Планирование эксперимента

в исследовании технологических процессов

ИЗДАТЕЛЬСТВО
„МИР“


ИЗДАТЕЛЬСТВО

Statistische Versuchsplanung und -auswertung in der Stoffwirtschaft

Von einem Autorenkollektiv

Herausgeber:
Prof. Dr.-Ing. Klaus Hartmann
Dr.-Ing. Eduard Lezki
Prof. Dr. rer.nat. habil.
Wolfgang Schäfer

VEB Deutscher Verlag für Grundstoffindustrie
Leipzig 1974

Планирование эксперимента в исследовании технологических процессов

К. ХАРТМАН (ГДР)
Э. ЛЕЦКИЙ (СССР)
В. ШЕФЕР (ГДР)
и коллектив авторов

Перевод с немецкого
Г. А. ФОМИНА и И. С. ЛЕЦКОЙ

Под редакцией
канд. техн. наук Э. К. ЛЕЦКОГО

ИЗДАТЕЛЬСТВО «МИР»
Москва 1977

УДК 658.57:519

Теория эксперимента — быстро развивающаяся область прикладной математики, направленная на повышение эффективности экспериментальных исследований.

Настоящая книга, посвященная вопросам статистической обработки данных и планирования эксперимента в технологических исследованиях, представляет собой коллективный труд советских и немецких специалистов. Материал книги охватывает широкий круг экспериментальных задач. Основное внимание уделяется статистическим методам, применяемым при отборе факторов, построении математических моделей различного вида, экспериментальной оптимизации.

Книга рассчитана на научных работников и инженеров, интересующихся современными методами повышения эффективности экспериментальных исследований, особенно в области непрерывных технологических процессов. Она может быть полезной аспирантам и студентам вузов различного профиля.

К. Хартман и др.

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА В ИССЛЕДОВАНИИ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Редактор В. Я. Фридман
Художник В. В. Шульженко
Художественный редактор В. Висенгалиев
Технический редактор Г. В. Алюшина

Сдано в набор 7/VII 1976 г. Подписано к печати 20/XII 1976 г. Бум. тип. № 2 60×90 1/16=17,25 бум. л. 34,50 печ. л. Уч.-изд. л. 32,06. Изд. № 20/8003 Цена 2 р. 42 к. Заказ № 530

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени Первая Образцовая типография имени А. А. Жданова Союзполиграфпрома при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли. Москва, М-54, Валовая, 28

Редакция литературы по новой технике

© 1974, VEB Deutscher Verlag für Grundstoffindustrie, Leipzig, Deutsche Demokratische Republik

© Перевод на русский язык с изменениями, «Мир», 1977

П 20204-145 / 041 (01)-77 145-77

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Статистические методы планирования эксперимента позволяют значительно интенсифицировать труд исследователя, сократить сроки и затраты на эксперимент, повысить достоверность выводов по результатам исследования.

Проблемам теории эксперимента, так же как и вопросам внедрения методов статистической теории эксперимента в практическую деятельность исследователей, в СССР за последние годы уделяется большое внимание. Достаточно напомнить о Всесоюзных совещаниях, школах, семинарах по этим проблемам, регулярно проводимым в нашей стране. В этом направлении интенсивно работают научные коллективы при МГУ, МЭИ и ряде других организаций, во многих вузах страны читаются соответствующие курсы и т. д.

Рост интереса к теории эксперимента вполне закономерен и объясняется как повышением требований к качеству экспериментальных исследований, так и широкими возможностями реализации часто трудоемких процедур планирования эксперимента и обработки его результатов, появившимися в связи с распространением средств вычислительной техники.

Развитие науки об эксперименте изменило сам характер подхода исследователя к эксперименту. Согласно новому подходу, эксперимент следует *проектировать*, стремясь обеспечить требуемое *качество исследования*. Цели и критерии оценки качества исследования необходимо выявлять, *системно* оценивая затраты, важность ожидаемых результатов, последствия недостаточной точности и ошибочных решений. Для вычисления показателей качества исследовательского процесса нужно уметь строить *модель* эксперимента (и убеждаться в ее правильности!). Решение всех этих задач базируется на результатах теории эксперимента, которая получила в последние годы мощные импульсы для своего развития благодаря успехам вычислительной техники и появлению автоматизированных систем обработки данных. Возможность обработки данных в реальном масштабе времени привела, в частности, к развитию *алгоритмических* методов планирования эксперимента, основанных на последовательных процедурах обработки данных и планирования.

За последние годы на русском языке появился ряд книг отечественных и зарубежных авторов по статистической теории и методам планирования эксперимента, рассчитанных на читателей с различным уровнем математической подготовки [1]. Предлагаемая вниманию читателей книга, представляющая собой коллективный труд советских и немецких авторов и первоначально изданная в 1974 г. в ГДР на немецком языке, явится хорошим дополнением к этим изданиям. В ней на достаточно строгой математической основе изложен широкий круг современных методов планирования эксперимента и статистической обработки данных, причем особый упор делается на технологические исследования в отраслях с непрерывным характером производства.

Чтение книги предполагает у читателя знакомство с математическим аппаратом теории вероятностей и линейной алгебры в объеме соответствующих курсов технических вузов.

В книге нашли отражение как традиционные методы планирования эксперимента, ориентированные на использование каталогов планов, так и алгоритмические методы (в задачах построения линейных и нелинейных моделей, в задачах классификации), связанные с применением средств вычислительной техники. Она охватывает достаточно широкий круг экспериментальных задач, хотя некоторые важные проблемы, такие, например, как дискриминация гипотез о виде модели, специальные задачи исследования моделей типа "состав—свойство" и т. п., все-таки остались за пределами рассмотрения.

При подготовке русского варианта книги в ее текст с согласия авторов были внесены некоторые изменения. Так, оказалось целесообразным исключить из книги раздел, посвященный методам автоматической оптимизации, поскольку эти вопросы более подробно рассмотрены в уже имеющихся на русском языке книгах [2]. Раздел 15 дополнен подразделом „Метод наименьших абсолютных отклонений и чебышевская аппроксимация", специально написанным В. Шефером для русского издания книги. Кроме того, были устранены отдельные неточности немецкого издания. Библиография к разделам пополнена списками дополнительной литературы.

Можно надеяться, что книга с интересом будет встречена советскими исследователями и окажет им помощь в практической работе.

Э. К. Лецкий

[1] См. списки основной и дополнительной литературы к разд. 1.

[2] См., например, „Теория автоматического управления", под ред. А. В. Нетушила, ч. 2, изд-во „Высшая школа", М., 1972; Александровский Н. М., Егоров С. В., Кузин Р. Е., „Адаптивные системы автоматического управления сложными технологическими объектами", изд-во „Энергия", М., 1973.

ПРЕДИСЛОВИЕ АВТОРОВ К НЕМЕЦКОМУ ИЗДАНИЮ

Статистические методы планирования эксперимента и обработки данных во все возрастающих масштабах используются в технологических и научных исследованиях. Опыт показывает, что с их помощью может быть существенно повышена эффективность исследований, базирующихся на экспериментальном подходе. В свою очередь результаты научных исследований создают основу для дальнейшего роста эффективности производства. Новые возможности для широкого внедрения методов статистической обработки данных и планирования эксперимента в практику исследований открываются в связи с распространением средств электронной обработки информации и автоматизированных систем управления.

В монографии предпринята попытка изложить для экспериментаторов, работающих в промышленности и в научно-исследовательских учреждениях, эффективные методы планирования эксперимента и статистической обработки данных. Авторы, длительное время работающие в области методологии экспериментальных исследований, а также обладающие опытом проведения практических исследований, связанных преимущественно с химической технологией, стремились данной работой восполнить пробел в научно-технической литературе, имеющейся по этим проблемам на немецком языке.

Монография является коллективной работой немецких и советских специалистов, представляющих Высшую техническую школу по химии „Карл Шорлеммер" в Лёйна-Мерзебурге, Высшую инженерную школу в Лейпциге, нефтехимический комбинат в г. Шведт и Московский энергетический институт.

Авторы выражают благодарность издательству, а также всем, кто помогал им в работе над книгой.

Все замечания и пожелания, направленные на улучшение книги, будут восприняты авторами с благодарностью.

Авторы

Авторский коллектив

К. Хартман, г. Мерзебург — разд. 1, 5, 7, 12.
Э. Лецкий, г. Москва — разд. 1, 4, 5, 6, 9, 11, 13.
В. Шефер, г. Лейпциг — разд. 2, 3, 5, 15.
Д. Бальцер, г. Шведт — разд. 14.
Н. Дамерт, г. Шведт — разд. 10.
Н. Дьякова, г. Москва — разд. 8.
Г. Фомин, г. Москва — разд. 13.
В. Каушус, г. Шведт — разд. 4, 7, 12.
Р. Кузин, г. Москва — разд. 14.

1.
ВВЕДЕНИЕ

Качественные и количественные изменения в производстве, науке и технике нашего времени, охватываемые понятием научно-технической революции, создают основу для значительного повышения производительности труда и эффективности всего народного хозяйства. Эти изменения возникают, однако, не сами по себе, а в результате сознательной, планомерной, научно обоснованной деятельности, охватывающей наряду с исследованиями, разработкой, проектированием и изготовлением изделий также запуск, поддержание, перестройку и рационализацию производственных процессов.

В эпоху научно-технической революции возрастает необходимость рационального использования в науке и технике рабочей силы — ученых, инженеров, рабочих — и средств производства — технической оснастки и оборудования. Одно из направлений повышения производительности научного труда заключается в применении современных математических методов и технических средств, таких, как планирование эксперимента, исследование операций, математическое моделирование, вычислительная техника.

Практическая полезность научных исследований в сильной степени зависит от методов их проведения и формы, в которой представляются результаты. Применение эффективной технологии исследований позволяет существенно сократить фазу внедрения, что приводит к экономии времени и средств.

Интенсификация действующих производственных процессов путем выявления всех резервов, нахождения и устранения узких мест, повышения качества продукции, а также экономии материалов и энергии является важным источником повышения производительности труда и эффективности народного хозяйства. Это касается как отдельных установок, так и производственных комплексов и систем.

Выявление резервов производства, как правило, связано с основательным анализом процесса и предполагает применение современных методов исследования и электронных систем обработки данных. В некоторых случаях для интенсификации высокопроизводительных процессов целесообразно использовать

специальные вычислительные машины, выполняющие функции автоматического управления и оптимизации режимов. Создание автоматизированных управляющих и регулирующих систем открывает новые возможности в этих областях. В некоторых отраслях производства, в частности в химической промышленности, производственные и исследовательские системы являются особенно сложными. Многие проблемы недостаточно проработаны теоретически, так что экспериментальные исследования на лабораторных, опытных и промышленных установках играют особую роль. На создание экспериментальных установок и на проведение экспериментов здесь приходится тратить много времени и средств.

С помощью традиционных методов не удается обеспечить требуемые темпы развития ни в фундаментальных или прикладных исследованиях, ни при модернизации или проектировании производственных установок. Поэтому в науке, технике и производстве при решении разнообразных задач во все больших масштабах применяются новые эффективные методы исследования. При этом особое внимание уделяется моделям процессов и способам их построения.

Для эффективного анализа механизма явлений и управления производственными процессами необходимо выявить взаимосвязи между факторами, определяющими ход процесса, и представить их в количественной форме — в виде математической модели. Математическая модель является математическим отображением наиболее существенных сторон процесса. Модель представляет собой совокупность уравнений, условий и алгоритмических правил и позволяет:

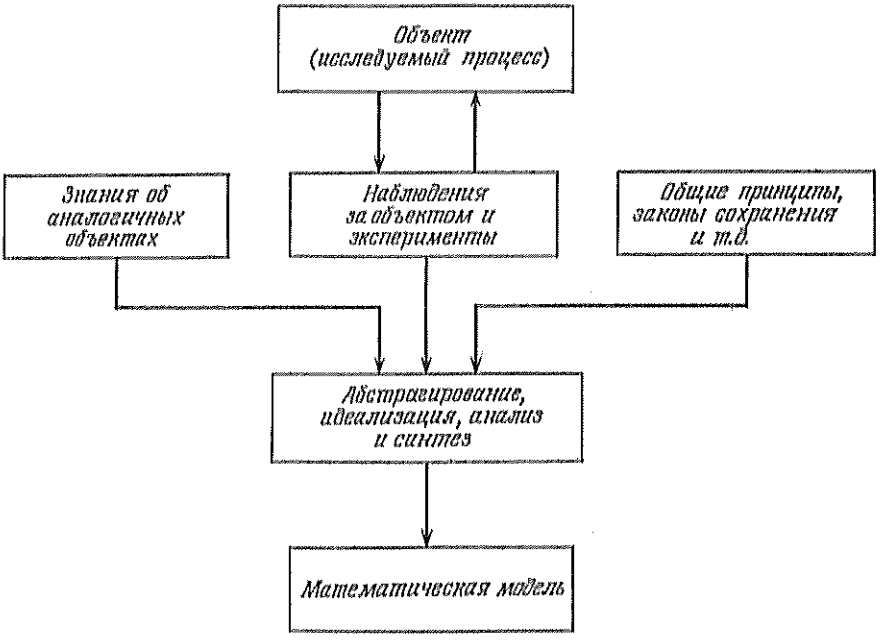
— получать информацию о процессах, протекающих в объекте;

— рассчитывать системы, т. е. анализировать и проектировать их;

— получать информацию, которая может быть использована для управления моделируемым объектом.

В зависимости от источника информации, используемого при построении математической модели (фиг. 1.1), различают физико-химические модели (называемые также иногда аналитическими или теоретическими) и статистические или эмпирические модели. В первом случае за основу берутся физико-химические закономерности моделируемых процессов, например в виде уравнений баланса или кинетических уравнений для превращений вещества. Построение теоретических моделей сопряжено с проведением обширных и длительных исследований, поскольку при этом необходимо выяснить природу микропроцессов, протекающих в объекте, и описать их математически. Как правило, модели процессов представляются в виде сложных систем уравнений (системы алгебраических, обыкновенных дифференциальных уравнений или уравнений в частных производных). Они позволяют очень

точно описать процессы, протекающие в объекте, и допускают экстраполяцию в точки факторного пространства, в которых невозможно непосредственное наблюдение этих процессов. Статистические модели получаются в результате статистической обработки экспериментальных данных, собранных на исследуемом объекте. Структура статистической модели может выбираться



Фиг. 1.1. Источники информации для построения математических моделей.

относительно произвольно. Соответствие модели объекту ограничивается лишь количественным аспектом.

Статистические модели имеют относительно простую структуру, очень часто они представляются в виде полиномов. Область их применения ограничивается ближайшей окрестностью рабочих точек, в которой проводятся эксперименты. Во многих случаях построение таких моделей можно выполнить ценой сравнительно небольших затрат времени и средств.

Принято различать статические (стационарные) и динамические модели (модели динамики). Первые из них описывают стационарные, т. е. не изменяющиеся во времени, соотношения в объекте, вторые—переходные процессы, т. е. нестационарные состояния. Модели динамики чаще всего имеют вид передаточных функций или систем дифференциальных уравнений. И те и дру-

гие модели могут относиться либо к статистическому, либо к физико-химическому типу.

После установления структуры модели необходимо численно оценить по экспериментальным данным свободные параметры модели. Смотря по тому, как эти параметры входят в модель, говорят о линейных или нелинейных по параметрам моделях. Кроме того, различают модели линейные или нелинейные по независимым переменным.

Можно выделить следующие этапы решения задачи построения модели:

1. Постановка задачи

Содержанием этого этапа является определение цели, выяснение исходной ситуации, оценка допустимых затрат времени и средств, а также установление типа задачи:

— выявление структуры или параметров;
— управление объектом;
— нахождение оптимальных условий.

2. Сбор априорной информации

Путем изучения литературы, опроса экспертов и т. д. необходимо собрать и оценить возможно более полным образом всю информацию, касающуюся решения таких же или сходных задач и имеющую значение для выбора дальнейшей стратегии.

3. Выбор способа решения и стратегии его реализации

На этом важном этапе устанавливается тип модели — статистический или физико-химический — и выявляются возможные влияющие переменные (факторы) и выходные переменные (отклики, целевые величины). При этом структура модели существенным образом зависит от цели ее построения. Например, для задачи проектирования имеет смысл рассматривать лишь теоретические модели. Одновременно формулируются статистические задачи (выбирается метод оценивания и постулируются гипотезы). Строится установка, необходимая для проведения экспериментов, и разрабатывается методика анализа их результатов. Составляются и отлаживаются алгоритмы и программы, требуемые для обработки экспериментальных данных. Здесь же необходимо решить, в каком режиме целесообразно применять непосредственно связанные с установкой устройства обработки данных.

4. Проверка выбранного способа решения

Для своевременного выявления и устранения возможных ошибок в постановке задачи, выбранной модели, экспериментальной установке, методике анализа и т. д., а также для экономии времени и средств проводятся предварительные эксперименты. С их помощью не только проверяются экспериментальная установка и методика, но и производится предварительная оценка качества модели.

5. Реализация выбранного способа решения

Реализация проверенной и скорректированной методики решения заключается в окончательном установлении целевых величин и факторов, объема выборок и плана эксперимента, кратности повторения опытов, а также в проведении экспериментов и статистической обработке их результатов.

6. Анализ и интерпретация результатов

На этом заключительном этапе производится статистический анализ результатов эксперимента и проверяются точность оценок параметров и работоспособность полученных моделей. Затем результаты интерпретируются с точки зрения их физико-химического или технико-экономического содержания. В том случае, если не достигнута цель исследования, необходимо проанализировать причины и обсудить новые варианты решения.

Одной из важнейших целей при разработке и запуске процессов с непрерывным характером производства является определение и поддержание оптимального режима процесса. Оптимизацию вновь создаваемых установок можно выполнить лишь с помощью описанных выше математических моделей, адекватно отражающих процесс. Для этого могут применяться различные методы оптимизации. Определение оптимальных параметров установок или процессов по сложным, многомерным, нелинейным моделям связано со значительными вычислительными трудностями. При оптимизации действующих установок находят применение модели самых различных типов. Кроме того, имеются методы, позволяющие найти оптимальные режимы процессов без моделей. Поскольку характеристики многих технологических процессов дрейфуют, т. е. изменяются во времени, то параметры моделей, а в некоторых случаях и их структура должны корректироваться и приспосабливаться к новой производственной ситуации.

Во многих случаях оптимальное управление производственными установками экономически целесообразно производить с помощью управляющих вычислительных машин. Это управле-

ние выполняется на основе соответствующих моделей и алгоритмов. Применение управляющих ЭВМ позволяет также автоматически подстраивать модели в условиях дрейфа.

Все упоминавшиеся до сих пор задачи и проблемы не разрешимы без проведения экспериментов. Эксперимент занимает главенствующее место среди способов получения информации о внутренних взаимосвязях явлений в природе и технике. Он является отправной точкой и критерием большинства наших знаний. Эксперименты и наблюдения явились основой для открытия большинства известных нам законов природы и для проверки теоретических гипотез. По мере роста сложности исследуемых процессов и явлений очень быстро возрастают также затраты на аппаратуру и проведение эксперимента. На экспериментальные установки, такие, например, как ускорители элементарных частиц в ядерной физике, затрачиваются огромные средства. Для проведения некоторых специальных экспериментов требуются количества энергии, которые были бы достаточны для энергоснабжения города средней величины. При этом постоянно возрастает сложность решаемых проблем, а большой объем информации, необходимой для выяснения внутренних взаимосвязей в природе и технике, заставляет применять электронные устройства обработки информации.

Значительная часть исследовательских усилий при разработке или модернизации технологических процессов определяется затратами на эксперименты. Разработка химических процессов в большинстве случаев включает следующие этапы: „лабораторная установка“ – „опытная установка“ – „промышленная экспериментальная установка“, хотя математические модели, построенные в лабораторных условиях, во многих случаях допускают экстраполяцию на промышленные установки.

Широкое применение экспериментальных методов привело к созданию теории эксперимента. Эта теория призвана дать экспериментатору ответы на следующие вопросы:

1. Как нужно организовать эксперимент, чтобы наилучшим образом решить поставленную задачу (в смысле затрат времени и средств или точности результатов)?

2. Как следует обрабатывать результаты эксперимента, чтобы получить максимальное количество информации об исследуемом явлении?

3. Какие обоснованные выводы можно сделать об исследуемом объекте по результатам эксперимента?

Основой теории эксперимента является математическая статистика, которая применима для анализа эксперимента в тех случаях, когда его результаты могут рассматриваться как случайные величины или случайные процессы. Это условие выполняется в большинстве исследований, поскольку, как правило, резуль-

таты эксперимента связаны с некоторой неопределенностью. Среди прочих причин такой неопределенности можно назвать случайный характер исследуемых процессов (например, при измерении физических явлений молекулярного характера с помощью макроскопических средств), влияние неконтролируемых факторов, неконтролируемые изменения условий эксперимента и ошибки наблюдений. Сюда же относятся измерительные ошибки, причины которых кроются в несовершенстве приборов, методов измерений и устройств передачи данных. Влияние этих возмущений на результат наблюдений может во многих случаях рассматриваться как случайное.

Отметим следующие новые возможности, которые дает исследователям теория эксперимента:

1. Статистическое представление об эксперименте образует основу для исследования сложных объектов и систем. Эти сложные объекты характеризуются большим числом факторов, воздействующих на результаты эксперимента. При классическом подходе к экспериментам исследование влияния совокупности факторов на результаты эксперимента производилось при условии, что изменяется только один из факторов и фиксируются значения всех остальных. В сложных системах, в которых большое число воздействий не может контролироваться или управляться, это условие не выполняется. Статистическая концепция учитывает влияние неконтролируемых факторов иным образом. Воздействие этих факторов рассматривается как дополнительный стохастический шум, наложенный на истинные результаты экспериментов. Для того чтобы воздействие сделать случайным, применяют специальные методы, называемые рандомизацией эксперимента [1, 2]. Благодаря этому удается надежно отделить факторы, интересующие экспериментатора, от шумового фона, обусловленного неконтролируемыми воздействиями.

2. Математическая статистика предоставляет в распоряжение экспериментатора методы анализа данных и принятия решений относительно исследуемого объекта на основании обработанных результатов эксперимента. Эти методы учитывают стохастический характер результатов и основываются на статистической проверке гипотез.

Идею рандомизации и статистической интерпретации результатов экспериментов можно пояснить на следующем простом примере.

Пусть необходимо сравнить качество конечного продукта двух различных производственных установок. Для этого исследуется некоторое число проб продукта с обеих установок и сравниваются средние значения соответствующих характеристик. Однако на качество продукта, помимо свойств самих установок, оказывают влияние также и другие факторы.

Предположим, например, что на качество продукта влияет температура сырья. При классическом методе проведения экспериментов в этом случае требуется, чтобы температура сырья была одной и той же для всех проб обеих установок. Но это требование, как правило, бывает невыполнимым, например, если исходные вещества для установок подаются из разных емкостей. Статистический учет влияния температуры делает необходимой рандомизацию эксперимента относительно этого фактора. Это означает, что средние характеристики по каждой из установок должны вычисляться по пробам, взятым при различных температурах. Изменение температуры при этом приводит к разбросу измеренных характеристик качества около некоторого среднего значения. Предположим, что выборочное качество конечного продукта на установке 1 выше, чем на установке 2. Можно спросить, является ли этот факт достаточным для того, чтобы принять решение в пользу первой установки? Ответ будет отрицательным. Принимая такое решение, необходимо учитывать случайный характер и параметры распределения вероятности полученных оценок. Решение должно быть найдено в результате проверки статистической гипотезы о равенстве математических ожиданий двух случайных величин по их выборочным оценкам. Лишь в том случае, если оценки средних значений показателей качества статистически значимо отличаются друг от друга, можно сделать вывод, что первая установка работает лучше второй.

