксимации при достижении наименьшего значения СКО δ_β^* . При этом, оказывается, число стадий β обычно не превышает $k - 1$.

Анализ особенностей МГУА

1. На начальных стадиях аппроксимации экспериментальных данных $[\bar{x}_i^3, y_i^3]$, $i = \overline{1, n_1}$, $\bar{x}_i = (x_{1i}, x_{2i}, \dots, x_{ki})$ отношение числа kn_1 этих данных к числу коэффициентов частных полиномов, определяемых в совокупности, намного больше единицы. Вследствие этого коэффициенты «наилучших» полиномов, приближающих точки $y_i^3 \in M_0$ и, отчасти, $y_i^3 \in M_n$, весьма слабо зависят от помехи, налагаемой на выходную координату объекта, а СКО ММ остается примерно одинаковой как на M_n , так и на M_k . С увеличением номера стадии β возрастает число аргументов полинома и общее число коэффициентов, определяемых по точкам $y_i^3 \in M_0$ и $y_i^3 \in M_n$, что, естественно, приводит к уменьшению величины σ_β^* . Одновременно с ростом β уменьшается отношение числа экспериментальных данных к числу определяемых коэффициентов (это отношение может стать < 1). Поэтому получаемая ММ фактически интерполирует $y_i^3 \in M_0$ и, отчасти, $y_i^3 \in M_n$, т.е. весьма точно описывает неточные экспериментальные данные, искаженные случайными помехами. ММ, коэффициенты которой найдены интерполяционным методом по случайным величинам, не может надежно описывать независимые экспериментальные данные $y_i^3 \in M_k$, поэтому во многих случаях величина δ_β^* будет нарастать с увеличением номера стадии β .

Такое явление будем называть неустойчивостью МГУА.

Неустойчивость МГУА наблюдается и при анализе величины σ_β^* , которая часто убывает с ростом β , а затем начинает также возрастать. Величина σ_β^* менее чувствительна к росту β , так как данные $y_i^3 \in M_n$ участвуют в формировании ММ.

Для предотвращения неустойчивости МГУА в работах Ивахненко А. Г. рекомендуется выбрать число стадий β из условия минимумам σ_β^* . Но во многих слу-

чаях величина δ_{β}^* более чувствительна к увеличению β и более целесообразно оканчивать процесс аппроксимации, как только δ_{β}^* достигнет своего минимума.

2. После завершения процесса аппроксимации строится окончательное выражение функции $f(\vec{x}, \vec{a})$ путем последовательного исключения промежуточных аргументов. При большом числе стадий структура модели $f(\vec{x}, \vec{a})$ может оказаться весьма сложной и неудобной в расчетах. В таких случаях $f(\vec{x}, \vec{a})$ задается набором полиномов.

3. Остановимся теперь на правиле отбора «наилучших» полиномов. По существу в МГУА отсутствует формализованная процедура выбора на каждой стадии таких полиномов, на базе которых можно построить $f(\vec{x}, \vec{a})$, описывающую экспериментальные данные $[\vec{x}_i, y_i^3]$ с заранее заданной погрешностью. Можно только предполагать, что подобные полиномы находятся среди тех частных полиномов, СКО аппроксимации которых точек $y_i^3 \in M_n$ относительно малы. Чрезмерное уменьшение (особенно на начальных стадиях) числа отбираемых полиномов увеличивает вероятность потери настоящих аргументов функции $f(\vec{x}, \vec{a})$; чрезмерное увеличение ведет к резкому росту объема вычислений. Поэтому на каждой стадии целесообразно отбирать примерно одно и то же число m_{β} полиномов с наименьшими значениями σ . Если при выбранном наугад числе m_{β} не удастся построить желаемую функцию $f(\vec{x}, \vec{a})$, то все расчеты повторяются при другом, несколько большем m_{β} .

4. В методе МГУА не требуется предварительно задавать структуру модели $f(\vec{x}, \vec{a})$. Она формируется в процессе многостадийного отбора «наилучших» частных полиномов.