В отношении статистических выводов необходимо всегда иметь в виду следующее обстоятельство: положительные результаты проверки некоторой статистической гипотезы означают лишь то, что постулируемая гипотеза не противоречит результатам эксперимента. Результаты проверки гипотезы никогда не могут служить доказательством абсолютной справедливости и правильности гипотезы. В приведенном выше примере результаты эксперимента могут в той же мере соответствовать гипотезе о равенстве средних значений, как и гипотезе о различии средних значений на некоторую величину.

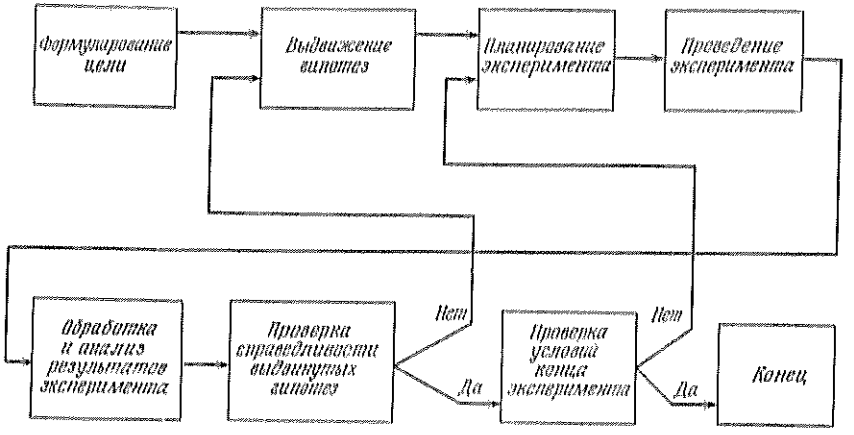
Аналогичная ситуация возникает при построении математических моделей процессов. Исследователь может постулировать целый ряд гипотез о виде модели и экспериментальным путем выбрать среди этих моделей ту, которая наилучшим образом соответствует результатам экспериментов. Эксперименты здесь выступают в роли третейского судьи, принимающего одно из многих возможных решений. Если среди постулированных гипотез не оказалось истинной, то и с помощью экспериментов ее также не удастся найти. Однако это ограничение экспериментального подхода ни в коем случае не принижает роли эксперимента при решении практических задач.

3. Большое значение имеет концепция оптимального использования области варьирования условий эксперимента при проведении исследования с целью построения математической модели. Как уже отмечалось выше, построение математической модели по экспериментальным данным выполняется следующим образом: постулируется вид модели; проводятся эксперименты в некотором числе N точек; на основании результатов экспериментов рассчитываются оценки неизвестных параметров модели и проверяется правильность постулированной модели. Статистический анализ характеристик экспериментальных моделей показывает, что точность модели существенным образом зависит от выбора условий проведения экспериментов, т. е. является функцией плана экспериментов. Это означает, что с помощью целенаправленного выбора условий проведения эксперимента из области возможных значений факторов, т. е. с помощью планирования эксперимента, требуемая точность решения может быть достигнута при минимальном числе экспериментов. Отсюда следует, что с помощью планирования во многих случаях могут быть снижены затраты времени и средств и повышена эффективность эксперимента. Целенаправленный выбор условий проведения экспериментов предполагает возможность активного воздействия на исследуемый объект, т. е. возможность активного эксперимента. Эта возможность обычно имеется при лабораторных исследованиях. На промышленных установках активный эксперимент применим не всегда, так как там часто имеется некоторое число факторов, изменять которые по желанию экспериментатора оказывается невозможно (например, состав сырья, содержащиеся в нем примеси и т. д. являются неуправляемыми факторами). В этих случаях применяют пассивный подход к сбору информации, т. е. регистрацию технологических факторов при нормальном функционировании объекта без целенаправленных возмущений.

Преимущества активного эксперимента, позволяющего применять наиболее целесообразно составленные планы, достаточно очевидны. Кроме всего прочего, при активном эксперименте можно оценить дисперсию ошибки, строго проверить адекватность модели и принять необходимые меры для выполнения условий, необходимых для применения метода множественного регрессионного анализа, используемого для обработки результатов экспериментов. Концепция оптимального (с точки зрения определенных характеристик точности решения) выбора условий проведения экспериментов имеет основополагающее значение в теории планирования эксперимента [3]. Понятие планирования эксперимента можно, впрочем, интерпретировать и в более широком смысле, подразумевая под этим также проблему оптимальной организации пассивного сбора информации (выбора интервала

времени между моментами регистрации данных, задания объема выборки и т. д.) [4].

Развитие статистических методов планирования эксперимента связаны с именем Р. А. Фишера. В 1935 г. Фишер опубликовал монографию „Планирование эксперимента“ („Design of Experiments“), давшую название новому направлению исследований. Работы Фишера были связаны с практическими проблемами агротехнических исследований. Планирование эксперимента применялось им в задачах с дискретными факторами, решаемых



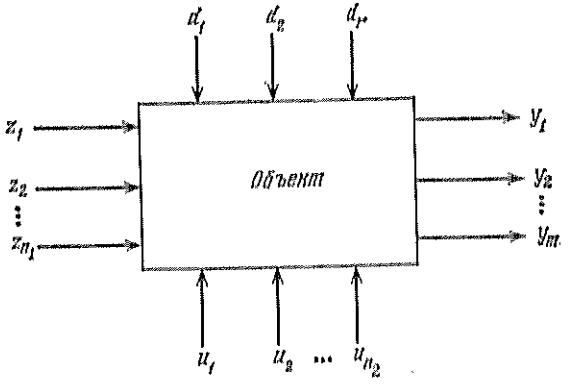
Фиг. 1.2. Стратегия построения математической модели.

с помощью аппарата дисперсионного анализа. Современное состояние проблемы планирования подобных экспериментов анализируется, например, в монографии [2]. В данной книге эти вопросы будут затронуты очень кратко.

Рассматриваемые в этой книге методы планирования эксперимента относятся к задачам исследования «поверхности отклика» и изучения механизма явлений. Эти методы получили особенно широкое распространение после 1951 г., когда появилась работа Бокса и Уилсона [5]. Развитие современных идей планирования эксперимента для оценивания параметров регрессионных уравнений и констант теоретических моделей связано с именем американского математика Кифера (см. [3]).

4. Теория эксперимента дает исследователю точную логическую схему и способ решения задач на разных этапах исследования. Каждое экспериментальное исследование состоит из ряда следующих друг за другом этапов (фиг. 1.2): формулирование цели, выдвижение гипотезы об исследуемом объекте, планирование экспериментов, проведение экспериментов, обработка и анализ результатов,

проверка правильности выдвинутой гипотезы, выдвижение новой гипотезы, проверка условий окончания эксперимента, планирование нового эксперимента. Из этой схемы ясно, что исследование объекта состоит из повторяющихся циклов, причем от цикла к циклу растет объем знаний об объекте, так что можно предположить, что выдвигаемые гипотезы все более приближаются к действительности. Вместе с тем возрастает также эффективность планирования эксперимента и всего исследования. Здесь прослеживается связь стратегии эксперимента с концепцией последова-



Фиг. 1.3. Схематическое представление объекта.

тельного анализа, разработанной в математической статистике [6]. Эта концепция практически воплощена в целом ряде методов, нашедших широкое применение при решении различных задач (экспериментальная оптимизация по методу Бокса—Уилсона, последовательное уточнение констант теоретических моделей, ЭВОП и т. п.). Новые возможности для распространения последовательных методов при исследовании промышленных процессов открылись благодаря внедрению вычислительных машин.

Область применения методов планирования эксперимента весьма обширна. Эти методы оказываются очень эффективными при проведении физико-химических исследований в лабораторных условиях, в экспериментах на опытных, полупромышленных и промышленных установках. Объектами исследований могут быть рецептуры смесей, аппараты и процессы химической технологии и других промышленных отраслей [7].

Все переменные, определяющие состояние объекта, независимо от их физико-химической сущности можно разделить на четыре группы (фиг. 1.3).

1. Группа $Z=(z_1, \ldots, z_{n_1})$. В эту группу входят факторы, которые характеризуют качество сырья или промежуточных про-

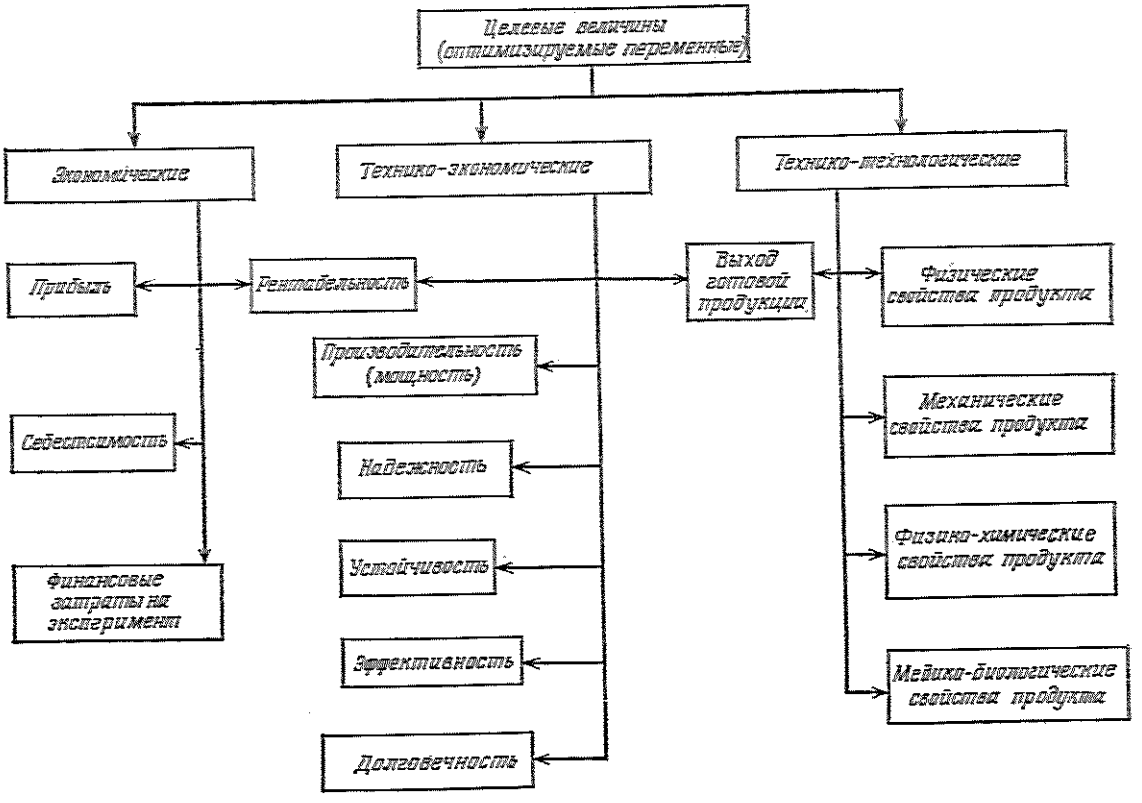
дуктов и не допускают целенаправленного изменения в ходе исследования. К входным и промежуточным продуктам можно отнести исходные вещества или продукты предыдущих звеньев технологической цепи. Информация о значениях переменных этой группы получается в результате лабораторных анализов, измерений и т. п.

2. Группа $U = (u_1, \ldots, u_{n_2})$. Эту группу образуют управляемые факторы процесса. С их помощью реализуется заданный технологический режим. К ним относятся показания расходомеров, положения уставок регуляторов и т. п. На значения управляемых факторов накладываются технологические ограничения, т. е. ограничивается область их допустимых значений. Переменные групп Z и U часто будут объединяться нами в группу $X = (x_1, \ldots, x_n)$ и называться контролируемыми входными или независимыми переменными процесса.

3. Переменные группы $Y = (y_1, \ldots, y_m)$ часто называют выходными. К их числу относятся величины, характеризующие экономическую эффективность процесса, а также технико-экономические параметры, технологические свойства и характеристики готовых продуктов (фиг. 1.4). Часто задача состоит в том, чтобы определить зависимость между входными и выходными переменными процесса или найти значения входных факторов, обеспечивающие достижение экстремума некоторой выходной переменной (возможно, при каких-то ограничениях). Это означает, что переменные этой группы часто выступают в качестве целевых величин при оптимизации процессов. Выходные переменные в некоторых случаях могут быть качественными и принимать лишь конечное число дискретных значений (например, устойчивость или неустойчивость какого-либо процесса).

4. Четвертую группу $D = (d_1, \ldots, d_r)$ образуют неконтролируемые факторы. Они характеризуют действующие на объект возмущения, которые не могут быть измерены количественно (неконтролируемые примеси в сырье, старение катализатора и т. д.). Воздействие неконтролируемых факторов, медленно изменяющихся во времени, приводит к дрейфу характеристик объекта.

Цель настоящей книги состоит в том, чтобы дать описание эффективных методов экспериментальных исследований, которые оказываются полезными при решении практических задач как в лабораторных, так и в промышленных исследованиях. В этой книге не только излагаются сами методы, но и поясняется их математическое содержание. Понимание математического содержания методов является необходимой предпосылкой их правильного применения. Все приводимые здесь методы иллюстрируются подробными примерами из практики экспериментальных исследований в области химии и химической технологии. Книга



Фиг. 1.4. Классификация целевых величин.

состоит из четырнадцати разделов и математического приложения (разд. 15).

В разд. 2 описаны основные понятия математической статистики: случайные выборки, оценки и их свойства, методы проверки гипотез и построения доверительных интервалов для математического ожидания и дисперсии. Кроме того, там же рассматриваются процедуры проверки гипотез о параметрах распределения случайных величин.

В разд. 3 рассматривается математический аппарат множественного регрессионного анализа. Этот метод является основой для расчета оценок параметров при построении модели процесса. Изложенный здесь материал применяется в последующих разделах.

В разд. 4 приведены некоторые процедуры выбора функций из заданного их множества, предназначенных для построения математической модели. Бо́льшая часть этого раздела посвящена описанию метода построения математической модели с помощью шаговой стратегии выбора функций, образующих ортогональную систему для заданного плана эксперимента.

В разд. 5 рассматриваются вопросы оптимального планирования эксперимента для построения математической модели в виде линейной комбинации линейных и квадратических функций от входных факторов с неизвестными параметрами. Там же вводятся основные понятия теории планирования эксперимента и обсуждаются критерии планирования. В этом же разделе описаны полные и дробные факторные планы, а также композиционные ортогональные и ротатабельные планы эксперимента для квадратичных моделей.

В разд. 6 изучаются планы эксперимента, оптимальные с точки зрения критерия D-оптимальности. В этом разделе изложены новые результаты теории планирования эксперимента и пути их практического использования. Здесь вводятся понятия точных (дискретных) и непрерывных планов, рассматриваются методы построения непрерывных D-оптимальных планов, приводятся и обсуждаются результаты исследования планов из разд. 5 с точки зрения D-оптимальности.

В разд. 7 дано описание методов выделения существенных факторов, которые необходимо учитывать при построении математических моделей. Эти методы имеют большое значение на этапе предварительного исследования, на котором из совокупности возможных факторов нужно выделить подмножество переменных, существенных для дальнейшего экспериментального исследования с целью построения модели процесса. Рассматриваются методы, основанные на применении насыщенных планов, а также метод случайного баланса и дисперсионный анализ, предназначенный для исследования дискретных факторов.

Раздел 8 является логическим продолжением разд. 7 и содержит описание метода решения задач разд. 7, основанного на использовании так называемых „психологических экспериментов". Для выделения существенных факторов в этом случае применяется статистическая обработка результатов опроса экспертов. Часто такой метод решения позволяет получить требуемые результаты без проведения физических экспериментов на объекте. В общем случае решение ищется на основе априорной информации, полученной от экспертов. Поэтому описанные здесь методы называют также методами априорного моделирования. В данном разделе нашли свое отражение лишь некоторые методы: ранговая корреляция, метод парных сравнений, конкордация. Это научное направление в настоящее время развивается весьма интенсивно.

Раздел 9 содержит методы построения моделей процессов в условиях неконтролируемого дрейфа характеристик объектов. Такой дрейф является одной из особенностей технологических процессов. Иногда влияние дрейфа на результаты экспериментов удается устранить с помощью специальных методов планирования эксперимента (планы, ортогональные к дрейфу). В других случаях, когда дрейф проявляется в изменении всех параметров модели (а не только свободного члена), требуется производить непрерывное отслеживание дрейфующих параметров, т. е. непрерывно корректировать модель. В этом разделе обсуждаются методы решения таких задач и способы установления факта отсутствия дрейфа.

Раздел 10 посвящен проблемам планирования эксперимента и расчета оценок параметров, входящих в модели нелинейно. С такими проблемами приходится сталкиваться, когда структура модели выбирается из физико-химических соображений (теоретические модели) и эксперименты применяются лишь для уточнения параметров теоретической модели. Применение теории планирования эксперимента в таких задачах основывается на линеаризации модели по неизвестным параметрам при некоторых начальных значениях этих параметров. Решение задач планирования и расчета оценок параметров предполагает использование вычислительных машин. В этом разделе рассматриваются практические пути решения указанных задач и описывается программа для ЭВМ, предназначенная для расчета оценок параметров и статистического анализа результатов.

В разд. 11 обсуждается последовательная методика расчета оценок параметров модели заданного вида. Здесь приводятся рекуррентные алгоритмы вычисления элементов дисперсионной матрицы, оценок параметров и остаточной суммы квадратов при добавлении новых экспериментальных точек. Кроме того, рассматривается метод последовательного планирования экспери-

мента для моделей, линейных и нелинейных по параметрам. Описывается также правило прекращения эксперимента при достижении требуемой точности оценок параметров.

В разд. 12 дано описание методов экспериментального поиска значений факторов, при которых целевая функция достигает экстремума. Здесь рассматриваются метод Бокса—Уилсона, один из вариантов метода случайного поиска, а также методы поиска экстремума в промышленных условиях (последовательный симплексный метод, ЭВОП). Далее приводятся методы анализа стационарной области, описываемой квадратичным полиномом, позволяющие установить характер поверхности отклика (каноническое преобразование и гребневый анализ).

Содержанием разд. 13 являются экспериментальные методы решения задач классификации. Эти методы применяются в тех случаях, когда необходимо найти модель объекта, у которого выходные переменные имеют качественный характер или могут принимать лишь небольшое число дискретных значений. Такими переменными являются, например, устойчивость процесса, работоспособность установки и т. п. В этом разделе описаны различные постановки задач (статистические и детерминированные) и этапы решения (выбор существенных признаков, планирование эксперимента, нахождение классифицирующего правила, оценка характеристик найденного правила).

В разд. 14 рассмотрено построение моделей динамических объектов. В основу излагаемого метода положена подача на объект входных возмущений в виде псевдослучайных двоичных последовательностей конечной длины. Этот способ называется активным корреляционным методом. Весовая функция объекта представляется в виде линейной комбинации функций Лагерра, параметры которой находятся методом наименьших квадратов. В этом разделе затронуты также практические проблемы, возникающие при использовании такого метода (выбор параметров функций Лагерра, определение тактовой частоты входного сигнала и длины входной последовательности).

Знание динамической модели процесса является необходимым условием для создания автоматических регуляторов, предназначенных для стабилизации различных характеристик объектов и для управления динамическими процессами (например, переходными процессами при запуске или включении установки). Здесь не рассматриваются проблемы динамической оптимизации, которые могли бы стать предметом отдельной монографии.

В разд. 15 приведены специальные распределения вероятности, которые используются при обработке экспериментальных данных. Кроме того, в этом разделе описан метод наименьших абсолютных отклонений применительно к задачам построения математических моделей.

С рядом проблем планирования эксперимента, не охваченных данной книгой, читатель может ознакомиться по источникам, указанным в списке литературы в конце этого раздела.

ЛИТЕРАТУРА

1. Налимов В. В., Теория эксперимента, изд-во „Наука", М., 1971.
2. Hicks C. R., Fundamental Concepts in the Design of Experiments, Holt, Rinfart a Winston, New York, 1964; русский перевод: Хикс Ч., Основные принципы планирования эксперимента, изд-во „Мир", М., 1967.
3. Федоров В. В., Теория оптимального эксперимента, изд-во „Наука", М., 1971.
4. Бородюк В. П., Лецкий Э. К., Статистическое описание промышленных объектов, изд-во „Энергия", М., 1971.
5. Box G. E. P., Wilson K. M., On the Experimental Attainment of Optimal Conditions, *J. Roy. Statist. Soc.*, Ser. B, 13, № 1 (1951).
6. Wald A., Sequential Analysis, London, 1947; русский перевод: Вальд А., Последовательный анализ, „Физматгиз", М., 1960.
7. Адлер Ю. П., Грановский Ю. В., Обзор прикладных работ по планированию эксперимента, Препринт 33, изд-во МГУ, М., 1972.
8. Bandemer H. u.a., Optimale Versuchsplanung, Akademie-Verlag, Berlin, 1973.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Кафаров В. В., Методы кибернетики в химии и химической технологии, изд-во „Химия", М., 1968.
2. Налимов В. В., Чернова Н. А., Статистические методы планирования экстремальных экспериментов, изд-во „Наука", М., 1965.
3. Химмельблау Д., Анализ процессов статистическими методами, изд-во „Мир", М., 1973.
4. Адлер Ю. П., Введение в планирование эксперимента, изд-во „Металлургия", М., 1969.
5. Маркова Е. В., Лисенков А. Н., Планирование эксперимента в условиях неоднородностей, изд-во „Наука", М., 1973.
6. Новые идеи в планировании эксперимента, под ред. В. В. Налимова, изд-во „Наука", М., 1969.
7. Планирование эксперимента, под ред. Г. К. Круга, изд-во „Наука", М., 1966.
8. Проблемы планирования эксперимента, под ред. Г. К. Круга, изд-во „Наука", М., 1969.
9. Планирование и автоматизация эксперимента в научных исследованиях, под ред. Г.К. Круга, изд-во „Советское радио", М., 1974.
10. Вознесенский В. А., Статистические методы планирования эксперимента в технико-экономических исследованиях, изд-во „Статистика", М., 1974.
11. Финни Д., Введение в теорию планирования эксперимента, изд-во „Наука", М., 1970.
12. Планирование эксперимента (библиографические указатели). Сост. Е. Г. Симонова, „Изд-во Гос. б-ки СССР им. В. И. Ленина", 1969, 1970 и 1972 гг.
13. Грановский Ю. В., Мурашов Т. И., Страхов А. В., Адлер Ю. П., Планирование эксперимента, Библиография прикладных работ за 1966—1968 гг., „Изд-во МГУ", 1971.

2.

ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ ОПРЕДЕЛЕНИЯ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК И ПРОВЕРКИ ГИПОТЕЗ

2.1. О задачах математической статистики

Реальные процессы химической технологии часто можно рассматривать как случайные и анализировать их с помощью вероятностных методов. Исследуемый процесс при этом считается реализацией некоторой случайной переменной X, и для его описания необходимо знать функцию $F(x)$ распределения вероятностей, функцию плотности распределения $f(x)$, или математическое ожидание, дисперсию и другие моменты распределения переменной X. Однако обычно на практике вид и параметры функции распределения точно не известны и информация о характеристиках случайной величины должна быть получена с помощью эксперимента.

Задачу математической статистики в общем виде можно сформулировать следующим образом: изучение случайных явлений, когда отсутствующая информация о распределениях вероятностей случайных величин, порождающих эти явления, может быть получена из эксперимента.

Пусть исход эксперимента (опыта, наблюдения) представляется некоторой случайной переменной X. Полагаем, что эксперимент может многократно повторяться. Реализация n экспериментов дает конкретный ряд значений $x_1, \ldots, x_n$. При этом говорят, что получена реализация $(x_1, x_2, \ldots, x_n)$ выборки объема n из заданной генеральной совокупности. На основе этой выборки необходимо построить оценки определенных параметров распределения случайной переменной X. Так, значение

$$\hat{m} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

может использоваться в качестве оценки математического ожидания m случайной переменной X.

Можно сказать, что математическая статистика среди других проблем изучает вопрос о том, насколько хорошо некоторый параметр a, истинное значение $\bar{a}$ которого не известно, может быть оценен с помощью реализации определенной выборочной функции

$$\hat{a} = \psi(x_1, x_2, \ldots, x_n).$$

Ряд вопросов подобного типа, относящихся к статистической оценке параметров распределения, рассмотрен ниже в разд. 2.3. Для получения практически полезного ответа на вопрос о свойствах оценок конкретные экспериментальные результаты x_i рассматриваются как реализации случайной переменной X. Вводится в рассмотрение вектор $(X_1, X_2, \ldots, X_n)$ случайных величин с одинаковыми распределениями, совпадающими с распределением переменной X, который представляет собой выборку, а также выборочная функция $\ddot{a} = \psi(X_1, X_2, \ldots, X_n)$, свойства которой должны быть исследованы. При этом $(x_1, x_2, \ldots, x_n)$ есть конкретная реализация выборки, а $\psi(x_1, x_2, \ldots, x_n)$ — реализация выборочной функции [1]). Строгое определение понятий „выборка", „выборочная функция" и некоторых других, важных для всех последующих разделов, будет дано в разд. 2.2.

При исследовании случайных явлений часто высказываются некоторые предположения, другими словами, формулируются некоторые гипотезы H. Такими гипотезами могут быть, например, следующие утверждения:

1. Некоторая случайная величина распределена по нормальному закону.
2. Случайная переменная имеет математическое ожидание m и (или) дисперсию σ^2.
3. Две случайные величины X и Y имеют одно и то же математическое ожидание m и (или) одинаковую дисперсию σ^2.

Вопросы проверки этих гипотез на основе выборочных данных излагаются в разд. 2.4.

В рамках этой главы, естественно, невозможно в полном объеме охватить все затронутые выше вопросы. Поэтому ниже задачи и методы их решения рассмотрены только с той степенью детализации, которая необходима для понимания дальнейших разделов книги. Это представляется целесообразным и по той причине, что обсуждаемые в гл. 2 методы и их практическое применение достаточно хорошо освещены в соответствующей литературе.

Один из важнейших разделов математической статистики — регрессионный анализ — излагается достаточно подробно в разд. 3.

2.2. Выборки и выборочные функции

В этом разделе мы дадим строгие определения введенным выше понятиям.

---

[1]) Здесь и ниже термины „реализация выборки" и „конкретная выборка", а также „реализация выборочной функции" и „конкретное значение выборочной функции" имеют один и тот же смысл. — *Прим. ред.*

Определение 2.1

Совокупность $(X_1, X_2, \ldots, X_n)$ n независимых случайных величин X_i, распределенных по одному и тому же закону, совпадающему с законом распределения случайной величины X, называется выборкой объема n из генеральной совокупности X (величины X_i представляют собой n экземпляров одной и той же случайной величины X).

Определение 2.2

Если $(X_1, X_2, \ldots, X_n)$ — выборка, согласно определению 2.1, то последовательность чисел $(x_1, x_2, \ldots, x_n)$ называется реализацией этой выборки.

Важнейшее значение для описания некоторой случайной переменной X имеет функция распределения вероятностей

$$F(x) = P\{X < x\}.$$

Если функция распределения не известна, можно на основе реализации $(x_1, x_2, \ldots, x_n)$ выборки объема n из генеральной совокупности X построить некоторое приближение к $F(x)$ с помощью функции

$$w_n(x) = \frac{n(x)}{n},$$

где $n(x)$ — число значений x_i, меньших, чем x. Эта функция представляет собой реализацию эмпирической функции распределения $W_n(x)$.

Приведем еще следующие определения:

Определение 2.3

Функция выборки $(X_1, X_2, \ldots, X_n)$ объема n

$$W_n(x) = \frac{N(x)}{n}, \tag{2.1}$$

где $N(x)$ — число тех X_i, для которых имеет место $X_i < x$, называется эмпирической функцией распределения выборки. Если $(x_1, x_2, \ldots, x_n)$ — реализация выборки, то функция

$$w_n(x) = \frac{n(x)}{n}, \tag{2.2}$$

где $n(x)$ — число значений x_i в рассматриваемой реализации, меньших x, называется реализацией эмпирической функции распределения $W_n(x)$.

В разд. 2.1 в связи с вопросами оценки параметров было введено понятие выборочной функции. Здесь мы дадим определение этого понятия.

Определение 2.4

Пусть $(X_1, X_2, \ldots, X_n)$ — некоторая выборка, а $(x_1, x_2, \ldots, x_n)$ — реализация этой выборки. Функция n переменных X_i

$$\tilde{a} = \psi(X_1, X_2, \ldots, X_n) \tag{2.3}$$

называется выборочной функцией, а ее конкретное значение для реализации выборки

$$\hat{a} = \psi(x_1, x_2, \ldots, x_n) \tag{2.4}$$

называется реализацией выборочной функции (2.3). Величина $\tilde{a}$ в (2.3) есть некоторая случайная переменная, а значение $\hat{a}$ в (2.4) представляет собой реализацию этой случайной переменной.

Рассмотрим теперь две специальные выборочные функции — среднее значение и дисперсию выборки — и их свойства.

Определение 2.5

Пусть $(X_1, X_2, \ldots, X_n)$ — некоторая выборка, а $(x_1, x_2, \ldots, x_n)$ — ее реализация. Тогда выборочная функция

$$\tilde{m} = \frac{X_1 + X_2 + \ldots + X_n}{n} \tag{2.5}$$

называется выборочным средним, а

$$\hat{m} = \frac{x_1 + x_2 + \ldots + x_n}{n} \tag{2.6}$$

реализацией выборочного среднего.

Связь между средним выборки и математическим ожиданием генеральной совокупности становится ясной из следующих утверждений:

Утверждение 2.1

Пусть выборка извлекается из генеральной совокупности случайной переменной X с математическим ожиданием

$$\overline{m} = E\{X\}. \tag{2.7}$$

Тогда

$$E\{\tilde{m}\} = \overline{m}. \tag{2.8}$$

Доказательство. Справедливость утверждения непосредственно вытекает из следующего:

$$E\{\tilde{m}\} = \frac{1}{n} E\left\{\sum_{i=1}^{n} X_i\right\} = \frac{1}{n} \sum_{i=1}^{n} E\{X_i\} = \frac{1}{n} \sum_{i=1}^{n} \overline{m} = \overline{m}.$$

Определение 2.6

Пусть $(X_1, X_2, \ldots, X_n)$ — некоторая выборка, а $(x_1, x_2, \ldots, x_n)$ — ее реализация. Тогда выборочная функция

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \tilde{m})^2 \tag{2.9}$$

называется выборочной дисперсией, а

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{m})^2 \tag{2.10}$$

— реализацией выборочной дисперсии.

Необходимость использования в знаменателе выражений (2.9) и (2.10) величины $(n-1)$, а не величины n следует из утверждения 2.2.

Утверждение 2.2

Пусть в определении 2.6 выборка извлекается из генеральной совокупности с математическим ожиданием $\bar{m}$ и дисперсией

$$\sigma^2 = D(X) = E\{(X - \bar{m})^2\}. \tag{2.11}$$

Тогда

$$E\{\tilde{s}^2\} = \sigma^2. \tag{2.12}$$

Доказательство. Имеем

$$\sum_{i=1}^{n} (X_i - \tilde{m})^2 = \sum_{i=1}^{n} (X_i - \bar{m} + \bar{m} - \tilde{m})^2 =$$

$$= \sum_{i=1}^{n} \{(X_i - \bar{m})^2 + (\bar{m} - \tilde{m})^2 + 2(X_i - \bar{m})(\bar{m} - \tilde{m})\} =$$

$$= \sum_{i=1}^{n} (X_i - \bar{m})^2 + n(\bar{m} - \tilde{m})^2 + 2(\bar{m} - \tilde{m}) \sum_{i=1}^{n} (X_i - \bar{m}) =$$

$$= \sum_{i=1}^{n} (X_i - \bar{m})^2 + n(\bar{m} - \tilde{m})^2 + 2(\bar{m} - \tilde{m})(n\tilde{m} - n\bar{m}) =$$

$$= \sum_{i=1}^{n} (X_i - \bar{m})^2 - n(\bar{m} - \tilde{m})^2.$$

В силу того, что

$$E\{(X_i - \bar{m})^2\} = \sigma^2,$$

$$E\{(\bar{m} - \tilde{m})^2\} = \frac{\sigma^2}{n},$$

получаем

$$E\left\{\sum_{i=1}^{n}(X_i-\tilde{m})^2\right\}=n\sigma^2-\sigma^2=(n-1)\sigma^2.$$

Следовательно,

$$E\{\tilde{s}^2\}=\sigma^2.$$

В случае если бы выборочная дисперсия определялась как

$$(\tilde{s}^*)^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\tilde{m})^2, \tag{2.13}$$

мы получили бы

$$E\{(\tilde{s}^*)^2\}=\frac{n-1}{n}\sigma^2. \tag{2.14}$$

Свойство (2.12), в силу которого выборочная функция (2.9) оказывается предпочтительнее функции (2.13) и которое в разд. 2.3 определено как свойство несмещенности оценок, непосредственно связано с понятием числа степеней свободы некоторой суммы квадратов. Пусть среди случайных переменных $Y_1, Y_2, \ldots, Y_n$ имеется ровно k попарно независимых, а остальные $(n-k)$ — линейно связаны [иными словами, существует $(n-k)$ линейных зависимостей, связывающих переменные Y_i]. Тогда говорят, что сумма квадратов

$$\sum_{i=1}^{n}Y_i^2$$

имеет k степеней свободы. В рассмотренном выше случае между случайными переменными $Y_i=X_i-\tilde{m}$ существует одна линейная связь (2.5). Поэтому в данном случае сумма квадратов связана не с n, а с $k=n-1$ степенями свободы.

Если сумму квадратов разделить на число степеней свободы, то полученная величина будет являться несмещенной оценкой дисперсии σ^2.

Вернемся еще раз к выборочной функции (2.5) и приведем без доказательства фундаментальную „центральную предельную теорему".

Утверждение 2.3 (центральная предельная теорема)

Арифметическое среднее некоторой выборки распределено асимптотически нормально с

$$E\{\tilde{m}\}=E\{X\}=\overline{m} \tag{2.15}$$

и

$$D(\tilde{m})=\frac{D(X)}{n}=\frac{\sigma^2}{n}. \tag{2.16}$$

Таким образом, случайная переменная $\tilde{m}$ в (2.5) имеет функцию распределения, которая при $n \to \infty$ стремится к нормальному закону с математическим ожиданием $\bar{m}$ и дисперсией σ^2/n вне зависимости от того, из какой генеральной совокупности была извлечена выборка. Имеет место и более сильное

Утверждение 2.4

Все моменты эмпирической функции распределения (2.1) имеют асимптотически нормальное распределение.

Этим свойством обладают также и другие симметричные выборочные функции вида (2.3), значения которых остаются неизменными при произвольном изменении порядка выборочных переменных. Такие симметричные выборочные функции имеют важное значение при статистической оценке параметров, так как оценки, естественно, должны зависеть только от значений, а не от порядка выборочных переменных.

Определение 2.7

Выборочная функция

$$\tilde{a} = \psi(X_1, X_2, \ldots, X_n) \tag{2.17}$$

называется симметричной, если

$$\tilde{a} = \psi(X_1, X_2, \ldots, X_n) = \psi(Y_1, Y_2, \ldots, Y_n), \tag{2.18}$$

где $(Y_1, Y_2, \ldots, Y_n)$ — произвольная перестановка элементов выборки $(X_1, X_2, \ldots, X_n)$.

2.3. Статистическая оценка параметров

Перейдем теперь к рассмотрению вопроса о том, как на основе некоторой выборки $(X_1, X_2, \ldots, X_n)$ найти оценку определенного параметра (например, математического ожидания или дисперсии) распределения случайной переменной X. Для лучшего понимания некоторых введенных выше обозначений приведем здесь их определения.

Определение 2.8

Истинное значение некоторого параметра a в дальнейшем обозначается через $\bar{a}$, его оценка с помощью некоторой выборочной функции — через $\tilde{a}$, а реализация выборочной функции — через $\hat{a}$ (см. разд. 2.2, определение 2.4).

Для оценки некоторого параметра a, естественно, может использоваться целый ряд различных выборочных функций, однако „хорошая" оценка должна удовлетворять определенным требованиям.

Свойства, характеризующие качество оценок, приведены в следующем определении:

Определение 2.9

Оценка (2.17)

$$\tilde{a} = \psi(X_1, X_2, \ldots, X_n)$$

некоторого параметра a называется симметричной, если выборочная функция $\psi(X_1, X_2, \ldots, X_n)$ является симметричной (см. разд. 2.2, определение 2.7).

Оценка (2.17) называется несмещенной, если

$$E\{\tilde{a}\} = \bar{a}. \tag{2.19}$$

Оценка (2.17) называется асимптотически несмещенной, если

$$E\{\tilde{a}\} \to \bar{a} \quad \text{при } n \to \infty. \tag{2.20}$$

Оценка (2.17) называется состоятельной, если ее значение при $n \to \infty$ с вероятностью 1 сходится к истинному значению $\bar{a}$, т. е.

$$P\left\{\lim_{n\to\infty} \tilde{a} = \bar{a}\right\} = 1. \tag{2.21}$$

Оценка (2.17) называется эффективной, если среди всех других возможных оценок она обладает наименьшей дисперсией:

$$D(\tilde{a}) = E\{(\tilde{a} - E\{\tilde{a}\})^2\} = \min! \tag{2.22}$$

Утверждение 2.5

Оценка (2.5) для математического ожидания некоторой случайной переменной X является симметричной, состоятельной и несмещенной, а для случая, когда величина X распределена по нормальному закону, также и эффективной.

Утверждение 2.6

Оценка (2.9) для дисперсии случайной переменной X является симметричной, состоятельной и несмещенной, а в случае нормального распределения величины X также и эффективной.

Утверждение 2.7

Оценка (2.13) для дисперсии случайной переменной X является смещенной; однако эта оценка состоятельная и асимптотически несмещенная.

Перейдем теперь к вопросу о том, как для параметра a или вектора параметров [1]

$$a = (a_1, a_2, \ldots, a_m)' \tag{2.23}$$

[1] Штрихом здесь и далее обозначается операция транспонирования. — *Прим. ред.*

построить некоторую подходящую оценку. Для этой цели может быть использован метод максимального правдоподобия, разработанный Р. А. Фишером.

Определение 2.10

Пусть $f(x; a)$ — функция плотности распределения вероятностей непрерывной случайной величины X, зависящая от вектора параметров a [см. (2.23)]. Тогда функция

$$L(X_1, X_2, \ldots, X_n; a) = f(X_1; a) f(X_2; a) \ldots f(X_n; a) \quad (2.24)$$

называется функцией правдоподобия.

Если X — дискретная случайная величина, принимающая значение y_i с вероятностью $p_i(a)$, $i = 1, 2, \ldots, k$, а q_i — число появлений значения y_i в реализации $(x_1, x_2, \ldots, x_n)$, то функция правдоподобия имеет вид

$$L(X_1, X_2, \ldots, X_n; a) = p_1^{q_1}(a) \cdot p_2^{q_2}(a) \cdot \ldots \cdot p_k^{q_k}(a). \quad (2.25)$$

Оценку

$$\tilde{a} = \psi(X_1, X_2, \ldots, X_n) \quad (2.26)$$

называют максимально правдоподобной оценкой параметра a, если она соответствует максимальному значению функции правдоподобия (2.24) или (2.25). Оценка $\tilde{a}$ при этом определяется из условия

$$\frac{\partial L}{\partial a_i}\Big|_{\tilde{a}_i} = 0, \quad i = 1, 2, \ldots, m. \quad (2.27)$$

Пример 2.1

Пусть X — нормально распределенная случайная величина, неизвестное математическое ожидание и дисперсию которой требуется оценить. Обозначим $a_1 = \overline{m}$, $a_2 = \sigma^2$ и введем эти обозначения в выражение $f_{m,\sigma}(x)$ плотности вероятностей нормального распределения (см. разд. 15.2):

$$f_{m,\sigma}(x) = f(x, a) = \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(x-a_1)^2}{2a_2}}.$$

Функция правдоподобия имеет вид

$$L(X_1, X_2, \ldots, X_n; a_1, a_2) = \left(\frac{1}{\sqrt{2\pi a_2}}\right)^n e^{-\frac{1}{2a_2}\sum_{i=1}^{n}(X_i - a_1)^2}.$$

Из условий

$$\left.\frac{\partial L}{\partial a_1}\right|_{\tilde{a}_1,\ \tilde{a}_2} = L\,\frac{1}{\tilde{a}_2}\sum_{i=1}^{n}(X_i-\tilde{a}_1)=0,$$

$$\left.\frac{\partial L}{\partial a_2}\right|_{\tilde{a}_1,\ \tilde{a}_2} = L\left\{-\frac{n}{2\tilde{a}_2}+\frac{1}{2\tilde{a}_2^2}\sum_{i=1}^{n}(X_i-\tilde{a}_1)^2\right\}=0$$

находим

$$\tilde{a}_1=\frac{1}{n}\sum_{i=1}^{n}X_i,\qquad \tilde{a}_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\tilde{a}_1)^2. \tag{2.28}$$

Полученные выражения совпадают с оценкой (2.5) для среднего и оценкой (2.13) для дисперсии.

Полученная с помощью некоторой выборочной функции $\tilde{a}=\psi(X_1, X_2, \ldots, X_n)$ конкретная оценка $\hat{a}=\psi(x_1, x_2, \ldots, x_n)$ параметра a не дает, однако, возможности судить о том, как точно найденная оценка воспроизводит истинное значение параметра a, даже если она является несмещенной и эффективной. Разумеется, величину разности $\hat{a}-\bar{a}$ нельзя определить точно, если распределение X и величина $\bar{a}$ не известны, так как $\hat{a}$ есть всего лишь реализация случайной величины $\tilde{a}$. Однако можно найти некоторую область, которая с вероятностью P содержит истинное значение параметра a. Если, например, $F(x)$—функция распределения величины $\tilde{a}-\bar{a}$, то можно рассчитать вероятность

$$P=P\{|\tilde{a}-\bar{a}|<\delta\}=F(\delta)-F(-\delta). \tag{2.29}$$

Если δ задано, то P определяет вероятность того, что случайная переменная $\tilde{a}$ отличается от $\bar{a}$ менее чем на δ, или вероятность того, что истинное значение параметра $\bar{a}$ менее чем на δ отличается от оценки $\hat{a}$ [1].

Истинное значение $\bar{a}$ с вероятностью P, задаваемой выражением (2.29), находится между $\hat{a}-\delta$ и $\hat{a}+\delta$, т. е. в интервале $\hat{I}=(\hat{a}-\delta, \hat{a}+\delta)$, представляющем собой реализацию интервала

$$\tilde{I}=(\tilde{a}-\delta, \tilde{a}+\delta). \tag{2.30}$$

Интервал $\tilde{I}$ называют обычно доверительным, а вероятность P — доверительной вероятностью.

[1] Здесь и ниже реализации оценок (т. е. значения $\hat{m}$, s^2) называются просто оценками.—*Прим. ред.*

Понятие доверительного интервала можно несколько обобщить, распространив его на интервалы, концы которых располагаются несимметрично относительно $\tilde{a}$.

Определение 2.11

Пусть $\tilde{a}_1$ и $\tilde{a}_2$ — выборочные функции, для которых

$$P\{\tilde{a}_1 < \bar{a} < \tilde{a}_2\} = P. \tag{2.31}$$

Тогда интервал

$$\tilde{I} = (\tilde{a}_1, \tilde{a}_2) \tag{2.32}$$

называется доверительным интервалом для параметра a, соответствующим доверительной вероятности P.

Пример 2.2

Оценивание с помощью доверительного интервала математического ожидания m нормально распределенной случайной величины X с известной дисперсией σ^2.

В качестве выборочной функции примем арифметическое среднее (2.5):

$$\tilde{m} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2.33}$$

Величина $\tilde{m}$ распределена по нормальному закону с математическим ожиданием $\bar{m}$ и дисперсией σ^2/n. Следовательно, величина

$$\tilde{M} = \frac{\tilde{m} - \bar{m}}{\sigma/\sqrt{n}} \tag{2.34}$$

имеет нормированное нормальное распределение (см. разд. 15.2) и, таким образом,

$$P\{|\tilde{M}| < e\} = \Phi(e) - \Phi(-e) = P. \tag{2.35}$$

Здесь $\Phi(x)$ — функция нормированного нормального распределения. Для всякой заданной вероятности P с помощью табл. 15.1 можно найти соответствующее значение $e = e(P)$. При этом доверительный интервал определяется как

$$\tilde{I} = (\tilde{m} - \delta,\ \tilde{m} + \delta), \tag{2.36}$$

где

$$\delta = \frac{e\sigma}{\sqrt{n}}. \tag{2.37}$$

Пусть из нормальной генеральной совокупности с неизвестным средним m и известной дисперсией $\sigma^2 = 25$ извлечена выборка объемом $n = 16$. На основе этой выборки рассчитана оценка $\tilde{m} = 20$.

Требуется построить доверительный интервал для m с доверительной вероятностью $P=0{,}95$. По табл. 15.1 из условия $\Phi(\varepsilon)=P/2=0{,}475$ находим $\varepsilon=1{,}96$ [1]. При этом с помощью (2.37) получаем

$$\delta=\frac{1{,}96\times 5}{4}=2{,}45.$$

Следовательно, математическое ожидание $\bar{m}$ с 95%-ной надежностью лежит в интервале

$$\hat{m}-\delta=17{,}55<\bar{m}<22{,}45=\hat{m}+\delta.$$

Пример 2.3

Оценивание с помощью доверительного интервала математического ожидания m нормально распределенной случайной величины X с неизвестной дисперсией σ^2.

В качестве выборочной функции для оценивания m используем арифметическое среднее (2.5):

$$\tilde{m}=\frac{1}{n}\sum_{i=1}^{n}X_i, \tag{2.38}$$

а для оценивания σ^2 — соответственно (2.9):

$$\tilde{s}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\tilde{m})^2. \tag{2.39}$$

Величина

$$\tilde{t}=\frac{\tilde{m}-\bar{m}}{\tilde{s}/\sqrt{n}} \tag{2.40}$$

при этих условиях имеет t-распределение (см. разд. 15.4) с числом степеней свободы $n-1$. Можно записать выражение

$$P\{|\tilde{t}|<\varepsilon\}=S_{n-1}(\varepsilon)-S_{n-1}(-\varepsilon)=P, \tag{2.41}$$

где S — функция распределения величины $\tilde{t}$. Для всякого заданного значения доверительной вероятности P с помощью табл. 15.3 можно определить соответствующее значение $\varepsilon=\varepsilon(P)$. Доверительный интервал при этом равен

$$\tilde{I}=(\tilde{m}-\delta,\ \tilde{m}+\delta), \tag{2.42}$$

---

[1] $\Phi(\varepsilon)=\int_0^{\varepsilon}\varphi(x)\,dx$, где $\varphi(x)$ — плотность нормированного нормального распределения. — *Прим. ред.*

где

$$\delta = \frac{\varepsilon \tilde{s}}{\sqrt{n}}. \tag{2.43}$$

Пусть из нормальной генеральной совокупности извлечена выборка объемом $n = 10$, с помощью которой рассчитаны оценки $\hat{m} = 20$ и $\hat{s} = \sqrt{10}$. Тогда для построения доверительного интервала с вероятностью $P = 0{,}99$ находим с помощью табл. 15.3 величину $\varepsilon = 3{,}25$. Используя выражение (2.43), вычисляем

$$\delta = \frac{3{,}25\sqrt{10}}{\sqrt{10}} = 3{,}25.$$

Следовательно, истинное значение m с 99%-ной надежностью находится в интервале

$$\hat{m} - \delta = 16{,}75 < m < 23{,}25 = \hat{m} + \delta.$$

Пример 2.4

Оценивание с помощью доверительного интервала дисперсии нормально распределенной случайной величины X при неизвестном математическом ожидании.