5. Полученная данным методом зависимость $f(\vec{x}, \vec{a})$ может быть функцией меньшего, чем k числа входных переменных x . При исследовании сложных объектов всегда имеется опасность включения в число входных координат таких факторов, влияние на y которых незначимо или не может быть оценено при данном уровне помех. Устранение из ММ таких незначимых факторов также осуществляется МГУА в процессе многостадийного отбора «наилучших» частных моделей.

6. Число стадий аппроксимации зависит от многочисленных факторов. Важнейшими из них являются число входных координат, допустимая погрешность Δ ММ, структура частных полиномов, число отбираемых полиномов, соотношение чисел n_1, n_2, n уровень помехи и др.

7. При использовании МГУА на каждой стадии имеем дело с матрицей системы нормальных уравнений МНК порядка, не превышающего 6×6 , что, в общем случае, способствует улучшению ее обусловленности. Если же размерность вектора \vec{a} частных полиномов равна $3 \div 4$, то появляется возможность

сравнительно просто оценивать степень обусловленности матрицы B системы нормальных уравнений МНК по критерию Фоке

$$F = |\det B| \cdot \|B\|_E^{-1} \sqrt{K_K},$$

где $\|B\|_E$ – эвклидова норма матрицы B с элементами $b_{ij}, i, j = \overline{1, k}$;

K_K – порядок матрицы B ;

$$\|B\|_E = \sqrt{\sum_{i,j=1}^{k_1} b_{ij}^2}.$$

Частные полиномы, матрицы B которых плохо обусловлены, т.е. величины F малы, не должны включаться в число отбираемых на следующую стадию.

В настоящее время многие вопросы применения и обоснования МГУА еще нуждается в детальной разработке, недостаточен и опыт его применения для решения задач моделирования объектов.

Практическая рекомендация по группировке исходных данных $[\bar{x}_i^3, y_i^3]$, $i = \overline{1, n_1}$ по множествам M_0, M_n, M_k . Определяются среднее арифметическое значение \bar{y}^3 по всем данным и отклонения случайных величин y_i^3 от \bar{y}^3 . Ординаты y_i^3 , обладающие наибольшими отклонениями включают в M_0 , наименьшими – в M_k , остальные – в M_n .

Структура алгоритма самоорганизации модели на ЭВМ. Многорядный алгоритм селекции с пороговым самоотбором реализует МГУА. Его можно эффективно применять для большого числа переменных (до 1000), когда таблица исходных экспериментальных данных содержит всего 15-20 точек.

Структурная схема алгоритма показана на рис. 4.5.