В качестве выборочных функций для оценивания m и σ^2 выберем, как и выше, функции (2.38) и (2.39) соответственно. Известно (см. разд. 15.3), что при этих условиях случайная величина

$$\tilde{\chi}^2 = \frac{(n-1)\tilde{s}^2}{\sigma^2} \tag{2.44}$$

распределена по закону χ^2 с $n-1$ степенями свободы. Здесь необходимо использовать расширенное понятие доверительного интервала, введенное в определении 2.11, так как применяемое выше выражение типа

$$P\{|\tilde{\chi}^2| < \varepsilon\} = P$$

для нахождения доверительного интервала в данном случае лишено смысла в силу неотрицательности $\tilde{\chi}^2$. Обозначим

$$P = 1 - \alpha \tag{2.45}$$

и найдем из табл. 15.2 два числа e_1 и e_2 из условия

$$P\{\tilde{\chi}^2 < e_1\} = P\{\tilde{\chi}^2 > e_2\} = \frac{\alpha}{2}. \tag{2.46}$$

Тогда

$$P\{e_1 < \tilde{\chi}^2 < e_2\} = P, \tag{2.47}$$

При доверительной вероятности P доверительный интервал для σ^2

$$\tilde{I} = (\tilde{a}_1,\ \tilde{a}_2), \tag{2.48}$$

где

$$\tilde{a}_i = \frac{(n-1)\,\tilde{s}}{v_i}. \tag{2.49}$$

С понятием доверительного интервала и доверительной вероятности связаны три основных типа задач. Первая задача состоит в том, чтобы при заданном интервале определить соответствующую ему доверительную вероятность. Особого практического значения эта задача не имеет. В этом случае сначала устанавливается интервал, а затем спрашивается, какова вероятность того, что истинное значение параметра находится в этом интервале.

Вторая задача состоит в том, чтобы по заданной величине доверительной вероятности построить соответствующий ей доверительный интервал. В этом случае сначала задаются степенью статистической надежности P, а затем выясняют вопрос о точности оценки с этой надежностью. Этот тип задач имеет важное практическое значение. В рассмотренных выше примерах 2.2—2.4 задача формулировалась именно таким образом.

До сих пор объем выборки считался заданным. Однако доверительная вероятность для некоторого интервала существенно зависит от n. При заданном интервале вероятность

$$P\{\tilde{a}_1 < \bar{a} < \tilde{a}_2\} = W(n)$$

есть монотонно возрастающая функция натуральных чисел $n = 1, 2, 3, \ldots$. Так как n — число экспериментов и увеличение n часто связано со значительным ростом затрат, следует стремиться к тому, чтобы число экспериментов не превышало минимального числа, необходимого для достижения требуемой точности при заданной степени надежности.

Третья основная задача, связанная с проблемой определения объема выборки, заключается в том, чтобы при заданной величине доверительного интервала и при заданной надежности P найти наименьшее натуральное число n, для которого

$$W(n) \geqslant P.$$

Пример 2.5

Вернемся к примеру 2.2. В этом примере была построена такая оценка $\tilde{m}$ среднего генеральной совокупности, что ее отклонение от истинного значения $\bar{m}$ с 95%-ной надежностью не превосходило величины $\delta = 2{,}45$. Если такая точность не достаточна и с 95%-ной надежностью требуется обеспечить ошибку $\delta \leqslant 1$,

возникает вопрос о том, какой минимальный объем выборки для этого потребуется. Задача третьего типа здесь может быть поставлена так: найти минимальное целое число n, для которого с вероятностью $P=0{,}95$ отклонение $\delta \leqslant 1$. Из (2.37) получаем

$$n \geqslant \frac{v^2 \sigma^2}{\delta^2} \geqslant (1{,}96)^2 \times 25 = 96{,}04.$$

Таким образом, для достижения требуемой точности здесь потребуется сделать около 100 опытов.

2.4. Статистическая проверка гипотез

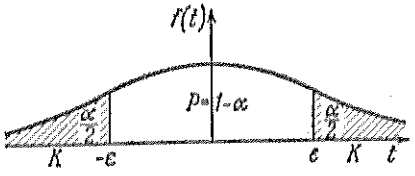
В разд. 2.1 уже отмечалось, что наряду со статистической оценкой параметров задачи проверки гипотез составляют один из важнейших разделов математической статистики. Ниже мы покажем, что оба упомянутых раздела математической статистики тесно взаимосвязаны.

Под гипотезой H мы в общем случае будем понимать некоторое предположение о случайной величине X (например, о виде функции распределения, параметрах распределения и т. п.). Путем статистической проверки необходимо установить, насколько данные, полученные из выборки $(X_1, X_2, \ldots, X_n)$, согласуются с высказанной гипотезой, т. е. можно ли на их основании принять или отвергнуть гипотезу.

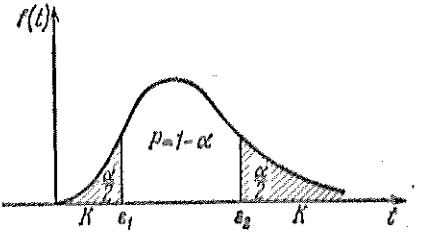
Абсолютно надежное решение относительно проверяемой гипотезы, вообще говоря, получить нельзя. Необходимо заранее допустить возможность ошибочного решения. Обозначим через α вероятность того, что гипотеза H будет отвергнута, хотя на самом деле она верна. Эту вероятность называют также уровнем значимости проверки гипотезы. При проверке гипотезы величина α или величина $P=1-\alpha$, характеризующая статистическую надежность, должна быть выбрана экспериментатором. При решении технических и экономических проблем в большинстве случаев выбирают $\alpha=0{,}05$ или $\alpha=0{,}01$ (что соответствует уровням значимости 5 и 1%), в медицинских исследованиях α часто принимают равной 0,001 (уровень значимости 0,1%). Процедура проверки гипотезы заключается в следующем: выбирается некоторая подходящая выборочная функция (критерий проверки гипотезы — Testgröße) $T(X_1, X_2, \ldots, X_n; H)$, которая определяется выборкой и выдвинутой гипотезой H. Затем устанавливается область K, в которую в случае справедливости гипотезы H значение функции T попадает с вероятностью P_H, равной заданной величине α. Область K называется критической областью. Если конкретное значение $T(x_1, x_2, \ldots, x_n; H)$ попадает в критическую область K, гипотеза H отклоняется; в противном случае считается, что оснований для отклонения гипотезы нет. При этом вероятность

того, что гипотеза H будет отвергнута в случае, когда на самом деле она верна, оказывается равной заданной вероятности α.

При любом значении α существует множество различных возможностей для выбора критической области. На фиг. 2.1 в качестве примера приведена функция плотности распределения



Фиг. 2.1. Симметричная критическая область.



Фиг. 2.2. Квазисимметричная критическая область.

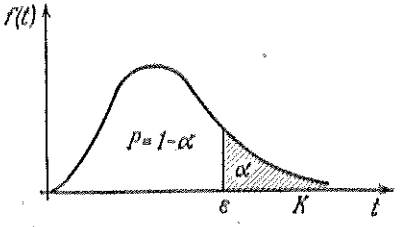
$f(t)$ для некоторого критерия T

$$P\{T < t\} = \int_{-\infty}^{t} f(t)\,dt,$$

которая симметрична относительно нуля и близка по виду к кривой нормального или t-распределения. Критическая область также выбрана здесь симметричной относительно нуля:

$$K\colon |t| \geqslant \varepsilon.$$

Эта область расположена под заштрихованными участками, площадь каждого из которых равна $\alpha/2$.



Фиг. 2.3. Односторонняя критическая область.

На фиг. 2.2 и 2.3 приведены функции плотности распределения, близкие по виду к χ^2-или F-распределениям. На фиг. 2.3 показана односторонняя критическая область, расположенная в диапазоне больших значений критерия:

$$K\colon t \geqslant \varepsilon.$$

Заштрихованная часть, расположенная над критической областью, имеет площадь, равную α. На фиг. 2.2 показана квазисимметричная критическая область

$$K\colon\ 0 \leqslant t \leqslant e_1,\ e_2 \leqslant t,$$

одна часть которой располагается правее нуля, а другая — в области больших значений критерия. Каждая из заштрихованных частей, расположенных над отрезками, соответствующими критической области, имеет площадь, равную $\alpha/2$.

Прежде чем перейти к конкретным задачам выбора критических областей, сформулируем приведенные выше соображения в виде определения и рассмотрим некоторые примеры.

Определение 2.12

Пусть $(X_1, X_2, \ldots, X_n)$ — выборка объема n из генеральной совокупности значений случайной переменной X, а $(x_1, x_2, \ldots, x_n)$ — реализация выборки. Допустим, что выдвинута гипотеза H и выбран критерий проверки (Testgröße)

$$T = T(X_1, X_2, \ldots, X_n; H), \tag{2.50}$$

зависящий от выборки и от высказанной гипотезы. Зададимся вероятностью α (уровнем значимости) или статистической надежностью

$$P = 1 - \alpha. \tag{2.51}$$

Подобласть K области изменения критерия T называется критической областью объема α, если вероятность $P_H\{T \in K\}$ того, что в случае справедливости гипотезы H T попадет в область K, равна α:

$$P_H\{T \in K\} = \alpha. \tag{2.52}$$

Если выборочное значение критерия

$$\hat{T} = T(x_1, x_2, \ldots, x_n; H) \tag{2.53}$$

попадет в область K $(T \in K)$, то гипотеза H отвергается. Подобная процедура называется статистической проверкой гипотезы.

Пример 2.6

Проверка гипотезы о среднем значении (математическом ожидании) нормально распределенной случайной величины X с известной дисперсией σ^2 (см. разд. 2.3, пример 2.2).

Гипотеза H утверждает, что среднее значение X равно заданной величине m_0, т. е.

$$H\colon\ m = m_0. \tag{2.54}$$

В качестве критерия проверки выберем функцию

$$\tilde{T} = \frac{\tilde{m} - m_0}{\sigma / \sqrt{n}}, \tag{2.55}$$

где

$$\tilde{m} = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{2.56}$$

Критерий $\tilde{T}$ в случае справедливости гипотезы H имеет нормированное нормальное распределение. Критическую область K, соответствующую уровню значимости α, выберем симметричной, согласно фиг. 2.1:

$$K: |t| \geqslant \varepsilon.$$

Для определения ε необходимо разрешить уравнение [см. (2.35)]

$$P\{|\tilde{T}| < \varepsilon\} = \Phi(\varepsilon) - \Phi(-\varepsilon) = P = 1 - \alpha. \tag{2.57}$$

Если значение

$$\hat{T} = \frac{\hat{m} - m_0}{\sigma / \sqrt{n}} \tag{2.58}$$

попадает в область K, гипотеза (2.54) отвергается.

Пусть из нормальной совокупности с дисперсией $\sigma^2 = 25$ извлечена выборка объема $n = 16$, с помощью которой получена оценка среднего $\hat{m} = 22$. Требуется проверить гипотезу $m = 20$. Для этого проделаем следующее.

При выбранном значении $\alpha = 0{,}05$ (это соответствует $P = 1 - 0{,}05 = 0{,}95$) найдем величину ε из уравнения (2.57), используя табл. 15.1:

$$\varepsilon = 1{,}96.$$

При этом

$$K: |t| \geqslant 1{,}96.$$

В данном случае

$$t = \hat{T} = \frac{22 - 20}{5/4} = \frac{8}{5} = 1{,}6 < 1{,}96.$$

Отсюда следует, что гипотеза H не отклоняется.

Пример 2.7

Проверка гипотезы о среднем значении нормально распределенной случайной величины X с неизвестной дисперсией (см. разд. 2.3, пример 2.3).

Проверяется та же гипотеза, что и в примере 2.6:

$$H: m = m_0. \tag{2.59}$$

В качестве критерия проверки выберем функцию

$$\tilde{T} = \frac{\tilde{m} - m_0}{s/\sqrt{n}}, \tag{2.60}$$

где

$$\tilde{m} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.61}$$

и

$$\tilde{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \tilde{m})^2. \tag{2.62}$$

Критерий $\tilde{T}$ в случае справедливости гипотезы H имеет t-распределение с числом степеней свободы $n-1$. Выберем критическую область симметричной относительно нуля

$$K: |t| \geqslant \varepsilon \tag{2.63}$$

и найдем величину ε из уравнения

$$P\{|\tilde{T}| < \varepsilon\} = S_{n-1}(\varepsilon) - S_{n-1}(-\varepsilon) = P = 1 - \alpha, \tag{2.64}$$

где S — функция распределения Стьюдента (t-распределения). Значение ε, удовлетворяющее условию (2.64), будем называть критическим значением t-распределения при числе степеней свободы $n-1$ и уровне значимости α и обозначим его $t_{кр}$.

Пусть из нормально распределенной генеральной совокупности извлечена выборка объема $n = 10$, с помощью которой найдены оценка среднего $\hat{m} = 24$ и оценка среднеквадратичного отклонения $\hat{s} = \sqrt{10}$. При этом величина критерия для проверки гипотезы $m = 20$ оказывается равной

$$t = \hat{T} = \frac{24 - 20}{\sqrt{10}/\sqrt{10}} = 4.$$

Из условия (2.64) с помощью табл. 15.3 для уровня значимости $\alpha = 0{,}01$ (или соответственно $P = 0{,}99$) находим

$$t_{кр} = \varepsilon = 3{,}25.$$

Так как $\hat{T} > \varepsilon$, то $\hat{T}$ попадает в критическую область и гипотезу $m = 20$ следует отклонить.

Пример 2.8

Проверка гипотезы о значении дисперсии σ^2 нормально распределенной случайной величины X при неизвестном среднем значении (см. разд. 2.3, пример 2.4).

Проверяется гипотеза

$$H\colon\ \sigma^2=\sigma_0^2. \tag{2.65}$$

В качестве критерия используем функцию

$$\hat{T}=\tilde{\chi}^2=\frac{(n-1)\,\tilde{s}^2}{\sigma_0^2}. \tag{2.66}$$

Эта величина, в случае когда H справедлива, подчиняется χ^2-распределению с числом степеней свободы $n-1$. Критическую область, соответствующую уровню значимости α, выберем квазисимметричной, согласно фиг. 2.2. Из табл. 15.2 найдем значения e_1 и e_2, соответствующие условию

$$P\{\hat{T}<e_1\}=P\{\hat{T}>e_2\}=\frac{\alpha}{2}. \tag{2.67}$$

Если выборочное значение $\hat{T}$ попадет в критическую область

$$K\colon\ 0\leqslant\hat{T}\leqslant e_1;\quad e_2\leqslant\hat{T}, \tag{2.68}$$

то гипотезу (2.65) следует отвергнуть.

Пусть из нормально распределенной совокупности извлечена выборка объема $n=40$, с помощью которой рассчитана оценка дисперсии $\tilde{s}^2=20{,}61$. Величина критерия (2.66) при этом оказывается равной

$$\chi^2=\hat{T}=\frac{39\times 20{,}61}{20}=40{,}2.$$

При уровне значимости $\alpha=0{,}05$ находим $e_1=24{,}4$; $e_2=59{,}3$. Так как $\hat{T}$ не попадает в критическую область (2.68), гипотеза $\sigma^2=20$ не отклоняется.

Пример 2.9

Проверка гипотезы о равенстве дисперсий σ_1^2 и σ_2^2 двух независимых нормально распределенных случайных величин X и Y при неизвестных средних значениях.

Проверяется гипотеза

$$H\colon\ \sigma_1^2=\sigma_2^2. \tag{2.69}$$

В качестве критерия проверки выбирают функцию

$$\hat{T}=\frac{\tilde{s}_1^2}{\tilde{s}_2^2}. \tag{2.70}$$

Здесь

$$\tilde{s}_1^2=\frac{1}{n-1}\sum_{l=1}^{n}(X_l-\tilde{m}_1)^2 \tag{2.71}$$

и

$$\hat{s}_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \tilde{m}_2)^2 \qquad (2.72)$$

представляют собой несмещенные оценки дисперсий, рассчитанные, согласно формуле (2.9), по выборке объема n для переменной X и выборке объема m для переменной Y. Величины $\tilde{m}_i$ являются оценками средних значений, которые определяются согласно выражению (2.5). В случае когда гипотеза H справедлива, критерий $\hat{T}$ имеет F-распределение с числами степеней свободы $n-1$ и $m-1$.

Обозначим, как и в разд. 15.5, через $F_{n-1,\,m-1}$ функцию распределения Фишера (F-распределения). Тогда, выбирая квазисимметричную критическую область согласно фиг. 2.2:

$$K:\ 0 \leqslant t \leqslant e_1,\ e_2 \leqslant t, \qquad (2.73)$$

находим для заданного уровня значимости α величины e_1 и e_2 из уравнения

$$F(e_1) = 1 - F(e_2) = \frac{\alpha}{2}. \qquad (2.74)$$

Если выборочное значение $\hat{T}$ попадает в область (2.73), гипотеза (2.69) отвергается.

Из (2.74) следует, что $e_2 = 1/e_1$, поэтому в табл. 15.4 и 15.5 содержатся только значения правых критических точек e_2 (величину e_2 мы ниже иногда будем обозначать также через $F_{\text{кр}}$). При вычислении критерия $F = \hat{T}$ в (2.70) в числитель всегда ставится бо́льшая из оценок $\hat{s}_1^2$ и $\hat{s}_2^2$. Если $F = \hat{T}$ оказывается больше, чем $F_{\text{кр}} = e_2$, то гипотеза (2.69) отвергается.

Замечание 2.1

При проверке адекватности модели используется F-критерий с несимметричной критической областью вида фиг. 2.3. Процедура решения этой задачи подробно рассматривается в разд. 3.2.5 и 3.3.7.

Перейдем теперь к вопросу о целесообразном выборе критической области K. При этом проанализируем возможные ошибки, возникающие при принятии или отклонении гипотезы, и значение этих ошибок.

Определение 2.13

Если гипотеза отклоняется, когда она на самом деле верна, говорят об ошибке первого рода. Если гипотеза не отклоняется, когда она на самом деле не верна, имеет место ошибка второго рода.

Утверждение 2.8

Уровень значимости α есть вероятность ошибки первого рода.

Определение 2.14

Пусть $\overline{H}$ — альтернативная гипотеза по отношению к гипотезе H. Тогда $P_{\overline{H}}\{T \in K\}$ есть вероятность того, что гипотеза H отвергается, когда она не верна, т. е. когда истинной является конкурирующая гипотеза $\overline{H}$. Эта вероятность характеризует избирательность критерия проверки и называется мощностью критерия. Чем больше мощность критерия проверки, тем меньше вероятность ошибки второго рода. При заданной вероятности ошибки первого рода критическую область выбирают так, чтобы обеспечить максимальную избирательность критерия:

$$P_{\overline{H}}\{T \in K\} = \max! \tag{2.75}$$

Замечание 2.2

В примерах 2.7—2.9 критическая область выбрана наилучшим образом в вышеуказанном смысле. Положение оптимальной области K в смысле (2.75) существенно зависит от выбора альтернативной гипотезы $\overline{H}$. В примерах 2.6 и 2.7 конкурирующей является гипотеза $\overline{H}$: $m \neq m_0$, в примере 2.8 — $\overline{H}$: $\sigma^2 \neq \sigma_0^2$, а в примере 2.9 — $\overline{H}$: $\sigma_1^2 \neq \sigma_2^2$. Если в примере 2.8 использовать альтернативную гипотезу $\overline{H}$: $\sigma^2 > \sigma_0^2$, то в качестве критической следует выбрать область, показанную на фиг. 2.3. Подобные же выводы могут быть сделаны и в отношении других примеров.

После рассмотрения общих проблем теории проверки гипотез и изложения ряда примеров рассмотрим кратко еще несколько важных задач проверки гипотез. Подробное их изложение и практические примеры можно найти в литературе по математической статистике.

Пример 2.10

Проверка гипотезы о равенстве средних значений m_1 и m_2 двух независимых нормально распределенных случайных величин X и Y при известных дисперсиях σ_1^2 и σ_2^2.

Проверяется гипотеза

$$H\colon\ m_1 = m_2. \tag{2.76}$$

В качестве критерия проверки выбираем функцию

$$\tilde{T} = \frac{\tilde{m}_1 - \tilde{m}_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}. \tag{2.77}$$

Здесь $\tilde{m}_1$ и $\tilde{m}_2$ — выборочные средние для выборки объема n_1 и n_2 соответственно, определяемые по формуле (2.5). Величина $\tilde{T}$

подчинена нормированному нормальному закону распределения. Критическая область выбирается симметричной, подобно показанной на фиг. 2.1.

Пример 2.11

Проверка гипотезы о равенстве дисперсий σ_1^2, σ_2^2, ..., σ_k^2 k независимых нормально распределенных случайных величин $X^{(1)}$, $X^{(2)}$, ..., $X^{(k)}$.

Проверяется гипотеза

$$H\colon\ \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2. \tag{2.78}$$

В качестве критерия выбирается функция

$$\tilde{T} = \max_{1 \leqslant l \leqslant k} \frac{\tilde{s}_l^2}{\tilde{s}_1^2 + \tilde{s}_2^2 + \ldots + \tilde{s}_k^2}. \tag{2.79}$$

Здесь $\tilde{s}_l^2$ —выборочная дисперсия переменной $X^{(l)}$, определяемая согласно выражению (2.9) при одинаковом объеме n выборки для каждой l-й переменной. Критические значения величины $\tilde{T}$ при различных значениях k и n для ряда значений α можно найти, например, в [3]. Критическая область выбирается несимметричной, подобно показанной на фиг. 2.3.

Пример 2.12

Проверка гипотезы о виде закона распределения $F(x)$ случайной переменной X.

Если закон распределения вероятностей $F(x)$ некоторой случайной переменной X не известен, можно получить представление о его форме, построив на основе выборки эмпирическую функцию распределения $W_n(x)$ (см. разд. 2.2). При этом оказывается возможным высказать предположение о том, что действительный вид закона распределения есть функция $F_0(x)$. Тогда требуется проверить гипотезу

$$H_0\colon\ F(x) = F_0(x). \tag{2.80}$$

Для проверки гипотезы делят интервал изменения случайной переменной X на k непересекающихся подынтервалов $\Delta_1, \Delta_2, \ldots, \Delta_k$.