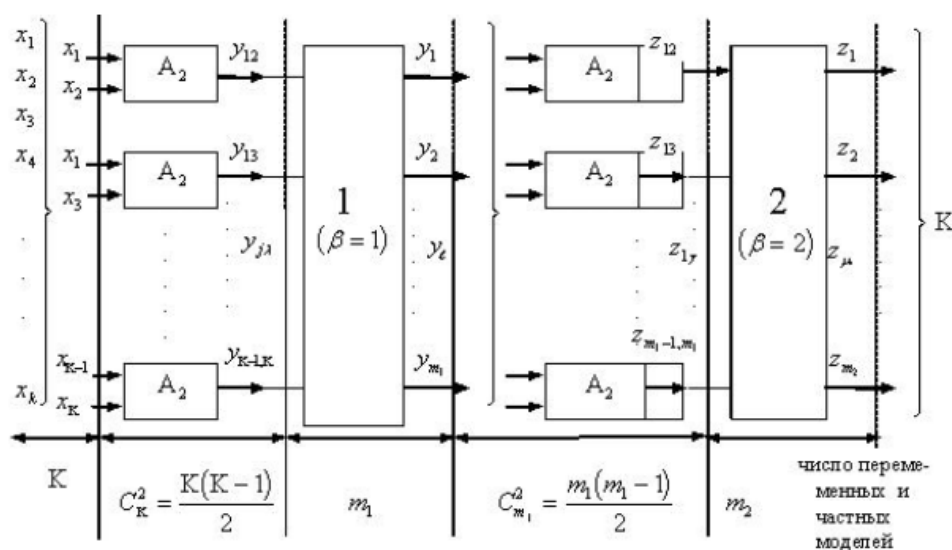


Рис. 4.5

Здесь: 1,2 – пороговые отборы $C_{m_\beta}^2 = \frac{m_\beta(m_\beta - 1)}{1 \cdot 2}$ лучших частных моделей $y_1(1 = \overline{1, m_1})$ и $z_\mu(1 = \overline{1, m_2})$ по критерию селекции;
 β – номер ряда селекции;
 A_2 – алгоритм попарной обработки переменных.

В первом ряду селекции образуется всевозможные пары аргументов, и для каждой из них находится частная модель вида

$$y_{j\lambda} = f_i(x_j, x_\lambda) = a_{0j\lambda} + a_{1j\lambda}x_j + a_{2j\lambda}x_\lambda + a_{3j\lambda}x_jx_\lambda + a_{4jj}x_j^2 + a_{5\lambda\lambda}x_\lambda^2,$$

где $i = \overline{1, d_1}$;

$$d_1 = C_k^2 = \frac{k(k-1)}{2};$$

$$j = \overline{1, k-1};$$

$$\lambda = j+1, j+2, \dots, k;$$

k – число входных переменных первого ряда селекции.

Оценки коэффициентов частных моделей $y_{j\lambda}$ рассчитываются МНК по части опытных данных, называемой обучающей последовательностью. Из всех частных моделей выбирается m_1 лучших моделей y_1, y_2, \dots, y_{m_1} .

Во втором ряду селекции образуются пары выходных переменных, прошедших первый ряд и для каждой из них находятся частные модели второго ряда $z_{\ell\gamma}$ и т.д.

Число предложенных алгоритмов МГУА непрерывно возрастает, но все они основаны на указанном выше едином принципе самоорганизации.

4.4. Определение наилучшей модели

Критерий выбора модели. При анализе сложных систем большое практическое значение имеют способы определения наилучшей модели объекта из некоторой их совокупности. Рассмотрим этот вопрос на примере регрессионных моделей.

Цель предварительного выбора наилучшей модели достигается применением в отдельности или в некоторой совокупности следующих критериев:

- наименьшее число коэффициентов модели, совместимое с заданной допустимой погрешностью;
- простота структуры модели, совместимая с допустимой погрешностью;
- разумные физические основания;

- минимальная сумма квадратов отклонений между предсказанными y_i^M и эмпирическими значениями y_i выходной переменной y ;
- минимальная оценка дисперсии σ_y^2 выходной переменной.

Анализ погрешностей. Прежде всего, при оценке моделей проводят анализ остатков (погрешностей) в виде разностей $y_i - y_i^M$. В регрессионном анализе используется ряд основных предположений:

- независимость ошибок;
- постоянства их дисперсий;
- нормальный закон распределения ошибок.

Если модель адекватно описывает экспериментальные данные, то погрешности должны не противоречить этим предположениям. Анализ погрешностей - это способ проверки того, что то или иное предположение не нарушено.

Можно выделить следующие наиболее характерные результаты анализа погрешностей:

1. Обнаружение выбросов, т.е. резко отличающихся, экстремальных значений y_i от предполагаемого значения y_i^M . В этом случае необходимо повторить эксперимент в точке выброса, предварительно проверив правильность его условий. Если выброс устойчив, то его надо или выбросить и пересчитать коэффициенты модели без него, или исследовать физическую природу его существования.

2. Обнаружение некоторого тренда в остатках, т.е. тенденции в их изменении с течением времени. Например, в остатках наблюдается тенденция к линейному росту. Для улучшения модели вычисляют поправку в виде разности

$\frac{y_i - y_i^M}{y_i^M} - \alpha_i$, $\alpha_i = kN$, $k = \text{tg } \alpha$, α – угол наклона усредненной линии остатков к оси абсцисс, N – номер эксперимента.

3. Обнаружение разного сдвига уровня процесса. В этом случае нужно выяснить причину резкого скачка погрешности, а затем разбить выборку на две и для каждого уровня построить модель.