Пусть p_l—вероятность того, что X попадает в подынтервал Δ_l в предположении, что закон распределения X описывается функцией $F_0(x)$:

$$p_l = P\{X \in \Delta_l \mid F_0(x)\}. \tag{2.81}$$

Очевидно, что

$$p_1 + p_2 + \ldots + p_k = 1. \tag{2.82}$$

Обозначим через $\tilde{M}_l$ число значений x_l в выборке объема n, попадающих в подынтервал Δ_l. Тогда

$$\tilde{M}_1 + \tilde{M}_2 + \ldots + \tilde{M}_k = n. \tag{2.83}$$

Согласно разд. 15.1, случайные величины $\tilde{M}_l$ в случае справедливости гипотезы H_0 подчинены биномиальному распределению с параметрами

$$E\{\tilde{M}_l\} = np_l, \tag{2.84}$$

$$D\{\tilde{M}_l\} = np_l(1-p_l) = \sigma_l^2. \tag{2.85}$$

При $n \to \infty$ распределение $\tilde{M}_l$ стремится к нормальному с теми же моментами.

При правильности гипотезы величины

$$\tilde{\psi}_l = \frac{\tilde{M}_l - np_l}{\sqrt{np_l}} \tag{2.86}$$

также будут асимптотически распределены нормально. Эти величины связаны между собой соотношением

$$\sum_{l=1}^{k} \tilde{\psi}_l \sqrt{p_l} = 0. \tag{2.87}$$

В качестве меры расхождения выборочных величин $\tilde{M}_l$ и их теоретических значений np_l рассмотрим величину

$$\tilde{T} = \sum_{l=1}^{k} \tilde{\psi}_l^2 = \sum_{l=1}^{k} \frac{(\tilde{M}_l - np_l)^2}{np_l} = \sum_{l=1}^{k} \frac{\tilde{M}_l^2}{np_l} - n. \tag{2.88}$$

Распределение случайной величины $\tilde{T}$ при $n \to \infty$ стремится к χ^2-распределению с числом степеней свободы $k-1$. Для выборок достаточно большого объема n принимается, что $\tilde{T}$ приближенно следует χ^2-распределению. При этом для выбранного уровня значимости α и критической области вида фиг. 2.3 с помощью таблиц χ^2-распределения (см. табл. 15.2) определяется величина e, соответствующая условию

$$P\{\tilde{T} > e\} = \alpha. \tag{2.89}$$

Гипотеза H_0 отклоняется, если $\tilde{T} > e$.

Пример 2.13

Выявление аномальных наблюдений (Ausreißertest).

В ряду измерений может появиться некоторый результат, который оказывается существенно большим или меньшим остальных значений. При этом может возникнуть подозрение, что этот результат ошибочен, не является типичным для выборки и возник вследствие грубой ошибки при измерении. Такие значения называют обычно аномальными (резко выделяющимися значе-

шиями, или выбросами). Для решения вопроса о том, является ли некоторое наблюдение аномальным или нет, используются специальные критерии. Существуют различные способы выявления аномальных наблюдений, отличающиеся имеющейся информацией о параметрах распределения случайной величины. В простейшем случае, полагая, что X подчиняется нормальному закону с известной дисперсией σ^2, для выявления аномальности может быть использована величина

$$\tilde{T} = \frac{X_l - \tilde{m}}{\sigma}, \tag{2.90}$$

которая имеет нормированное нормальное распределение (при достаточно больших выборках, используемых для вычисления $\tilde{m}$). При заданном уровне значимости α из условия

$$P\{|\tilde{T}| > \varepsilon\} = \alpha \tag{2.91}$$

может быть найдена величина ε, определяющая положение критической области вида фиг. 2.1. Если

$$\frac{|x_l - \hat{m}}{\sigma} > \varepsilon, \tag{2.92}$$

наблюдение x_l считается аномальным.

Пример 2.14

Проверка гипотезы о независимости двух нормально распределенных случайных величин X и Y.

При выяснении вопроса о независимости нормально распределенных случайных величин X и Y проверяется гипотеза

$$H_0: \rho = 0, \tag{2.93}$$

где ρ — коэффициент корреляции. При указанных выше предположениях выборочный коэффициент корреляции

$$\tilde{r} = \frac{\tilde{m}_{1,2}}{\tilde{s}_1 \tilde{s}_2}, \tag{2.94}$$

где

$$\tilde{m}_{1,2} = \frac{1}{n}\left\{\sum_{l=1}^{n} X_l Y_l - \frac{\left(\sum_{l=1}^{n} X_l\right)\left(\sum_{l=1}^{n} Y_l\right)}{n}\right\}, \tag{2.95}$$

$$\tilde{s}_1^2 = \frac{1}{n}\sum_{l=1}^{n}(X_l - \tilde{m}_1)^2, \tag{2.96}$$

$$\tilde{s}_2^2 = \frac{1}{n}\sum_{l=1}^{n}(Y_l - \tilde{m}_2)^2, \tag{2.97}$$

распределен нормально с нулевым средним и дисперсией

$$\sigma_r^2 \approx \tilde{s}_r^2 = \frac{(1-\tilde{r}^2)^2}{n}. \tag{2.98}$$

При этом величина

$$\tilde{T} = \frac{\tilde{r}}{\sigma_r} \tag{2.99}$$

подчинена нормированному нормальному распределению. При заданном уровне значимости α из условия

$$P\{|\tilde{T}| > e\} = \alpha \tag{2.100}$$

может быть найдено значение e.

Гипотеза $\rho = 0$ отклоняется, если $\hat{r} > e\sigma_r$.

ЛИТЕРАТУРА

1. Müller P. H. (Hrsg.), Wahrscheinlichkeitsrechnung und mathematische Statistik, Lexikon, Akademie-Verlag, Berlin, 1970.
2. Fisz M., Rachunek prawdopodobeństwa i statistyka matematyczna, Biblioteka Matematyczna, t. 18, Państwowe Wydawnictwo Naukowe, Warszawa, 1958.
3. Смирнов Н. В., Дунин-Барковский И. В., Курс теории вероятностей и математической статистики, изд-во „Наука", М., 1969.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Химмельблау Д., Анализ процессов статистическими методами, изд-во „Мир", М., 1973.
2. Митропольский А. К., Техника статистических вычислений, изд-во „Наука", М., 1971.
3. Леман Э., Проверка статистических гипотез, изд-во „Наука", М., 1964.
4. Худсон Д., Статистика для физиков, изд-во „Мир", М., 1970.
5. Налимов В. В., Применение математической статистики при анализе вещества, Физматгиз, М., 1960.
6. Большев Л. Н., Смирнов Н. В., Таблицы математической статистики, изд-во „Наука", М., 1965.
7. Хан Г., Шапиро С., Статистические модели в инженерных задачах, изд-во „Мир", М., 1969.
8. Шор Я. Б., Бендерский А. Н., Методы оценки аномальных результатов измерений, М., 1953.
9. Вентцель Е. С., Теория вероятностей, Физматгиз, М., 1962.

3.

ПОСТРОЕНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ С ПОМОЩЬЮ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

3.1. Введение

На практике в производственных или лабораторных условиях встречаются процессы, характер протекания которых детерминированным образом зависит от определенных величин

$$x_1, x_2, \ldots, x_n.$$

Например, выход у некоторого химического реактора может определяться значениями давления P, температуры T и времени протекания реакции t. Переменные $x_1, x_2, \ldots, x_n$ обычно называют входными контролируемыми или независимыми переменными, и их возможные значения принадлежат некоторой области Ω_x n-мерного пространства. Выходную переменную y мы в дальнейшем будем называть зависимой переменной, целевой величиной или выходом процесса, даже если она не обозначает буквально выход продукта. В общем случае можно сказать, что между независимыми переменными и выходом процесса существует функциональная взаимосвязь

$$y = y(x), \tag{3.1}$$

где

$$x = (x_1, x_2, \ldots, x_n)' \tag{3.2}$$

— вектор значений независимых переменных. Зависимость (3.1) на практике часто бывает не известна, и тогда ее пытаются найти путем обработки экспериментальных данных.

Так как всякий эксперимент связан с появлением случайных ошибок, то при построении математических моделей на основе экспериментальных данных необходимо использовать методы математической статистики.

Наиболее часто при решении этой задачи применяют метод наименьших квадратов. Метод наименьших квадратов позволяет построить оптимальную в определенном смысле оценку моментов распределения ошибки эксперимента, а также решить вопрос о том, является ли полученная модель адекватной (т. е. соответствует ли она действительности). Этот метод будет рассмотрен в разд. 3.2. Учитывая важное значение метода наименьших квадратов в теории планирования эксперимента и вообще в задачах статистической обработки данных, основные выводы разд. 3.2 в краткой форме будут повторены в разд. 3.3.

3.2. Метод наименьших квадратов

3.2.1. Постановка задачи

Пусть требуется на основе экспериментальных данных построить модель некоторого процесса. При этом прежде всего необходимо составить себе какое-то представление о структуре этой модели. Из физических соображений можно, например, предположить, что взаимосвязь между y и x_i линейна:

$$y(a, x) = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n. \tag{3.3}$$

При этом a_i являются неизвестными параметрами процесса, оценки которых требуется найти путем обработки экспериментальных данных. В случае если характер связи описывается квадратичной функцией, имеем

$$\begin{aligned} y(a, x) = {} & a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n + \\ & + a_{n+1}x_1x_1 + a_{n+2}x_2x_2 + \ldots + a_{2n}x_nx_n + \\ & + a_{2n+1}x_1x_2 + a_{2n+2}x_1x_3 + \ldots + a_{3n-1}x_1x_n + \\ & + a_{3n}x_2x_3 + \ldots + \\ & \cdots\cdots\cdots\cdots \\ & + a_kx_{n-1}x_n. \end{aligned} \tag{3.4}$$

Здесь

$$k + 1 = \binom{n+2}{2} = \frac{(n+2)(n+1)}{2}. \tag{3.5}$$

Обычно коэффициенты квадратичной модели вида (3.4) нумеруются не по порядку, а так, что коэффициент при функции x_ix_j обозначается через a_{ij}.

Модели полиномиального вида имеют большое значение в связи с тем, что с их помощью любая аналитическая функция (3.1) может быть описана как угодно точно. Однако с увеличением степени полинома весьма существенно увеличивается число оцениваемых параметров модели и соответственно возрастают затраты на эксперимент. Так, если степень полинома есть m, то

$$k + 1 = \binom{n+m}{m}. \tag{3.6}$$

В дальнейшем мы будем иметь дело с моделями вида

$$y = y(x, a), \tag{3.7}$$

где a — вектор параметров модели,

$$a = (a_0, a_1, \ldots, a_k)'. \tag{3.8}$$

Примем, что модель (3.7) линейна относительно коэффициентов a_l, т. е.

$$y(a, x) = a_0 f_0(x) + a_1 f_1(x) + \ldots + a_k f_k(x). \tag{3.9}$$

При этом $f_l(x)$ — известные функции, являющиеся компонентами вектора

$$f(x) = (f_0(x), f_1(x), \ldots, f_k(x))'. \tag{3.10}$$

Используя векторные обозначения, вместо (3.9) можно записать

$$y = a' f(x) = f'(x)\, a. \tag{3.11}$$

В случае модели вида (3.3) или (3.4) получаем соответственно следующие выражения для компонент $f(x)$:

$$f(x) = (1, x_1, x_2, \ldots, x_n)' \tag{3.12}$$

и

$$f(x) = (1, x_1, x_2, \ldots, x_n, x_1x_1, x_2x_2, \ldots, x_nx_n, x_1x_2, x_1x_3, \ldots, x_1x_n, x_2x_3, \ldots, x_{n-1}x_n)'. \tag{3.13}$$

Для истинных (действительных) значений вектора коэффициентов a в (3.11), которые мы в последующем будем обозначать через $\bar{a}$, требуется найти оценки $\hat{a}$, используя для этой цели результаты эксперимента. При этом оценка $\hat{y}$ для y рассчитывается по формуле

$$\hat{y} = \hat{a}' f(x) = f'(x)\, \hat{a}. \tag{3.14}$$

Эксперимент приводится в N точках

$$x^1, x^2, \ldots, x^N$$

с координатами

$$x^l = (x_1^l, x_2^l, \ldots, x_n^l)'. \tag{3.15}$$

Результаты наблюдений $\tilde{y}^l$ в точках x^l представляются с помощью вектора наблюдений

$$\tilde{Y} = (\tilde{y}^1, \tilde{y}^2, \ldots, \tilde{y}^N)'. \tag{3.16}$$

В каждой точке x^l может быть поставлено ν опытов, результатами которых будут

$$\tilde{y}^{l1}, \tilde{y}^{l2}, \ldots, \tilde{y}^{l\nu}.$$

В этом случае в качестве $\tilde{y}^l$ используется среднее значение наблюдений в точке x^l:

$$\tilde{y}^l = \nu^{-1}(\tilde{y}^{l1} + \tilde{y}^{l2} + \ldots + \tilde{y}^{l\nu}). \tag{3.17}$$

Задача состоит теперь в том, чтобы на основе результатов (3.16) найти наилучшие в определенном смысле оценки $\hat{a}$ и $\hat{y}$.

3.2.2. Решение задачи

Чтобы решить задачу, сформулированную в разд. 3.2.1, необходимо сначала выяснить, что следует понимать под наилучшими оценками.

Будем исходить из того, что модель вида (3.11) является адекватной. Вопросы о том, что произойдет, если (3.11) не соответствует действительности, и как на основе экспериментальных данных проверить адекватность модели, будут обсуждаться в разд. 3.2.4 и 3.2.5 соответственно.

Сопоставим теперь друг с другом экспериментальные результаты (3.16), отражающие действительность, и значения

$$\hat{Y} = (\hat{y}^1, \hat{y}^2, \ldots, \hat{y}^N)', \tag{3.18}$$

рассчитанные с помощью $\hat{a}$ и представляющие модель (3.14). Имеем

$$\hat{y}^l = \hat{a}' f(x^l) = f'(x^l)\,\hat{a} \tag{3.19}$$

или соответственно

$$\hat{Y} = F\hat{a}, \tag{3.20}$$

где матрица F определяется следующим образом:

$$F = (f_j(x^l)) = \begin{bmatrix} f_0(x^1)\, f_1(x^1) \ldots f_k(x^1) \\ f_0(x^2)\, f_1(x^2) \ldots f_k(x^2) \\ \vdots \\ f_0(x^N)\, f_1(x^N) \ldots f_k(x^N) \end{bmatrix}. \tag{3.21}$$

Результат наблюдения $\tilde{y}^l$ в некоторой точке зависит от случайной ошибки $\tilde{e}^l$:

$$\tilde{e}^l = \tilde{y}^l - \bar{y}^l. \tag{3.22}$$

Множество значений ошибок в N экспериментальных точках может быть представлено вектором

$$\tilde{e} = \tilde{Y} - \bar{Y}. \tag{3.23}$$

Здесь черта сверху, как и выше, означает истинное значение соответствующей переменной.

Наложим на результаты наблюдений три условия, которые на практике, как правило, выполняются:

1. Результаты эксперимента свободны от систематических ошибок, или, иными словами, математическое ожидание величины $\tilde{y}^l$ равно действительному значению $\bar{y}^l$:

$$E\{\tilde{Y}\} = \bar{Y} = (\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^N) = F\bar{a}, \tag{3.24}$$

т. е.

$$E(\tilde{e}) = 0. \tag{3.25}$$

2. Результат наблюдения в точке x^j не зависит от результата в точке x^l, т. е.

$$E\{(\tilde{y}^j - \bar{y}^j)(\tilde{y}^j - \bar{y}^j)\} = 0 \quad \text{для} \quad i \neq j, \tag{3.26}$$

или

$$E\{\tilde{e}^j \cdot \tilde{e}^j\} = 0 \quad \text{для} \quad i \neq j. \tag{3.27}$$

3. Дисперсия результатов наблюдений во всех точках x^l одинакова, т. е.

$$D(\tilde{y}^l) = \sigma^2 \text{ для всех } l, \tag{3.28}$$

или

$$D(\tilde{e}^l) = \sigma^2 \text{ для всех } l. \tag{3.29}$$

Условия 2 и 3 выполняются, если

$$E\{\tilde{e} \cdot \tilde{e}^T\} = \sigma^2 I, \tag{3.30}$$

где I — единичная матрица.

Наложим еще два условия на оценки $\hat{a}$, обратив внимание на то, что оценки $\hat{a}$, полученные на основе обработки случайных результатов наблюдений, представляют собой некоторый случайный вектор.

1. Оценка $\hat{a}$ не должна содержать систематических ошибок (т. е. оценка $\hat{a}$ должна быть несмещенной):

$$E\{\hat{a}\} = \bar{a}. \tag{3.31}$$

2. Дисперсия σ_l^2 оценки $\hat{a}_l$ должна быть минимальной:

$$\sigma_l^2 = D(\hat{a}_l) = D(\hat{a}_l - \bar{a}_l) = E\{(\hat{a}_l - \bar{a}_l)^2\} = \min!$$
$$\text{для } l = 0, 1, 2, \ldots, k. \tag{3.32}$$

При этом рассматривается класс оценок, образуемых линейными комбинациями результатов наблюдений $\tilde{y}^l$.

Эти два требования представляют собой только одну из многих возможных конкретизаций понятия „наилучшая оценка". Вместо второго требования могут быть использованы такие условия, как

$$\max_{0 \leqslant l \leqslant k} |\hat{a}_l - \bar{a}_l| = \min!$$

или

$$\max_{1 \leqslant l \leqslant N} |\hat{y}^l - \bar{y}^l| = \min!$$

и т. д.

Условие (3.32) является наиболее приемлемым, так как оно приводит к методу наименьших квадратов и при нормально рас-

пределенных результатах наблюдений в каждой точке x^l позволяет провести статистический анализ полученных оценок, а также проверить адекватность модели.

Теоретическую основу метода наименьших квадратов составляют следующие утверждения:

Утверждение 3.1

Оценка $\hat{a}$ удовлетворяет условиям (3.31) и (3.32), если сумма

$$S = \sum_{l=1}^{N} (\tilde{y}^l - y^l)^2 = |\tilde{Y} - Y|^2 = (\tilde{Y} - Y)'(\tilde{Y} - Y) \qquad (3.33)$$

минимальна, т. е.

$$S(\hat{a}) = \min_{a} S(a). \qquad (3.34)$$

Очевидно, что S в (3.33) является функцией a, причем в силу (3.20) можно записать, что

$$S = (\tilde{Y} - Fa)'(\tilde{Y} - Fa),$$

т. е.

$$S = S(a) = \tilde{Y}'\tilde{Y} + a'F'Fa - 2\tilde{Y}'Fa. \qquad (3.35)$$

S является расширенной квадратичной формой a_l, которая в случае невырожденности матрицы $F'F$ имеет единственный минимум при

$$\hat{a} = (F'F)^{-1}F'\tilde{Y}. \qquad (3.36)$$

Матрица $F'F$ невырожденна, т. е.

$$|F'F| \neq 0,$$

если матрица F имеет ранг $k+1$. Имеет место также

Утверждение 3.2

Если матрица F имеет ранг $(k+1)$, то сумма квадратов (3.35) достигает минимума при

$$\hat{a} = (F'F)^{-1}F'\tilde{Y} = CF'\tilde{Y}. \qquad (3.37)$$

Матрица C в (3.37) размера $(k+1)\times(k+1)$:

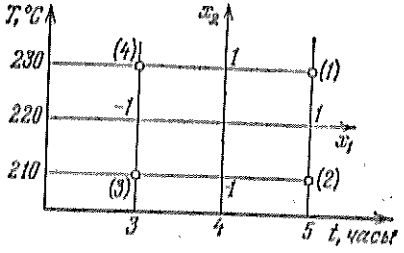
$$C = (F'F)^{-1} \qquad (3.38)$$

называется дисперсионной матрицей.

Для пояснения утверждения 3.2 рассмотрим пример.

Пример 3.1

Выход y некоторой химической реакции зависит от длительности реакции t и температуры T. Эту зависимость можно считать линейной в окрестности точки $t = 4$ ч, $T = 220$ °C.



Фиг. 3.1. План эксперимента для линейной модели примера 3.1.

Меняя t с шагом в 1 ч и температуру с шагом в 10 °C, поставим опыты в следующих точках (фиг. 3.1):

$$t = (4 \pm 1) \text{ ч}, \quad T = (220 \pm 10)\ ^\circ\text{C}.$$

В дальнейшем мы для упрощения вычислений введем нормированные переменные x_1 и x_2 (фиг. 3.1):

$$x_1 = t - 4, \quad x_2 = \frac{T - 220}{10}.$$

Для оценивания коэффициентов линейной модели

$$y = a_0 + a_1 x_1 + a_2 x_2$$

проводятся опыты в следующих точках:

i	x_1^i	x_2^i
1	1	1
2	1	−1
3	−1	−1
4	1	1

Результаты опытов (в процентах) представлены вектором [1])

$$\tilde{Y}' = (65{,}5;\ 55;\ 44{,}9;\ 55)'.$$

[1]) В целях упрощения символики здесь и ниже в ряде случаев не делается различий в обозначениях случайной величины и ее реализации. Так, оценка вектора параметров (случайная величина) и ее реализация обозначаются через $\hat{a}$, а результат наблюдения и его реализация — через $\tilde{y}$ и т. д.— *Прим. ред.*

Пользуясь формулами (3.12), (3.21), (3.37), (3.38) и (3.19), получаем следующие результаты:

$$F = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{bmatrix},$$

$$F'F = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} = 4I_3,$$

$$C = (F'F)^{-1} = \frac{1}{4} I_3,$$

$$CF' = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

$$\hat{a} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = CF'\tilde{Y} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 65{,}5 \\ 55 \\ 44{,}9 \\ 55 \end{bmatrix} = \begin{bmatrix} 55{,}1 \\ 5{,}15 \\ 5{,}15 \end{bmatrix},$$

$$\hat{y} = 55{,}1 + 5{,}15x_1 + 5{,}15x_2.$$

3.2.3. Ошибки оценивания

Оценки $\hat{a}_i$, рассчитанные согласно (3.37), отличаются от истинных значений коэффициентов, причем ошибка тем больше, чем больше дисперсия ошибок наблюдений. Показателями точности оценок $\hat{a}_i$ и величины $\hat{y}$ являются дисперсии σ^2_i и $\sigma^2_{\hat{y}}$ соответственно. Эти дисперсии зависят не только от дисперсии ошибок наблюдений σ^2, но и от выбранной структуры модели и точек постановки опытов, т. е. от матрицы F (см. выражение 3.21). Для определения этой зависимости необходимо прежде всего найти выражение для ковариационной матрицы $\operatorname{cov}(\hat{a})$:

$$\operatorname{cov}(\hat{a}) = \| E\{(\hat{a}_i - \bar{a}_i)(\hat{a}_j - \bar{a}_j)\} \| = E\{(\hat{a} - \bar{a})(\hat{a} - \bar{a})'\}.$$

Из (3.37), (3.31) и (3.25) следуют выражения

$$\hat{a} = CF'\tilde{Y},$$
$$\bar{a} = E\{\hat{a}\} = CF'E\{\tilde{Y}\} = CF'\bar{Y}.$$

Используя эти выражения, получаем

$$E\{(\hat{a} - \bar{a})(\hat{a} - \bar{a})'\} = E\{CF'(\tilde{Y} - \bar{Y})(\tilde{Y} - \bar{Y})'FC'\} =$$
$$= CF'E\{(\tilde{Y} - \bar{Y})(\tilde{Y} - \bar{Y})'\}FC'.$$

В силу (3.23), (3.24) и (3.30) имеем

$$\operatorname{cov}(\hat{a}) = CF'E\{(\hat{Y}-\bar{Y})(\hat{Y}-\bar{Y})'\}FC' = CF'FC'\sigma^2.$$

Так как матрица $F'F$ симметрична, то

$$CF'FC' = (F'F)^{-1}F'F(F'F)^{-1} = C,$$

и для $\operatorname{cov}(\hat{a})$ получаем

$$\operatorname{cov}(\hat{a}) = C\sigma^2. \tag{3.39}$$

При этом, в частности, для дисперсии σ_l^2 оценки $\hat{a}_l$ имеем

$$\sigma_l^2 = c_{ll}\sigma^2,$$

а для коэффициента корреляции между оценками $\hat{a}_l$ и $\hat{a}_j$

$$\rho_{lj}^2 = \frac{c_{lj}^2}{c_{ll}c_{jj}}.$$

Дисперсию $\sigma_{\hat{y}}^2$ для величины $\hat{y}$ получаем с помощью следующих преобразований:

$$\begin{aligned}\sigma_{\hat{y}}^2 &= E\{(\hat{y}-\bar{y})^2\} = E\{((\hat{a}-\bar{a})'f(x))^2\} = \\ &= E\{f'(x)(\hat{a}-\bar{a})(\hat{a}-\bar{a})'f(x)\} = \\ &= f'(x)E\{(\hat{a}-\bar{a})(\hat{a}-\bar{a})'\}f(x) = \\ &= f'(x)\operatorname{cov}(\hat{a})f(x) = f'(x)C\sigma^2 f(x).\end{aligned}$$

Таким образом,

$$\sigma_{\hat{y}}^2 = f'(x)Cf(x)\sigma^2.$$

Полученные результаты позволяют сформулировать утверждение, которое, кстати, поясняет, почему матрица C была названа дисперсионной.