4. Обнаружение изменений в дисперсии ошибки. Анализ, который проводят при построении регрессионных моделей, позволяет найти некоторое среднее значение дисперсии ошибки (например, дисперсию воспроизводимости). Если дисперсии ошибки неоднородны, то найденная средняя дисперсия может неверно описывать часть экспериментальных данных.

5. Исследование остатков для проверки того, описываются ли они нормальным законом распределения. В этом случае можно проверить случайность значений остатков и воспользоваться кривой нормального распределения.

Сравнение двух регрессионных моделей. Рассмотрим некоторые методы сравнения регрессионных моделей, применимые как к линейным, так и нелинейным моделям.

1. Критерий Хоэла. Предполагается, что имеется адекватная м модель y_1^M , которая удовлетворительно описывает экспериментальные данные в области ее применимости. С ней конкурирует другая модель y_2^M . Ставится задача: не следует ли отказаться от модели y_1^M и отдать предпочтение модели y_2^M . Строят тестовую линейную зависимость в виде уравнения в параметрической форме

$$y - y_1^M = \lambda(y_2^M - y_1^M), \quad (4.69)$$

где $\lambda > 0$. Проверка сводится к оценке в уравнении (4.69) углового коэффициента (параметра) λ . Если λ значимо положителен ($\lambda \geq 1$), от модели y_1^M отказываются в пользу второй модели y_2^M . Действительно, из (4.69) имеем:

$$y - y_1^M = \lambda(y_2^M - y_1^M);$$

$$\lambda(y_2^M - y) = (\lambda - 1)(y_1^M - y).$$

Обозначим $\Delta_1 = y_1^M - y$, $\Delta_2 = y_2^M - y$ – ошибки модели y_1^M и y_2^M .

Тогда $\lambda\Delta_2 = (\lambda - 1)\Delta_1$.

Отсюда при $\lambda \geq 1$ имеем

$$|\Delta_2| < |\Delta_1|.$$

Если λ незначимо положителен ($\lambda < 1$), то нельзя определить, какая из моделей лучше. Действительно, $\lambda < 1$ не дает определенного результата, так как возможно выполнение и неравенства $|\Delta_2| < |\Delta_1|$ (в пользу модели y_1^M) и противоположного ему неравенства $|\Delta_2| < |\Delta_1|$ (в пользу модели y_2^M), т.е. неизвестно условие нарушения неравенств. В этом случае отдается предпочтение модели y_1^M .

Критерий Хоэла называется несимметричным, так как он может использоваться только при $\lambda > 0$. Таким образом, подставляя в уравнение (4.69) значения y_i , y_{1i}^M , y_{2i}^M можно дать оценку погрешностей ММ регрессионных моделей y_1^M , y_2^M и выбрать путем оценки параметра λ лучшую из них.

2. Критерий Вильяма и Клута. Для сравнения двух регрессионных моделей, которые, по крайней мере, первоначально представляются равноценными, можно использовать симметричный критерий Вильяма и Клута. Проверка осуществляется путем оценки параметра λ в тестовом уравнении

$$y - \frac{1}{2}(y_1^M + y_2^M) = \lambda(y_2^M - y_1^M), \quad (4.70)$$

где y принимает эмпирические значения. Предположим, что модель y_1^M удовлетворительна, тогда

$$y = \eta_1 + \delta, \quad (4.71)$$

где η_1 – теоретическое значение, предсказанное первой моделью в предположении, что ее коэффициенты регрессии вычислены точно, δ – погрешность модели. Подставляя (4.71) в (4.70) получим

$$\eta_1 + \delta - \frac{1}{2}(y_1^M + y_2^M) = \lambda(y_2^M - y_1^M). \quad (4.72)$$

Вследствие того, что модель y_1^M достаточно адекватна, можно положить $\eta_1 \approx y_1^M$ и преобразовать выражение (4.72) к виду

$$-\frac{1}{2}(y_2^M - y_1^M) + \delta = \lambda(y_2^M - y_1^M).$$

Отсюда при малых ошибках δ следует, что модели y_1^M отдается предпочтение при $\lambda \approx -\frac{1}{2}$.