Утверждение 3.3

При предположениях разд. 3.2.2 и при известной дисперсии ошибки наблюдений σ^2 дисперсия σ_l^2 оценки $\hat{a}_l$ рассчитывается по формуле

$$\sigma_l^2 = c_{ll}\sigma^2. \tag{3.40}$$

Коэффициент корреляции между $\hat{a}_l$ и $\hat{a}_j$ определяется согласно выражению

$$\rho_{lj}^2 = \frac{c_{lj}^2}{c_{ll}c_{jj}}, \tag{3.41}$$

а дисперсия $\sigma_{\hat{y}}^2$ величины $\hat{y}$ по формуле

$$\sigma_{\hat{y}}^2 = f'(x)Cf(x)\sigma^2. \tag{3.42}$$

Ковариационная матрица $\hat{a}$ имеет вид

$$\operatorname{cov}(\hat{a}) = C\sigma^2. \tag{3.43}$$

При дополнительном предположении о нормальности закона распределения результатов наблюдений в любой точке постановки эксперимента величина

$$X = \frac{\hat{a}_l - \bar{a}_l}{\sigma_l} \tag{3.44}$$

распределена нормально и имеет место соотношение

$$P(|X| \leqslant \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon), \tag{3.45}$$

где $\Phi(\varepsilon)$—функция Лапласа (см. разд. 15.2).

Если величина ε задана, то с помощью (3.45) можно рассчитать вероятность того, что $|X| \leqslant \varepsilon$. Однако обычно задается вероятность $P = P\{|X| \leqslant \varepsilon\}$ и из (3.45) находят соответствующую величину $\varepsilon = \varepsilon(P)$. При этом используют табл. 15.1. Например, для $P = 0{,}95$ получаем $\varepsilon = 1{,}96$ [1]. Таким образом, имеет место следующее неравенство:

$$|\hat{a}_l - \bar{a}_l| \leqslant \sigma_l \varepsilon = \varepsilon \sqrt{c_{ll}}\sigma. \tag{3.46}$$

Если известна дисперсия наблюдений σ^2, выражение (3.46) может быть использовано при построении доверительного интервала для истинного значения l-го коэффициента при заданной доверительной вероятности P. Однако в общем случае дисперсия ошибок наблюдений σ^2 не известна и должна быть оценена с помощью полученных экспериментальных данных. При этом может быть использована остаточная сумма квадратов $S(\hat{a})$ [см. (3.33) и (3.34)]:

$$S(\hat{a}) = S_R = \sum_{l=1}^{N} (\tilde{y}^l - \hat{y}^l)^2.$$

Эта сумма квадратов имеет

$$\varphi = N - k - 1 \tag{3.47}$$

степеней свободы [N слагаемых, между которыми существует $k+1$ линейная связь, определяемая системой уравнений (3.37)]. Величина

$$s^2 = \frac{S_R}{\varphi} \tag{3.48}$$

[1] Величина ε определяется из условия $\bar{\Phi}(\varepsilon) = P/2$, где $\bar{\Phi}(\varepsilon) = \int_0^{\varepsilon} \varphi(x)\,dx$, $\varphi(x)$—функция плотности нормированного нормального распределения. Функция $\bar{\Phi}(\varepsilon)$ табулирована (см. табл. 15.1). Отметим также, что иногда функцией Лапласа называют функцию $\bar{\Phi}(\varepsilon)$, а не $\Phi(\varepsilon) = 0{,}5 + \bar{\Phi}(\varepsilon)$.— *Прим. ред.*

является несмещенной оценкой дисперсии ошибок наблюдений σ^2 [1]. Оценка s_l^2 для дисперсии σ_l^2 вычисляется аналогично (3.40) с помощью выражения

$$s_l^2 = c_{ll} s^2. \tag{3.49}$$

Эта формула справедлива и для любой иной несмещенной оценки дисперсии σ^2.

Величина

$$t = \frac{\hat{a}_l - \bar{a}_l}{s_l} \tag{3.50}$$

подчиняется t-распределению (распределению Стьюдента) с φ степенями свободы. При этом справедливо соотношение

$$P\{|t| < e\} = S_\varphi(e) - S_\varphi(-e). \tag{3.51}$$

Если задана вероятность $P = P\{|t| < e\}$, то величину $e = e(P)$ в (3.51) можно найти с помощью табл. 15.3. Например, при вероятности $P = 0{,}9$ и при числе степеней свободы $\varphi = 1$ получаем $e = 6{,}3$. С вероятностью P имеем

$$|\hat{a}_l - \bar{a}_l| < s_l e = e\sqrt{c_{ll}}\, s. \tag{3.52}$$

Выражение (3.52) определяет доверительный интервал для истинного значения l-го коэффициента в случае, когда известна лишь оценка s^2 дисперсии наблюдений.

Если в (3.52) $e\sqrt{c_{ll}}\, s > |\hat{a}_l|$, то нуль-гипотеза $H_0: \bar{a}_l = 0$ принимается и коэффициент a_l считается незначимым. При проверке значимости коэффициентов величину e, соответствующую выбранному уровню значимости $\alpha = 1 - P$ и числу степеней свободы φ, мы будем обозначать через $t_{кр}$ и называть критическим значением распределения Стьюдента. Заметим, что в общем случае оценка $\hat{a}_l$ и ее дисперсия зависят от оценок всех других коэффициентов.

Если тот или иной член исключается из уравнения (например, из-за незначимости соответствующего коэффициента), необходимо пересчитать оценки всех остальных коэффициентов и их дисперсии. При этом могут измениться как доверительные интервалы для коэффициентов, так и выводы относительно коэффициентов значимости.

---

[1] Величина s_l^2, определяемая выражением (3.48), является несмещенной оценкой дисперсии ошибок наблюдений при условии, что модель является адекватной, т. е. выбранный вид модели совпадает с истинным характером зависимости. О проверке адекватности см. ниже (разд. 3.2.5). — *Прим. ред.*

Пример 3.2

Продолжим теперь рассмотрение примера 3.1. Рассчитаем прежде всего

$$\hat{Y} = F\hat{a} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 55{,}1 \\ 5{,}15 \\ 5{,}15 \end{bmatrix} = \begin{bmatrix} 65{,}4 \\ 55{,}1 \\ 44{,}8 \\ 55{,}1 \end{bmatrix}.$$

Найдем остаточную сумму квадратов S_R:

$$S_R = 4\,(0{,}1)^2 = 0{,}04.$$

Мы имеем $N = 4$ эксперимента и $k + 1 = 3$ оцениваемых параметра a_0, a_1, a_2. Тогда число степеней свободы равно $\varphi = 1$. В силу (3.48) получаем

$$s = \sqrt{0{,}04} = 0{,}2.$$

В данном случае $c_{ii} = 1/4$ и

$$s_i = \sqrt{c_{ii}} \cdot s = 0{,}1.$$

При доверительной вероятности $P = 0{,}9$ и $\varphi = 1$ находим $\varepsilon = 6{,}3$ и $|\hat{a}_i - \bar{a}_i| < 0{,}63$.

Таким образом, с 90%-ной надежностью имеем

$$\bar{a}_0 = 55{,}1 \pm 0{,}63,$$
$$\bar{a}_1 = \bar{a}_2 = 5{,}15 \pm 0{,}63.$$

3.2.4. Ошибки при выборе вида модели

Результаты, относящиеся к методу наименьших квадратов и изложенные в разд. 3.2.2 и 3.2.3, были получены в предположении, что модель вида (3.11) является адекватной. Метод, позволяющий проверить, можно ли некоторую модель рассматривать как адекватную, будет рассмотрен ниже в разд. 3.2.5. Здесь же мы проанализируем лишь последствия, возникающие при неправильном выборе вида модели.

Рассмотрим случай, когда модель имеет следующий вид:

$$\bar{y} = \bar{a}_0 f_0(x) + \bar{a}_1 f_1(x) + \ldots + \bar{a}_k f_k(x) + \bar{b}_1 g_1(x) + \ldots + \bar{b}_l g_l(x). \quad (3.53)$$

Здесь $\bar{a}_i$ и $\bar{b}_i$ — истинные значения параметров модели, а $f_i(x)$ и $g_i(x)$ — известные функции независимых переменных $x_1, x_2, \ldots, x_n$. Выражение (3.53) можно записать в форме

$$\bar{y} = \bar{a}' f(x) + \bar{b}' g(x), \quad (3.54)$$

где $\bar{a}$ и $f(x)$ определены выше [см. (3.8) и (3.10)], а $\bar{b}$ и $g(x)$ задаются следующим образом:

$$\bar{b} = (\bar{b}_1, \bar{b}_2, \ldots, \bar{b}_l)', \tag{3.55}$$

$$g(x) = (g_1(x), g_2(x), \ldots, g_l(x))'. \tag{3.56}$$

При этом для вектора $\bar{Y}$ истинных значений выхода имеем

$$\bar{Y} = F\bar{a} + G\bar{b}, \tag{3.57}$$

где

$$G = (g_j(x^i)) = \begin{bmatrix} g_1(x^1) & g_2(x^1) & \ldots & g_l(x^1) \\ g_1(x^2) & g_2(x^2) & \ldots & g_l(x^2) \\ \ldots & \ldots & \ldots & \ldots \\ g_1(x^N) & g_2(x^N) & \ldots & g_l(x^N) \end{bmatrix}. \tag{3.58}$$

Если теперь вектор параметров $\bar{a}$ заменить на вектор оценок $\hat{a}$, рассчитанный согласно (3.36) в предположении адекватности модели (3.11), то выражение $E(\hat{a}) = \bar{a}$ и соответственно $E(\hat{y}) = \bar{y}$ уже не будут справедливы. Для $E(\hat{a})$ в этом случае получаем

$$\begin{aligned} E(\hat{a}) &= E(CF'\bar{Y}) = CF'E(\bar{Y}) = CF'\bar{Y} = \\ &= CF'(F\bar{a} + G\bar{b}) = (F'F)^{-1}F'F\bar{a} + CF'G\bar{b} = \\ &= \bar{a} + CF'G\bar{b} = \bar{a} + A\bar{b}. \end{aligned} \tag{3.59}$$

Матрицу

$$A = CF'G \tag{3.60}$$

будем называть матрицей смещения. Эта матрица характеризует смещение в оценках коэффициентов. Исследование ошибок, вызванных смещением оценок, имеет большое значение для построения рациональных экспериментальных планов.

Если бы, например, функции $f(x)$ были заданы согласно (3.12), а истинная модель имела вид (3.54), причем часть $\bar{b}'g(x)$ содержала бы нелинейные функции x, то эту часть можно было бы исключить из рассмотрения, если в (3.59)

$$E(\hat{a}_i) = \bar{a}_i \text{ для } i = 1, 2, \ldots, n.$$

В этом случае ошибка в выборе вида модели не сказывается на математических ожиданиях оценок коэффициентов.

В общем случае можно поставить вопрос о том, когда условие $E(\hat{a}_i) = \bar{a}_i$ справедливо и когда нет.

Чтобы ответить на этот вопрос, введем прежде всего следующие обозначения:

$$H = F'G = (h^1, h^2, \ldots, h^l). \tag{3.61}$$

i-й вектор-столбец матрицы $H = F'G$ будем обозначать через h^i. Положим далее

$$\bar{c} = A\bar{b}. \tag{3.62}$$

В силу того, что

$$A = CF'G = CH, \tag{3.63}$$

можно записать

$$\bar{c} = C(h^1\bar{b}_1 + h^2\bar{b}_2 + \ldots + h^l\bar{b}_l). \tag{3.64}$$

Отсюда следует, что величина $E(\hat{a})$ не зависит от $\bar{b}_l$, если $Ch^l = 0$.

Так как C — невырожденная матрица, можно сформулировать

Утверждение 3.4

Оценка $\hat{a}_l$ является несмещенной в случае неправильного выбора вида модели, если l-й вектор-столбец матрицы $H = F'G$ равен нулю.

Пример 3.3

Продолжим рассмотрение примеров 3.1 и 3.2, приняв, что линейное описание

$$y = a_0 + a_1x_1 + a_2x_2$$

недостаточно и истинная модель представляется полиномом второй степени

$$y = a_0 + a_1x_1 + a_2x_2 + b_1x_1x_1 + b_2x_2x_2 + b_3x_1x_2.$$

В этом случае для (3.56) получаем

$$g(x) = (x_1x_2,\ x_2x_3,\ x_1x_2)',$$

а матрица G принимает вид

$$G = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

В силу (3.61) имеем

$$G = F'G = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 4 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

а из (3.63) получаем

$$A = CH = \frac{1}{4}IH = \frac{1}{4}H = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Следовательно, согласно (3.59),

$$E(\hat{a}_0) = \bar{a}_0 + \bar{b}_1 + \bar{b}_2,$$
$$E(\hat{a}_1) = \bar{a}_1, \quad E(\hat{a}_2) = \bar{a}_2.$$

Несмотря на то что линейная модель не адекватна, оценки $\hat{a}_1$ и $\hat{a}_2$ получаются несмещенными и, следовательно, несмещенными являются оценки градиента функции $y(a, x)$ в точке $x = 0$.

3.2.5. Проверка адекватности модели

После того как в разд. 3.2.4 были проанализированы последствия неправильного выбора вида модели, обсудим теперь метод, который дал бы нам возможность принимать решения о том, адекватной или неадекватной является принятая модель. Основополагающие идеи этого метода уже были изложены выше в гл. 2.

Для проверки гипотезы об адекватности модели необходимо сопоставить достигнутую точность модели с величиной, характеризующей точность наблюдений. Если ошибки, характеризующие точность модели, превосходят ошибки наблюдений, то гипотеза об адекватности модели отклоняется. В этом случае уже нельзя оценивать ошибку наблюдений путем нахождения разности между результатом наблюдения выходной переменной и результатом ее расчета по модели, так как в случае неправильного выбора вида модели определяемая по модели величина $\hat{y}^l$ уже не может служить достаточно хорошей оценкой среднего значения наблюдений, поскольку

$$E(\hat{y}^l) \neq \bar{y}^l.$$

Поэтому дисперсия ошибок наблюдений может быть оценена лишь путем сравнения результатов нескольких параллельных опытов, проведенных в каждой экспериментальной точке. Мы опишем методику проверки адекватности, полагая, что в каждой из N точек x^l реализуется ν экспериментов. Результаты этих экспериментов для каждой точки x^l представляются рядом

$$\tilde{y}^{l1}, \tilde{y}^{l2}, \ldots, \tilde{y}^{l\nu}.$$

Для расчета оценок коэффициентов мы будем использовать средние значения $\tilde{y}^l$ ряда наблюдений для каждой точки x^l, определяемые согласно (3.17). Для проверки гипотезы адекватности необходимо сравнить две суммы квадратов:

1) сумму квадратов, характеризующую неадекватность (дефект) модели (Defektquadratsumme),

$$S_D = \sum_{l=1}^{N} \nu (\tilde{y}^l - \hat{y}^l)^2 = \nu S_R. \tag{3.65}$$

Эта сумма зависит от разности между рассчитанными по модели и наблюдаемыми значениями выходной переменной;

2) сумму квадратов, характеризующую ошибки наблюдений (Fehlerquadratsumme),

$$S_e = \sum_{l=1}^{N} \sum_{j=1}^{\nu} (\tilde{y}^{lj} - \bar{y}^l)^2. \tag{3.66}$$

Сумма S_D состоит из N слагаемых, между которыми имеет место $(k+1)$ линейных связей, определяемых выражением (3.37). Поэтому с S_D связано

$$\varphi_1 = N - k + 1 \tag{3.67}$$

степеней свободы.

Сумма S_e состоит из νN слагаемых, между которыми, согласно (3.17), существует N линейных связей. При этом S_e имеет

$$\varphi_2 = N\nu - N = N(\nu - 1) \tag{3.68}$$

степеней свободы.

Для оценки s^2 дисперсии ошибок наблюдений σ^2 выражение (3.48) в силу приведенных выше соображений не может быть использовано. Оценка s^2 получается теперь с помощью суммы квадратов S_e по формуле

$$s^2 = \frac{S_e}{\nu\varphi_2}. \tag{3.69}$$

Заметим, что s^2 в (3.69) является оценкой дисперсии величины $\bar{y}$, которая представляет собой среднее по ν параллельным наблюдениям. Это означает, что s^2 в данном случае не является оценкой дисперсии ошибки единичного наблюдения, а определяет величину дисперсии среднего, рассчитанного по ν наблюдениям. Легко видеть, что оценка дисперсии s_e^2 ошибки единичного наблюдения равна νs^2 [1]).

Перейдем теперь к проверке адекватности. Частное от деления оценки дисперсии неадекватности на оценку дисперсии ошибки единичного наблюдения

$$F = \frac{S_D/\varphi_1}{S_e/\varphi_2}, \tag{3.70}$$

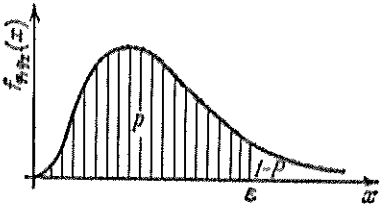
в случае, когда модель адекватна, является случайной величиной, подчиненной F-распределению с числами степеней свободы

[1]) Ниже через s^2 всюду обозначается оценка дисперсии ошибок наблюдений, причем результатом наблюдения считается значение выходной переменной в точке плана, используемое при расчетах по методу наименьших квадратов. Таким образом, если для каждой точки плана проводится усреднение по ν параллельным опытам, дисперсия ошибок наблюдений есть дисперсия среднего по ν параллельным опытам.— *Прим. ред.*

φ_1 и φ_2. При этом можно определить значение $F_{кр}$, соответствующее условию

$$P(F > F_{кр}) = 1 - P = \alpha, \tag{3.71}$$

где α — заданный уровень значимости проверки гипотезы об адекватности. Задавшись надежностью $P = P(F < F_{кр})$ (обычно P выбирают равной 0,95 или 0,99), значение $F_{кр}$ можно найти с помощью табл. 15.4 или 15.5. Например, при $\varphi_1 = 1$ и $\varphi_2 = 4$



Фиг. 3.2. Вид функции плотности распределения F-критерия.

для $P = 0{,}99$ из табл. 15.5 находим $F_{кр} = 21{,}2$; при $\varphi_1 = 3$ и $\varphi_2 = 9$ для $P = 0{,}99$ получаем $F_{кр} = 6{,}99$.

Для иллюстрации процедуры проверки адекватности с помощью F-критерия на фиг. 3.2 показан вид функции $f_{\varphi_1\varphi_2}(x)$ плотности распределения величины F. Заштрихованная часть имеет площадь, равную P, а $\varepsilon = F_{кр}$. Если рассчитанное согласно (3.70) значение F лежит левее $F_{кр}$, т. е. находится в допустимой области, гипотеза об адекватности принимается. Если F попадает в область, находящуюся правее $F_{кр}$, т. е. в область, вероятность попадания в которую равна $\alpha = 1 - P$ (обычно $\alpha = 0{,}05$ или $0{,}01$), гипотеза об адекватности отклоняется.

Способ проверки адекватности для другой ситуации (когда число опытов в точках x^l неодинаково, а матрица $\tilde{Y}$ содержит результаты каждого наблюдения, а не их средние значения) будет описан в разд. 5.6 [1].

Пример 3.4

Для пояснения процедуры проверки адекватности модели вернемся вновь к примерам 3.1 и 3.2, сняв, разумеется, предположение о том, что линейная модель является адекватной. Чтобы проверить адекватность линейной модели, необходимо добавить по крайней мере еще по одному эксперименту в каждой точке плана примера 3.1. После проведения этих опытов получим следующие результаты:

[1] См. по этому вопросу Горский В. Г., Адлер Ю. П., *Заводск. лаб.*, № 3 (1971). — *Прим. ред.*

t	$\tilde{y}^{(1)}$	$\tilde{y}^{(2)}$	$\tilde{y}'=\frac{1}{2}(\tilde{y}^{(1)}+\tilde{y}^{(2)})$
1	65,5	65,6	65,55
2	55	55,2	55,1
3	44,9	45	44,95
4	55	54,8	54,9

Здесь $N=4$, $k=2$, $\nu=2$ и в соответствии с выражениями (3.67) и (3.68) $\varphi_1=1$, $\varphi_2=4$. При новых значениях $\tilde{Y}$ рассчитываем, согласно (3.37), оценки

$$\hat{a}=\begin{bmatrix}55,125\\5,2\\5,1\end{bmatrix}.$$

При этом

$$\hat{y}=55,125+5,2x_1+5,1x_2$$

и

$$\hat{Y}=(65,425;\ 55,225;\ 44,825;\ 55,025)'.$$

Теперь можно рассчитать суммы квадратов S_D и S_e по формулам (3.65) и (3.66):

$$S_D=0,125,\quad S_e=0,05.$$

Согласно (3.69),

$$s^2=\frac{S_e}{\nu\varphi_2}=0,056$$

и для среднеквадратичных ошибок оценок коэффициентов получаем

$$s_l=\frac{s}{2}=0,028.$$

Для 90%-ной надежности при $\varphi_2=4$ из табл. 15.3 получаем $\varepsilon=2,1$ и соответственно

$$|\hat{a}_l-\bar{a}_l|<0,028\times 2,1=0,06.$$

В данном случае получаются существенно более точные оценки, чем в примере 3.2. Для проверки адекватности примем

$$P(F>F_{кр})=1-0,99=0,01.$$

Для $\varphi_1=1$ и $\varphi_2=4$ выше с помощью табл. 15.5 было найдено значение

$$F_{кр}=21,2.$$

Согласно (3.70), вычисляем

$$F=0,125/0,0125=10<F_{кр}=21,2.$$

Таким образом, результаты не противоречат предположению об адекватности модели.

Пример 3.5

Введем нормированные переменные

$$x_1 = t - 8, \quad x_2 = \frac{T - 260}{10}$$

и выберем модель в виде квадратичного уравнения [см. формулы (3.4) и (3.13)]:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 (x_1)^2 + a_4 (x_2)^2 + a_5 x_1 x_2.$$

Линейная модель здесь уже оказывается недостаточной, так как целевая функция в окрестности оптимума имеет существенную кривизну. В каждой из указанных точек поставим по два эксперимента:

t	x^t		$\tilde{y}^{t1}$	$\tilde{y}^{t2}$	$\tilde{y}^t$
1	0	0	88	88,2	88,1
2	1	0	89,2	89,1	89,15
3	0	1	82,2	82,1	82,15
4	−1	0	83,7	83,8	83,75
5	0	−1	87,3	87,4	87,35
6	1	1	82,6	82,7	82,65
7	1	−1	89,6	89,6	89,6
8	−1	−1	82,2	82,1	82,15
9	−1	1	79,3	79,1	79,2

Здесь $N = 9$, $\nu = 2$, $k = 5$, $\varphi_1 = 3$, $\varphi_2 = 9$. Выберем 95%-ную надежность для t-критерия, а для F-критерия положим $P(F < F_{кр}) = 0{,}99$. При этом соответствующие значения ε и $F_{кр}$ (см. табл. 15.3 и 15.5) оказываются равными

$$\varepsilon = 2{,}26,$$
$$F_{кр} = 6{,}99.$$

Согласно (3.13), (3.21), (3.37) и (3.38), находим

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix},$$

$$F'F = \begin{bmatrix} 9 & 0 & 0 & 6 & 6 & 0 \\ 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 6 & 0 & 0 & 6 & 4 & 0 \\ 6 & 0 & 0 & 4 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix},$$

$$C = (F'F)^{-1} = \begin{bmatrix} 5/9 & 0 & 0 & -1/3 & -1/3 & 0 \\ 0 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/6 & 0 & 0 & 0 \\ -1/3 & 0 & 0 & 1/2 & 0 & 0 \\ -1/3 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/4 \end{bmatrix},$$

$$CF' = \begin{bmatrix} 5/9 & 2/9 & 2/9 & 2/9 & 2/9 & -1/9 & -1/9 & -1/9 & -1/9 \\ 0 & 1/6 & 0 & -1/6 & 0 & 1/6 & 1/6 & 1/6 & -1/6 \\ 0 & 0 & 1/6 & 0 & -1/6 & 1/6 & -1/6 & -1/6 & 1/6 \\ -1/3 & -1/6 & 1/3 & -1/6 & -1/3 & 1/6 & 1/6 & 1/6 & 1/6 \\ -1/3 & -1/3 & 1/6 & -1/3 & +1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1/4 & -1/4 & 1/4 & -1/4 \end{bmatrix},$$

$$\hat{a} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \end{bmatrix} = CF'\tilde{Y} = \begin{bmatrix} 87{,}97 \\ 2{,}717 \\ -2{,}517 \\ -1{,}45 \\ -3{,}15 \\ -1 \end{bmatrix}.$$

Таким образом,

$$\hat{y} = 87{,}97 + 2{,}717x_1 - 2{,}517x_2 - 1{,}45\,(x_1)^2 - 3{,}15\,(x_2)^2 - x_1x_2$$

и

$$\hat{Y} = (87{,}97;\ 89{,}24;\ 82{,}3;\ 83{,}8;\ 87{,}34;\ 82{,}57;\ 89{,}6;\ 82{,}17;\ 79{,}14)'.$$

Суммы квадратов S_D и S_e рассчитываем по формулам (3.65) и (3.66)

$$S_D = 0{,}121, \qquad S_e = 0{,}07.$$

Далее, согласно (3.69),

$$s^2 = \frac{S_e}{\nu\varphi_0} = 0{,}0039.$$

Для дисперсий оценок коэффициентов, согласно (3.49), получаем

$$s_0^2 = (5/9)\, s^2 = 0{,}00217,$$
$$s_1^2 = (1/6)\, s^2 = 0{,}00065,$$
$$s_2^2 = (1/6)\, s^2 = 0{,}00065,$$
$$s_3^2 = (1/2)\, s^2 = 0{,}00195,$$
$$s_4^2 = (1/2)\, s^2 = 0{,}00195,$$
$$s_5^2 = (1/4)\, s^2 = 0{,}01.$$

В силу (3.52) с 95%-ной надежностью имеем

$$|\hat{a}_0 - \bar{a}_0| < s_0 \varepsilon = 0{,}105,$$
$$|\hat{a}_1 - \bar{a}_1| < s_1 \varepsilon = 0{,}0576,$$
$$|\hat{a}_2 - \bar{a}_2| < s_2 \varepsilon = 0{,}0576,$$
$$|\hat{a}_3 - \bar{a}_3| < s_3 \varepsilon = 0{,}1,$$
$$|\hat{a}_4 - \bar{a}_4| < s_4 \varepsilon = 0{,}1,$$
$$|\hat{a}_5 - \bar{a}_5| < s_5 \varepsilon = 0{,}0715.$$

Такую точность можно считать достаточной. Из (3.70) получаем

$$F = \frac{S_D/\varphi_1}{S_e/\varphi_2} = \frac{0{,}0403}{0{,}0078} = 5{,}2 < 6{,}99.$$

Отсюда следует вывод об адекватности квадратичной модели.

3.3. Краткое изложение основных положений метода наименьших квадратов

Принимается, что целевая величина y зависит от n независимых переменных x_i, которые представляются вектором

$$x = (x_1, x_2, \ldots, x_n)'. \tag{3.72}$$

С помощью эксперимента требуется получить приближенное описание функции

$$y = y(x). \tag{3.73}$$

3.3.1. Выбор структуры модели

Предполагается, что модель имеет следующий вид:

$$y = a' f(x). \tag{3.74}$$

Здесь

$$a = (a_0, a_1, \ldots, a_k)' \tag{3.75}$$

— вектор неизвестных параметров модели, а

$$f(x) = (f_0(x), f_1(x), \ldots, f_k(x))' \tag{3.76}$$

— вектор заданных функций $f_i(x)$.

3.3.2. Выбор экспериментальных точек

В m-мерном пространстве Ω_x переменных x_l выбираем $N > k+1$ точек

$$x^1, x^2, \ldots, x^N,$$

где

$$x^t = (x_1^t, x_2^t, \ldots, x_n^t)', \quad t = 1, 2, \ldots, N. \tag{3.77}$$

Тогда можно вычислить матрицу

$$F = (f_j(x^t)) = \begin{bmatrix} f_0(x^1) & f_1(x^1) & \ldots & f_k(x^1) \\ f_0(x^2) & f_1(x^2) & \ldots & f_k(x^2) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ f_0(x^N) & f_1(x^N) & \ldots & f_k(x^N) \end{bmatrix}. \tag{3.78}$$

Экспериментальные точки должны быть расположены так, чтобы матрица F имела ранг $k+1$. В этом случае существует матрица

$$C = (F'F)^{-1}, \tag{3.79}$$

которую будем называть дисперсионной матрицей.

3.3.3. Проведение эксперимента

В каждой экспериментальной точке x^t проводятся ν опытов, по результатам которых

$$\tilde{y}^{t1}, \quad \tilde{y}^{t2}, \quad \ldots, \quad \tilde{y}^{t\nu}$$

рассчитывается среднее значение

$$\tilde{y}^t = \nu^{-1} \sum_{l=1}^{l} \tilde{y}^{tl}. \tag{3.80}$$

Из средних значений составляется вектор результатов экспериментов

$$\tilde{Y} = (\tilde{y}^1, \tilde{y}^2, \ldots, \tilde{y}^N)'. \tag{3.81}$$

При этом предполагается, что результаты наблюдений независимы, распределены нормально и имеют одну и ту же дисперсию σ^2. Кроме того, предполагается, что математические ожидания результатов наблюдений равны истинным значениям целевой величины.

3.3.4. Оценивание параметров модели

Оценки $\hat{a}$ коэффициентов модели рассчитываются по методу наименьших квадратов в соответствии с формулой

$$\hat{a} = CF'\hat{Y}. \tag{3.82}$$

При этом для дисперсии оценок имеем

$$\sigma_l^2 = c_{ll}\sigma^2 \tag{3.83}$$

(через c_{ll} обозначен l-й элемент главной диагонали дисперсионной матрицы C),

$$\operatorname{cov}\{\hat{a}\} = C\sigma^2, \tag{3.84}$$

$$\sigma_{\hat{y}}^2 = E\{(\hat{y}-\bar{y})^2\} = f'(x)\,C\,f(x)\,\sigma^2. \tag{3.85}$$

С помощью оценок $\hat{a}$ определяется оценка функции y:

$$\hat{y} = \hat{a}'f(x). \tag{3.86}$$

При этом можно рассчитать значения выходной переменной в точках x^l:

$$\hat{y}^l = \hat{a}'f(x^l). \tag{3.87}$$

Рассчитанные значения выхода представляются с помощью вектора

$$\hat{Y} = (\hat{y}^1, \hat{y}^2, \ldots, \hat{y}^N)'. \tag{3.88}$$

3.3.5. Вычисление сумм квадратов и оценка дисперсий

3.3.5.1. Оценка дисперсии в случае, когда в каждой точке проводится один эксперимент

Оценка дисперсии ошибок наблюдений вычисляется с помощью остаточной суммы квадратов

$$S_R = \sum_{l=1}^{N} (\tilde{y}^l - \hat{y}^l)^2 = (\tilde{Y}-\hat{Y})'(\tilde{Y}-\hat{Y}) \tag{3.89}$$

с числом степеней свободы

$$\varphi = N - k - 1 \tag{3.90}$$

по формуле

$$s^2 = \frac{S_R}{\varphi}. \tag{3.91}$$

Отсюда получают оценки s_l^2 дисперсий σ_l^2 величин $\hat{a}_l$ по формуле

$$s_l^2 = c_{ll}s^2, \quad l = 0, 1, 2, \ldots, k. \tag{3.92}$$

3.3.5.2. Оценка дисперсии в случае, когда в каждой точке проводится ν экспериментов

Оценка дисперсии ошибок наблюдений вычисляется с помощью суммы квадратов ошибок S_e

$$S_e = \sum_{l=1}^{N} \sum_{j=1}^{\nu} (\tilde{y}^{lj} - \tilde{y}^{l})^2 \tag{3.93}$$

с числом степеней свободы

$$\varphi_2 = N(\nu - 1) \tag{3.94}$$

по формуле

$$s^2 = \frac{S_e}{\nu \varphi_2}. \tag{3.95}$$

Оценки s_l^2 в этом случае получают также по формуле (3.92).

3.3.5.3. Вычисление суммы квадратов, характеризующей неадекватность модели (Defektquadratsumme), и значения F-критерия

Сумма квадратов S_D, характеризующая неадекватность модели, рассчитывается по формуле

$$S_D = \sum_{l=1}^{N} \nu (\tilde{y}^{l} - \hat{y}^{l})^2 \tag{3.96}$$

и имеет

$$\varphi_1 = N - k - 1 \tag{3.97}$$

степеней свободы.

Величину F-критерия находят по формуле

$$F = \frac{S_D/\varphi_1}{S_e/\varphi_2}. \tag{3.98}$$

3.3.6. Статистический анализ точности

3.3.6.1. Дисперсия наблюдений σ^2 известна и модель (3.74) адекватна

Доверительный интервал для коэффициентов с доверительной вероятностью P определяется следующим образом:

$$|\hat{a}_l - \bar{a}_l| < \sigma_l \varepsilon. \tag{3.99}$$

Значение ε находится из табл. 15.1 при условии

$$\Phi(\varepsilon) = \frac{P}{2}. \tag{3.100}$$

3.3.6.2. Дисперсия ошибки наблюдений σ^2 не известна и в каждой точке плана проводится один эксперимент

Доверительный интервал для коэффициентов с вероятностью P определяется из выражения

$$|\hat{a}_l - \bar{a}_l| < s_l e. \tag{3.101}$$

s_l^2 есть оценка дисперсии $\hat{a}_l$, рассчитываемая по формуле (3.92). e находится из табл. 15.2 при $\varphi = N - (k+1)$ степенях свободы и заданной доверительной вероятности P.

3.3.6.3. Дисперсия ошибки наблюдений σ^2 не известна и в каждой точке плана проводится несколько экспериментов

В этом случае также применяется t-распределение, но в качестве оценки дисперсии наблюдений используется величина s^2, рассчитанная по формуле (3.95) и имеющая φ_2 степеней свободы [см. (3.94)]. Все изложенное в разд. 3.3.6.2 остается в силе, только φ заменяется на φ_2.

3.3.7. F-критерий

В случае нескольких опытов в каждой точке экспериментального плана по формуле (3.98) можно получить величину F. Эта величина подчиняется F-распределению с числами степеней свободы φ_1 и φ_2.

Для заданной надежности $P = P\{F < F_{кр}\}$ с помощью табл. 15.4 или 15.5 можно определить соответствующее значение $F_{кр}$. Если $F \leqslant F_{кр}$, модель (3.74) считается адекватной; если $F > F_{кр}$, гипотеза об адекватности модели отклоняется.

3.3.8. Проверка значимости коэффициентов

Коэффициент a_l считается значимо отличающимся от нуля, если

$$|\hat{a}_l| > t_{кр} s_l, \tag{3.102}$$

где $t_{кр}$—критическое значение распределения Стьюдента для заданного уровня значимости α и φ степеней свободы (если оценка s_l^2 имеет φ степеней свободы); $t_{кр}$ находится с помощью табл. 15.2 из условия

$$P(|t| > t_{кр}) = \alpha. \tag{3.103}$$

ЛИТЕРАТУРА

1. Müller P. H. (Hrsg.), Wahrscheinlichkeitsrechnung und mathematische Statistik, Lexikon, Akademie-Verlag, Berlin, 1970.
2. Линник Ю. В., Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений, Физматгиз, М., 1962.
3. Налимов В. В. Чернова Н. А., Статистические методы планирования экстремальных экспериментов, изд-во „Наука", М., 1965.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Бородюк В. П., Лецкий Э. К., Статистическое описание промышленных объектов, изд-во „Энергия", М., 1971.
2. Дрейпер Н., Смит Г., Прикладной регрессионный анализ, изд-во „Статистика", М., 1973.
3. Хальд А., Математическая статистика с техническими приложениями, ИЛ, М., 1956.
4. Plackett R. L., Regression Analysis, Clarendon Press, Oxford, 1960.
5. Айвазян С. А., Статистическое исследование зависимостей, изд-во „Металлургия", М., 1968.
6. Рао С. Р., Линейные статистические методы и их применение, изд-во „Наука", М., 1968.

4.

ПРИМЕНЕНИЕ МЕТОДОВ ОРТОГОНАЛИЗАЦИИ ПРИ ПОСТРОЕНИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

4.1. Выбор вида модели

В этой главе будут рассмотрены методы построения регрессионных моделей вида

$$y = a' f(x). \tag{4.1}$$

с использованием ортогональных функций или процедур ортогонализации функций. Если вектор $f'(x) = (f_1(x), \ldots, f_k(x))$ функций переменных $x' = (x_1, \ldots, x_n)$ задан, а модель (4.1) линейна относительно коэффициентов $a' = (a_1, a_2, \ldots, a_k)$, то при наличии множества наблюдений (выборки) оценки коэффициентов a могут быть найдены с помощью метода множественной регрессии. Обычно компоненты вектора $f(x)$ выбираются по определенному критерию из достаточно широкого множества функций. Это множество должно включать в себя типы функций, принятых для представления модели. Например, модель может быть представлена с помощью нормальных полиномов вида $x_1^{q_1} x_2^{q_2} \ldots x_n^{q_n}$ ($q_l \geqslant 0$, q_l — целые числа). При этом выражение (4.1) является линейной комбинацией множества различных полиномов от переменных $x_1, x_2, \ldots, x_n$. Могут быть приняты функции и иного вида: $\exp(\alpha + \beta x_l)$, $\ln x_l$, $\sin \alpha x_l$, $\cos \alpha x_l$, $\operatorname{tg} \alpha x_l$, $\operatorname{arctg} \alpha x_l$, $1/(\alpha + \beta x_l)$ и т. д., причем величины α и β в перечисленных функциях предполагаются известными. Для функций, применяемых в данной конкретной задаче, при программной реализации метода построения модели необходимо в соответствующих записях указать номер функции, а также дополнительные данные, касающиеся показателей степени q_l, номера переменной i и численных значений коэффициентов α и β. Другая возможность программной реализации состоит в том, чтобы представить все множество допустимых функций в виде пакета программ.

В дальнейшем множество функций, которые могут быть использованы при построении модели, будет считаться заданным. При применении метода множественной регрессии необходимо из этого множества выбрать некоторое количество функций $f_l(x)$, которые следует включить в уравнение регрессии.

Один из возможных способов решения этой задачи заключается в следующем [1]. Пусть имеется выборка наблюдений, содержащая N экспериментальных точек. Допустим, что задано

некоторое множество функций $f_0(x)$, $f_1(x)$, ..., $f_k(x)$, причем число функций в этом множестве $k+1 \leqslant N$. Далее:

1. Находим оценки $\hat{a}_l$ коэффициентов уравнения регрессии вида (4.1) с помощью метода наименьших квадратов.

2. Для каждого коэффициента $\hat{a}_l$ рассчитываем величину

$$t_l = \frac{\hat{a}_l}{s_l}. \tag{4.2}$$

Здесь s_l — среднеквадратичная ошибка оценки $\hat{a}_l$, равная

$$s_l = \sqrt{c_{ll}}\, s, \tag{4.3}$$

где c_{ll} — диагональный элемент матрицы $C = (F'F)^{-1}$, а s^2 — оценка дисперсии ошибки наблюдений. Предположим, что оценка s^2 имеет число степеней свободы φ_e, а ошибки наблюдений распределены нормально с нулевым математическим ожиданием. При этом величина t_l подчинена закону распределения Стьюдента с числом степеней свободы φ_e.

3. Находим величину $t_j = \min\limits_l t_l$ и сопоставляем ее с критическим значением $t_{кр}$ распределения Стьюдента для числа степеней свободы φ_e при заданном уровне значимости α.

4. Если $t_j \geqslant t_{кр}$, полученную модель считаем окончательной.

5. Если $t_j < t_{кр}$, функция $f_j(x)$ из уравнения (4.1) исключается и все действия повторяются, начиная с п. 1, до тех пор, пока условие п. 4 не окажется выполненным.

Если ожидаемое число функций в окончательном уравнении существенно меньше числа функций в исходном множестве, более предпочтительным может оказаться иной метод выбора функций $f_l(x)$ [2]. Этот метод состоит в том, чтобы, начиная с некоторой функции $f_1(x)$, дополнять уравнение модели наиболее удачными функциями, выбираемыми из заданного множества. Пусть на некотором j-м цикле этой процедуры уравнение модели содержит k_j функций. Дальнейший алгоритм отбора включает следующие действия:

1. Методом наименьших квадратов вычисляем оценки $\hat{a}_l$ коэффициентов модели, содержащей k_j отобранных функций. Для каждого коэффициента $\hat{a}_l$ находим величину

$$t_l = \frac{\hat{a}_l}{s_l} = \frac{\hat{a}_l}{\sqrt{c_{ll}}\, s}. \tag{4.4}$$

Здесь s^2 — оценка дисперсии наблюдений с числом степеней свободы φ_e. Если в качестве оценки дисперсии наблюдений используется остаточная дисперсия s_k^2, то $\varphi_e = N - k_j$, где N — число точек измерений в выборке, а k_j — число функций в уравнении регрессии на j-м шаге.

2. Находим $t_j = \min_l t_l$ и сопоставляем величину t_j со значением $t_{кр}$. Вместо распределения Стьюдента здесь можно воспользоваться распределением Фишера с числами степеней свободы $\varphi_1 = 1$ и $\varphi_2 = \varphi_e$, которому подчинена величина $t_j^2 = F_j$.

3. Если $t_j < t_{кр}$ (или $F_j < F_{кр}$), функция $f_j(x)$ из уравнения регрессии исключается и процедура повторяется, начиная с п. 1.

4. Если $t_j \geqslant t_{кр}$ (или $F_j \geqslant F_{кр}$), среди функций исходного множества, не включенных в уравнение регрессии, ищем такую функцию $f_j(x)$, введение которой в модель приводит к максимальному уменьшению остаточной суммы квадратов. При этом для каждой из возможных функций $f_l(x)$ рассчитываем величину

$$V_l = \frac{\hat{a}_l^2}{c_{ll}} \tag{4.5}$$

и находим такую функцию $f_j(x)$, которой соответствует значение

$$V_j = \max_l V_l. \tag{4.6}$$

5. Проверяем целесообразность включения функции $f_j(x)$ в уравнение регрессии. Для этого вычисляем значение

$$t_j = \frac{\hat{a}_j}{\sqrt{c_{jj}}\, s} \quad \text{или} \quad F_j = \frac{\hat{a}_j^2}{c_{jj} s^2}. \tag{4.7}$$

Если $t_j \geqslant t_{кр}$, функция $f_j(x)$ включается в уравнение регрессии и процесс повторяется, начиная с п. 1 (при k_j, увеличенном на единицу).

6. Если $t_j < t_{кр}$, процедура заканчивается и полученное уравнение, включающее k_j функций, считается окончательным.

Заметим, что при реализации описанных выше процедур выбора вида уравнения регрессии приходится многократно вычислять элементы матрицы $C = (F'F)^{-1}$. Так как на каждом шаге исключается или добавляется в уравнение одна функция, то целесообразно воспользоваться рекуррентными формулами для коррекции элементов матрицы C (метод шаговой регрессии [2]).

Описанные процедуры существенно упрощаются, если функции $f_l(x)$ ортогональны на точках плана [1]. В этом случае

$$\sum_{l=1}^{N} f_l(x^l) f_j(x^l) = \begin{cases} 0 & \text{при } l \neq j, \\ \lambda_l & \text{при } l = j \end{cases} \tag{4.8}$$

и матрицы $F'F$ [и соответственно $(F'F)^{-1}$] являются диагональными. При этом устраняются недостатки процедуры, вызванные

[1] План эксперимента — это множество точек измерения зависимой переменной (выходной переменной, отклика). В данном разделе план эксперимента обозначается $\{x^\nu,\ \nu = \overline{1, N}\}$. — *Прим. ред.*

статистической связью между оценками $\hat{a}_l$. Если в условиях (4.8) новая функция $f_l(x)$ включается в уравнение регрессии, то оценки остальных коэффициентов $\hat{a}_l$ не изменяют своей величины.

Использование ортогональных систем функций связано с решением одной из следующих задач:

1. Для заданной системы функций $f_0(x)$, $f_1(x)$, ..., $f_k(x)$ построить такой план эксперимента, чтобы на точках плана заданная система функций была бы ортогональной. Это, по существу, задача ортогонального планирования эксперимента. Некоторые планы, получаемые при ее решении для линейных и квадратичных моделей, описываются в разд. 5 (планы полного и дробного факторного эксперимента, ортогональные планы второго порядка).

2. Для заданного плана эксперимента найти k линейно независимых функций n переменных, ортогональных на точках плана. Эта задача имеет в общем случае множество решений, поэтому необходимо наложить дополнительные условия на вид функций. В разд. 4.4 будет рассмотрен метод построения ортогональной системы функций вида

$$g_{q_1 q_2 \ldots q_n}(x) = \prod_{l=1}^{n} P_{q_l}(x_l). \tag{4.9}$$

Здесь каждая функция $P_{q_l}(x_l)$ есть полином от x_l степени q_l, причем $0 \leqslant q_l \leqslant Q_l$. План эксперимента задается в виде полного перебора всех возможных комбинаций переменных x_l на K_l уровнях.

3. Заданы план эксперимента и система функций, в общем случае не ортогональная на точках плана. В этом случае, исходя из заданной системы функций $f_0(x)$, $f_1(x)$, ..., $f_k(x)$, можно построить некоторый ортогональный базис $g_0(x)$; $g_1(x)$, ..., $g_k(x)$.

В следующем разделе рассматривается метод, основанный на сочетании приема ортогонализации и описанной выше последовательной процедуры отбора функций.

4.2. Построение модели в случае произвольных функций и произвольного плана

4.2.1. Постановка задачи

При нахождении оценок $\hat{a}$ коэффициентов модели (4.1) на основе метода наименьших квадратов

$$\hat{a} = (F'F)^{-1}F'\tilde{Y} \tag{4.10}$$

иногда пользуются специальными планами, позволяющими упростить вычисления.