Теперь предположим, что удовлетворительна модель y_2^M , тогда

$$y = \eta_2 + \delta. \quad (4.73)$$

Подставляя (4.73) в (4.71) и предполагая, что $\eta_2 \approx y_2^M$, получим

$$\frac{1}{2}(y_2^M - y_1^M) + \delta = \lambda(y_2^M - y_1^M).$$

Следовательно, если верна модель y_2^M , то получим оценку $\lambda \approx \frac{1}{2}$. Таким образом, значимое отрицательное значение параметра $\lambda \left(\lambda \approx -\frac{1}{2} \right)$ указывает на

то, что модель y_1^M лучше модели y_2^M . Если λ значимо положителен, то окажется лучше модель y_2^M . Если λ незначительно отличается от нуля, то выбора между двумя моделями сделать нельзя - они равнозначны.

Итак, подставляя в уравнение (4.70) значения y_i , y_{1i}^M , y_{2i}^M взятые из допустимой области применения регрессионных моделей y_1^M , y_2^M можно путем количественной оценки параметра λ выбрать лучшую из этих моделей. критерий Вильяма и Клута называется симметричным, так как он может применяться и при положительных и при отрицательных значениях λ .

Контрольные вопросы

1. Опишите выбор существенных переменных модели объекта.
2. На какие основные этапы разбивается эксперимент?
3. Как осуществляется нормирование переменных модели?
4. Опишите полный факторный эксперимент. Основные стадии. Особенности.
5. Каким образом вычисляются коэффициенты модели? Как происходит оценка значимости коэффициентов модели?
6. Приведите пример проверки адекватности модели.
7. В чем особенности проведения дробного факторного эксперимента?
8. Как осуществляется учет влияния помех?
9. Опишите метод группового учета аргументов.
10. Как осуществляется определение наилучшей модели? Какие критерии при этом используются?

ЗАКЛЮЧЕНИЕ

Отличительная особенность моделирования как метода исследования заключается в возможности изучения, прогнозирования и оптимизации объектов, физический эксперимент с которыми очень затруднителен, опасен или не выгоден. Вычислительный эксперимент с моделью объекта в подобных ситуациях значительно сокращает сроки и стоимость исследований, повышает обоснованность принимаемых решений. По отношению к модели исследователь является экспериментатором, проводящим эксперимент не с реальным объектом, а с его моделью.

По результатам опытов с моделью количественно предсказывают поведение объекта в рабочих условиях, определяют оптимальное решение, результаты моделирования распространяют на оригинал. Причем распространение на оригинал выводов, полученных в опытах с моделью, не обязательно должно означать простое равенство тех или иных параметров оригинала и модели. Достаточно получить правило расчета интересующих нас параметров оригинала.

Специфика работы исследователя моделей заключается в том, что решаемые им задачи весьма разнообразны, поскольку моделирование объектов реализуется в любой области человеческой деятельности. Это приводит к широкому спектру математических методов, применяемых для их решения. Поэтому специалист по моделированию объектов должен владеть достаточным арсеналом разнообразных математических методов – теории вероятностей, математической статистики, математического анализа и дискретной математики, численного решения дифференциальных уравнений, структурной и параметрической оптимизации, теории случайных процессов и массового обслуживания.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып. 1 и 2. – М.: Мир, 1974. Вып. 1–408 с. Вып. 2–200 с.
2. Сейдж Э. П., Мелса Дж. Л. Идентификация систем управления. – М.: Наука, 1974. – 248 с.
3. Эйкхофф П. Основы идентификации систем управления. – М.: Мир, 1975. – 683 с.
4. Химмельблау Д. Анализ процессов статистическими методами. – М.: Мир, 1973. – 957 с.
5. Растрингин Л. А. Введение в идентификацию объектов управления. – М.: Энергия, 1977. – 216 с.
6. Брикман М. С., Кристинков Д. С. Аналитическая идентификация управляемых систем. – Рига: Зинатне, 1974. – 206 с.
7. Адаптивные системы идентификации [Текст] : научное издание /А. Г. Кику [и др.] ; под ред. В. И. Костюк. – Киев: Техника, 1975. – 284 с.
8. Перельман И. И. Оперативная идентификация объектов управления. – М.: Энергоиздат, 1982. – 272 с.
9. Райбман Н. С. Что такое идентификация? – М.: Наука, 1970. – 118 с.
10. Себер Дж. Линейный регрессионный анализ. – М.: Мир, 1980. – 456 с.
11. Горский В. Г., Адлер Ю. П. Планирование промышленных экспериментов (модели статики). – М.: Metallургия, 1974. – 264 с.
12. Горский В. Г., Адлер Ю. П. Планирование промышленных экспериментов (модели динамики). – М.: Metallургия, 1978. – 112 с.
13. Адлер Ю. П., Маркова Е. В., Грановский Ю. В. Планирование эксперимента при поиске оптимальных условий. – М.: Наука, 1976. – 280 с.
14. Дейч А. М. Методы идентификации динамических объектов. – М.: Энергия, 1979. – 240 с.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
1. СВЕДЕНИЯ О МЕТОДАХ МОДЕЛИРОВАНИЯ И ИДЕНТИФИКАЦИИ...	4
1.1. Понятие модели, типы моделей и их использование.....	4
1.2. Идентификация как метод построения моделей.....	8
1.3. Структурная идентификация.....	14
1.4. Методы упрощения моделей.....	20
2. ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ МОДЕЛИРОВАНИЯ.....	27
2.1. Адекватность модели и объекта.....	27
2.2. Характеристика проблем и методов моделирования объектов.....	32
2.3. Этапы и принципы моделирования объектов и систем на ЭВМ.....	37
2.4. Особенности фиксации и обработки результатов моделирования.....	43
3. АЛГОРИТМИЗАЦИЯ ПРОЦЕДУР ИДЕНТИФИКАЦИИ ОБЪЕКТОВ УПРАВЛЕНИЯ.....	47
3.1. Постановка задачи идентификации объектов управления.....	47
3.2. Идентификация статики линейного детерминированного объекта.....	50
3.3. Идентификация статики нелинейного детерминированного объекта.....	55
3.4. Идентификация объекта моделирования при наличии случайных помех...64	
3.5. Построение регрессионной модели стохастического объекта.....	71
3.6. Идентификация динамических объектов.....	91
3.7. Идентификация закона распределения случайной величины	97
4. ПОСТРОЕНИЕ МОДЕЛЕЙ СТАТИКИ СЛОЖНЫХ МНОГОМЕРНЫХ ОБЪЕКТОВ.....	114
4.1. Выбор существенных переменных модели объекта моделирования.....	114
4.2. Оптимальное планирование эксперимента.....	120
4.2.1. Нормирование переменных модели	122
4.2.2. Полный факторный эксперимент.....	123
4.2.3. Вычисление коэффициентов модели.....	126
4.2.4. Оценка значимости коэффициентов модели.....	129
4.2.5. Проверка адекватности модели.....	130
4.2.6. Дробный факторный эксперимент	132
4.2.7. Учет влияния помех.....	136
4.2.8. Построение модели с квадратичными эффектами.....	139
4.3. Метод группового учета аргументов.....	142
4.4. Определение наилучшей модели.....	150
ЗАКЛЮЧЕНИЕ.....	155
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	156

Учебное издание

Бурковский Виктор Леонидович
Мурзинов Павел Валерьевич
Матвеевко Иван Михайлович

**МЕТОДЫ ИДЕНТИФИКАЦИИ ОБЪЕКТОВ
СИСТЕМ УПРАВЛЕНИЯ**

Учебное пособие

Редактор Ю. А. Ярцева

Подписано к изданию 19.04.2022.
Объем данных 2,9 Мб

ФГБОУ ВО «Воронежский государственный технический университет»
394006 Воронеж, ул. 20-летия Октября, 84