В химической технологии для достижения необходимой точности воспроизведения планов эксперимента часто требуются слишком большие и экономически неоправданные затраты, а иногда план вообще реализовать не удается. Поэтому информация о действующей технологической установке часто появляется лишь в результате наблюдений за естественными изменениями в режимах процесса (пассивный эксперимент). В таких случаях оказывается невозможным априори выбрать систему функций, ортогональную на множестве точек наблюдений. Выходом из этой ситуации является применение методов ортогонализации.

Ортогонализация осуществляется последовательно (разд. 4.2.2), т. е. в ортогональный базис $g_l(x)$ поочередно включаются функции $f_l(x)$. После каждого шага можно найти коэффициенты модели, построенной на основе ортогональных функций $g_l(x)$ (разд. 4.2.3). Так как на каждом шаге остаточная сумма квадратов уменьшается на неотрицательную величину (разд. 4.2.4), оказывается возможным ввести простой принцип последовательного выбора, с помощью которого в модель включается „наилучшая" функция $f_l(x)$. В окончательном виде легко получить коэффициенты модели, представленной с помощью произвольных функций $f_l(x)$ (разд. 4.2.5).

Алгоритмически процедура оформляется таким образом, что ортогонализированные функции $g_l(x)$ в явном виде в ней не фигурируют. Применение вычислительной процедуры и принципа отбора функций, основанных на использовании приема ортогонализации, обеспечивают получение достаточно хорошего набора функций $f_l(x)$ из заданного множества функций (разд. 4.2.6 и 4.2.7).

4.2.2. Ортогонализация

Пусть известны результаты опытов в N точках x^ν, $\nu = 1\,(1)\,N$ [1], причем не обязательно требовать, чтобы в каждой точке был поставлен только один опыт. Обозначим через $\{x^\nu\}$ множество значений вектора x. Поясним понятие ортогональности двух функций $f_l(x)$ и $f_j(x)$, являющихся компонентами заданного вектора функций $f(x)$.

Определение 4.1

Две заданные функции $f_j(x)$ и $f_i(x)$ называются ортогональными на множестве точек $\{x^\nu,\ \nu = 1\,(1)\,N\}$, если

$$[f_i f_j] = \sum_{\nu=1}^{N} f_i(x^\nu)\, f_j(x^\nu) = 0. \qquad (4.11)$$

[1] $\nu = 1\,(1)\,N$ означает, что индекс ν меняется с шагом 1 от 1 до N.

В случае когда функции одинаковы, результат суммирования (4.11) является суммой квадратов QS.

Определение 4.2

Сумма квадратов QS некоторой известной функции $f_l(x)$ на множестве точек $\{x^\nu,\ \nu = 1\,(1)\,N\}$ есть

$$QS(f_l) = [f_l f_l] = [f_l^2]. \tag{4.12}$$

В частности, функции $f_l(x) = 1$ и $f_l(x) = 1$ не являются ортогональными. Для функции $f_l(x) = 1$ имеем $QS(1) = N$. При введении среднего значения для (4.11) множитель N^{-1} обычно опускают, так как его включение приводит лишь к усложнению вычислений. Отметим также, что при недостаточно хорошо обусловленном множестве точек наблюдений сумма квадратов QS может оказаться равной нулю, хотя функция $f_l(x)$ не равна тождественно нулю во всей области изменения переменной x. Например, для функции $f_1(x) = x'x\,(1 - x'x)$ при $x^1 = 0$, $x^2 = 1$ имеем $QS(f_1) = 0$. Подобные ситуации устраняются либо путем отказа от использования таких функций, либо путем добавления новых точек в выборку наблюдений. Аналогичным образом следует поступать и тогда, когда множество отобранных функций оказывается линейно зависимым.

Определение 4.3

Функции $f_l(x)$, $l = 1\,(1)\,k$, называются линейно зависимыми на множестве точек $\{x^\nu,\ \nu = 1\,(1)\,N\}$, если

$$\left[\left(\sum_{l=1}^{k} a_l f_l\right)^2\right] = 0, \tag{4.13}$$

причем по крайней мере один из коэффициентов

$$a_l \neq 0, \quad l = 1\,(1)\,k.$$

Это условие линейной зависимости слабее аналогичного условия в непрерывном случае. В качестве примера рассмотрим функции $f_1(x) = x'x$ и $f_2(x) = x'x^0$ на множестве точек $x^1 = 0$; $x^2 = x^0$. В силу того, что $f_1(x) - f_2(x) = x'(x - x^0) = 0$ для $x = x^1$ и $x = x^2$, имеет место $[(a_1 f_1 + a_2 f_2)^2] = 0$ для $a_1 = -a_2 = 1$ и функции являются линейно зависимыми.

Условие (4.13) охватывает также рассмотренные выше особые случаи линейной зависимости. Линейная зависимость обусловливает следующее свойство функций $f_l(x)$. Так как существует по крайней мере один коэффициент $a_l \neq 0$, можно считать, что $a_k = -1$. Обозначая величину суммы в выражении (4.13) через

$f(x)$, запишем

$$f_k(x) = f(x) + \sum_{i=1}^{k-1} a_i f_i(x) \text{ при } [f^2] = 0. \tag{4.14}$$

Функция типа $f(x)$ всегда существует, например:

$$f(x) = f_0(x) \cdot \prod_{\nu=1}^{N} x'(x - x^\nu),$$

где $f_0(x)$ означает некоторую произвольную функцию. Эта «остаточная функция» $f(x)$ ортогональна ко всем $f_i(x)$ на множестве точек измерений. В этом смысле условие $[f_k^2] = 0$ можно понимать как условие линейной зависимости между $f_k(x)$ и остаточной функцией $f(x)$.

После этих замечаний можно перейти к рассмотрению метода ортогонализации. Пусть задано k функций $f_i(x)$, линейно независимых на множестве точек $\{x^\nu,\ \nu = 1(1)N\}$. Построим k функций $g_i(x)$ по следующей рекуррентной формуле:

$$g_j(x) = f_j(x) - \sum_{i=1}^{j-1} g_i(x)\, a_{ij}, \quad j = 1(1)k. \tag{4.15}$$

Коэффициенты a_{ij} выбираются таким образом, чтобы функция $g_j(x)$ была ортогональной согласно определению 4.1. В качестве $g_1(x)$ примем

$$g_1(x) = f_1(x). \tag{4.16}$$

Пусть ортогональные функции $g_1(x)$, $g_2(x)$, ..., $g_{j-1}(x)$ уже построены и, следовательно, коэффициенты a_{il} для $1 \leqslant i < l \leqslant j-1$ известны. Имея в виду выражение (4.15), запишем уравнение для вычисления коэффициентов a_{ij}, $l = 1(1)j-1$:

$$[g_j g_l] = \left[\left(f_j - \sum_{i=1}^{j-1} g_i a_{ij}\right) g_l\right] = [f_j g_l] - \sum_{i=1}^{j-1} [g_i g_l]\, a_{ij},$$
$$[g_j g_l] = [g_l f_j] - a_{lj}[g_l^2] = 0 \text{ для } l < j.$$

Обозначая сумму квадратов значений ортогональной функции g_l через $a_{ll} = \mathrm{QS}(g_l) = [g_l^2]$, получаем выражение

$$a_{lj} = a_{ll}^{-1}[g_l f_j],\ a_{ll} = [g_l^2], \quad i = 1(1)j-1, \tag{4.17}$$

с помощью которого находятся коэффициенты a_{ij} для $i < j$. Сумму квадратов $\mathrm{QS}(q_j)$, равную $a_{jj} = [g_j^2]$, рассчитываем, используя (4.15):

$$a_{jj} = \left[\left(f_j - \sum_{i=1}^{j-1} g_i a_{ij}\right)^2\right],$$
$$a_{jj} = [f_j^2] - 2\sum_{i=1}^{j-1} [f_j g_i]\, a_{ij} + \sum_{i=1}^{j-1}\sum_{l=1}^{j-1} [g_i g_l]\, a_{ij} a_{lj}.$$

В силу ортогональности функций $g_i(x)$ при $i=1(1)j-1$ и учитывая (4.17), получаем

$$a_{jj}=[f_j^2]-2\sum_{i=1}^{j-1}a_{ij}a_{ij}^2+\sum_{i=1}^{j-1}a_{ij}a_{ij}^2, \tag{4.18}$$

$$a_{jj}=[f_j^2]-\sum_{i=1}^{j-1}a_{ij}a_{ij}^2, \quad j=1(1)k.$$

Выражения (4.15), (4.17) и (4.18) позволяют последовательно вычислять коэффициенты a_{ij}, $1\leqslant i\leqslant j\leqslant k$. Если в результате ортогонализации нужно получить лишь значения коэффициентов a_{ij}, то вычисления проводят без введения функций $g_i(x)$. При этом в (4.17) можно, используя (4.15), записать величину $[g_if_j]$ в виде

$$[g_if_j]=\left[\left(f_i-\sum_{l=1}^{i-1}g_la_{li}\right)f_j\right]=[f_if_j]-\sum_{l=1}^{i-1}[g_lf_j]a_{li}.$$

Отсюда следует, что [1]

$$a_{ij}\Leftarrow a_{ii}^{-1}\left([f_if_j]-\sum_{l=1}^{i-1}a_{li}a_{lj}\right), \quad 1\leqslant i<j\leqslant k. \tag{4.19}$$

Уравнения (4.18) и (4.19) определяют последовательный процесс вычисления коэффициентов a_{ij}, множество значений которых может быть представлено с помощью верхней треугольной матрицы A:

$$A=\begin{bmatrix} a_{11}\cdots a_{1j} & a_{1j+1} & \cdots \\ \vdots & \vdots & \\ a_{ii}\cdots a_{ij} & a_{ij+1} & \cdots \\ & \ddots & \\ a_{i+1i+1}\cdots a_{i+1j} & a_{i+1j+1} & \cdots \\ \ddots & \vdots & \\ a_{jj} & & \end{bmatrix}$$

Заполнение матрицы A осуществляется столбцами слева направо. Если $(j-1)$-й столбец заполнен, элементы j-го столбца рассчитываются с помощью уравнения (4.19), начиная с верхнего, причем элемент, находящийся на главной диагонали, определяется по формуле (4.18). Ввиду того что с точностью до множителя a_{ii}^{-1} уравнения (4.18) и (4.19) имеют одинаковую струк-

[1] Символ $\Leftarrow$ означает „присвоить значение".

туру, вычисления можно проводить в такой последовательности, когда перед началом счета элементы главной диагонали устанавливаются равными $a_{ii}=1$, а затем применяется выражение (4.19) для $1\leqslant i\leqslant j\leqslant k$. В результате элементам главной диагонали будут присвоены их действительные значения.

Выражение (4.19) содержит деление на сумму квадратов $a_{ii}=[g_i^2]$. Для осуществления этой операции требуется выявить случаи, когда $a_{ii}=0$. Заметим, что всегда $a_{ii}\geqslant 0$, так как a_{ii} представляет собой сумму квадратов. Ниже будет показано, что $a_{ii}=0$ тогда и только тогда, когда $f_i(x)$ линейно связана с $f_j(x)$, $j=1\,(1)\,i-1$, в соответствии с определением 4.3. При проведении ортогонализации такие функции $f_i(x)$ должны быть исключены из рассмотрения.

4.2.3. Оценка параметров модели, представленной с помощью ортогональных функций

Модель, представленная с помощью ортогональных функций, полученных в предыдущем разделе, имеет вид

$$y(x)=\sum_{i=1}^{k} b_i g_i(x). \tag{4.20}$$

Дополним множество точек измерения в пространстве независимых переменных упорядоченным множеством результатов наблюдений $\tilde{y}$ до выборки наблюдений $\{x^\nu,\ \tilde{y}^\nu,\ \nu=1\,(1)\,N\}$. Применяя метод наименьших квадратов при величинах a_{ii}, рассчитанных по формуле (4.17), получим оценки $\hat{b}_i$:

$$\hat{b}_i=a_{ii}^{-1}[\tilde{y}g_i],\quad i=1\,(1)\,k. \tag{4.21}$$

Чтобы не пользоваться ортогональными функциями $g_i(x)$, запишем $[\tilde{y}g_i]$, применяя формулу (4.15), следующим образом:

$$[\tilde{y}g_i]=\left[\tilde{y}\left(f_i-\sum_{j=1}^{i-1} g_j a_{ji}\right)\right]=[yf_i]-\sum_{j=1}^{i-1}[\tilde{y}g_j]\,a_{ji}.$$

Учитывая (4.21), получаем

$$\hat{b}_i=a_{ii}^{-1}\left([\tilde{y}f_i]-\sum_{j=1}^{i-1}\hat{b}_j a_{jj}a_{ji}\right),\quad i=1\,(1)\,k. \tag{4.22}$$

С помощью выражения (4.22) оценки $\hat{b}_i$ можно рассчитывать последовательно.

4.2.4. Принцип отбора функций

За меру оценки качества модели (4.20) примем остаточную сумму квадратов S_R. Величина S_R находится по формуле

$$S_R = [\tilde{y}] - 2\sum_{i=1}^{k} \hat{b}_i [\tilde{y} g_i] + \sum_{i=1}^{k}\sum_{j=1}^{k} \hat{b}_i \hat{b}_j [g_i g_j].$$

Имея в виду выражение (4.21) и условие $[g_i g_j] = 0$ при $i \neq j$, получаем

$$S_R = [\tilde{y}^2] - \sum_{i=1}^{k} a_{ii} (\hat{b}_i)^2. \tag{4.23}$$

Существенным является то, что выражение (4.23) строится как сумма слагаемых вида $V_i = a_{ii} (\hat{b}_i)^2$.

Приписывая индексе величине S, получим

$$S_i = S_{i-1} - V_i, \quad V_i = a_{ii} (\hat{b}_i)^2, \quad i = 1\,(1)\,k, \tag{4.24}$$
$$S_0 = [y^2], \quad S_R = S_k.$$

Так как уравнения (4.19) и (4.22) имеют рекуррентную форму, то, в случае когда определены $(i-1)$ коэффициентов $\hat{b}_j$, $j = 1\,(1)\,i-1$, вычисление i-го коэффициента не оказывает влияния на величины ранее рассчитанных коэффициентов. Величина V_i характеризует собой „приращение точности", которое зависит только от функции $f_i(x)$, использованной при построении ортогонального базиса на i-м шаге [а также, разумеется, от ранее использованных для построения этого базиса функций $f_j(x)$, $j = 1\,(1)\,i-1$]. Указанные обстоятельства могут быть использованы для обоснования принципа выбора, с помощью которого из множества функций $\{f_i(x),\ i = 1\,(1)\,K\}$ следует отбирать некоторое подмножество $k < K$ функций для включения в уравнение модели.

Определение 4.4. Принцип отбора

Пусть в уравнение модели уже включено $j-1 < k$ функций из множества $\{f_i(x),\ i = 1\,(1)\,K,\ K > k\}$. Тогда из оставшихся $K-j+1$ функций на текущем шаге выбирается такая функция $f_j(x)$, которой соответствует максимальная величина $V_j = a_{jj} (\hat{b}_j)^2$.

Выбор функций производится путем последовательного вычисления соответствующих коэффициентов. Вычислительные затраты можно измерять числом рассчитанных столбцов треугольной матрицы A (независимо от их длины), которое равно

$$\sum_{j=1}^{k} (K-j+1) = k(K+1) - \binom{k+1}{2} = \frac{k}{2}(2K+1-k).$$

В случае если для нескольких функций $f_j(x)$ на некотором шаге величины $a_{jj}(b_j)^2$ одновременно достигают максимума, выбирается любая из этих функций.

Процедуру добавления функций в уравнение модели целесообразно прекращать тогда, когда величина V_j становится меньше некоторой малой положительной величины ε, либо когда вклад V_j в остаточную сумму квадратов оказывается статистически незначимым. Если известна оценка дисперсии ошибки наблюдений s^2, то на каждом j-м шаге можно рассчитать величину $F_j = V_j/s^2$. Эта величина в предположении нормальности распределения ошибок наблюдений подчинена F-распределению с числами степеней свободы $\varphi_1 = 1$ и $\varphi_2 = \varphi_e$. Пользуясь табл. 15.4—15.5, при заданном уровне значимости α находим значение $F_{кр}$. Выбор функций прекращают, если $F_j < F_{кр}$. Для вычисления оценки s^2 необходимо, чтобы по крайней мере в одной экспериментальной точке было проведено несколько опытов. Если дублирование наблюдений отсутствует, то в качестве оценки дисперсии ошибки наблюдений s^2 используется остаточная дисперсия $s_j^2 = S_j/\varphi_j$ с числом степеней свободы $\varphi_j = N - j$.

Наряду с описанным возможны и другие способы отбора функций. Например, из K заданных функций можно извлекать одновременно по k произвольных функций. При этом число различных возможностей выбора составляет $\binom{K}{k}$, а число рассчитываемых столбцов матрицы A — $k\binom{K}{k}$ [1]. В этой процедуре уже для небольших значений k и K вычислительные затраты оказываются неоправданно большими. Так, для выбора $k = 10$ функций из множества $K = 20$ функций требуется рассчитать около $2 \cdot 10^6$ столбцов матрицы A. Применение принципа отбора, основанного на определении 4.4, потребует в этом случае вычисления 155 столбцов. Без реализации же какого-либо выбора необходимо вычислить всего 10 столбцов. Таким образом, принцип отбора, сформулированный в определении 4.4, представляет собой приемлемый компромисс по отношению к вычислительным затратам.

4.2.5. Оценка коэффициентов модели

Ортогональные функции $g_i(x)$, полученные в разд. 4.2.2 для описания модели, мало пригодны для практического применения, так как, согласно (4.15), они представляют собой линейную

[1] Здесь $\binom{K}{k}$ означает число сочетаний из K по k. Ниже эта величина обозначается также через C_K^k. — *Прим. ред.*

комбинацию исходных функций $f_j(x)$, $j = 1(1)i$, и коэффициентов, зависящих от точек проведения эксперимента. При этом для разных множеств точек измерений ортогональные функции будут получаться различными. Кроме того, для их полного описания необходимо задание треугольной матрицы A с $\binom{k-1}{2}$ элементами. Желательно представлять модель в виде линейной комбинации исходных функций $f_i(x)$:

$$y = \sum_{i=1}^{k} a_i f_i(x), \tag{4.25}$$

так чтобы для ее описания требовалось знание только k коэффициентов a_i, $i = 1(1)k$.

Возврат к исходной системе функций может быть осуществлен с помощью следующих вычислений. Согласно уравнению (4.15), имеем

$$f_j(x) = \sum_{i=1}^{j} g_i(x)\, a_{ij}, \quad j = 1(1)k, \tag{4.26}$$

причем $a_{ij} = 0$ для $i > j$ и $a_{jj} = 1$. При $i < j$ в треугольной матрице A элементы a_{ij} отсутствуют. Если положить

$$g_i(x) = \sum_{l=1}^{i} b_{li} f_l(x), \quad i = 1(1)k, \tag{4.27}$$

где $b_{li} = 0$ для $l > i$, то, подставив (4.27) в (4.26), получаем тождество

$$f_j(x) \equiv \sum_{i=1}^{j} a_{ij} \sum_{l=1}^{i} b_{li} f_l(x) = \sum_{l=1}^{j} f_l(x) \sum_{i=l}^{j} a_{ij} b_{li}.$$

Сравнение значений коэффициентов позволяет найти

$$\sum_{i=l}^{j} a_{ij} b_{li} = \begin{cases} 1 & \text{для } l = j, \\ 0 & \text{для } l < j. \end{cases} \tag{4.28}$$

Поскольку A, так же как и $B = (b_{ij})$, является треугольной матрицей, то система (4.28) легко разрешается. Для $l = j$ из (4.28) следует

$$a_{jj} b_{jj} = 1 \to b_{jj} = 1, \quad j = 1(1)k,$$

а для $l < j$

$$\begin{aligned} a_{lj} b_{ll} + a_{jj} b_{jl} + \sum_{l=l+1}^{j-1} a_{lj} b_{ll} &= 0, \\ b_{lj} &= -\left(a_{jl} + \sum_{l=j+1}^{i-1} a_{ll} b_{lj} \right), \quad j < i. \end{aligned} \tag{4.29}$$

Для оценок $\hat{a}_i$ модели вида (4.25) аналогичным путем находим

$$\sum_{i=1}^{k} \hat{a}_i f_i(x) \equiv \sum_{i=1}^{k} \hat{b}_i g_i(x) \equiv \sum_{i=1}^{k} \hat{b}_i \sum_{l=1}^{i} b_{il} f_l(x),$$

$$\sum_{i=1}^{k} \hat{a}_i f_i(x) \equiv \sum_{l=1}^{k} f_l(x) \sum_{i=l}^{k} \hat{b}_i b_{il},$$

$$\hat{a}_l = \sum_{i=l}^{k} \hat{b}_i b_{il}, \tag{4.30}$$

$$\hat{a}_l = \hat{b}_l + \sum_{j=l+1}^{k} \hat{b}_j b_{jl}, \quad l = 1\,(1)\,k.$$

При реализации на цифровой вычислительной машине целесообразно записывать коэффициенты b_{ij} в свободной нижней части матрицы A, расположенной слева от главной диагонали:

$$a_{ij} = -\left(a_{ji} + \sum_{l=j+1}^{i-1} a_{jl} a_{li}\right), \quad 1 \leqslant j < i \leqslant k,$$

$$\hat{a}_l = \hat{b}_l + \sum_{j=l+1}^{k} \hat{b}_j a_{jl}, \quad l = 1\,(1)\,k. \tag{4.31}$$

Уравнения (4.31) представляют собой систему для последовательного вычисления оценок $\hat{a}_l$ модели вида (4.25). Так как для вычисления $\hat{a}_l$ требуется знание элементов a_{jl}, то вычисление элементов a_{ij}, $i > j$, осуществляется лишь после заполнения верхней (по отношению к главной диагонали) части треугольной матрицы, причем a_{ij}, $i > j$, вычисляются по столбцам сверху вниз, начиная с $(k-1)$-го столбца. В случае если представляет интерес знание ортогонализированных функций, согласно (4.27), находим

$$g_l(x) = f_l(x) + \sum_{i=1}^{l-1} a_{il} f_i(x). \tag{4.32}$$

Уравнения (4.31) позволяют дать ответ на вопрос о необходимых и достаточных условиях линейной зависимости функций $f_i(x)$, поставленный в разд. 4.2.2. Основываясь на рекуррентном характере процедуры отбора функций, можно доказать следующее утверждение:

Утверждение 4.1. Линейная зависимость

Если $(j-1)$ функций $f_i(x)$, $i = 1\,(1)\,j-1$, линейно независимы, то функция $f_j(x)$ линейно зависима от $f_i(x)$, $i = 1\,(1)\,j-1$, тогда и только тогда, когда $a_{jj} = [g_j^*]$ равно нулю.

Для доказательства предположим, что $a_{ii} > 0$ при $i = 1\,(1)\,j-1$. Пусть $a_{jj} = 0$; тогда из (4.15) следует

$$0 = a_{jj} = [g_j^2] = \left[\left(f_j - \sum_{i=1}^{j-1} g_i(x)\, a_{ij}\right)^2\right].$$

Так как $a_{ii} > 0$ при $i = 1\,(1)\,j-1$, то можно однозначно определить коэффициенты в следующем выражении, записанном аналогично (4.14):

$$f_j(x) = f(x) + \sum_{i=1}^{j-1} g_i(x)\, a_{ij}, \qquad [f^2] = 0. \tag{4.33}$$

С помощью уравнения (4.27) функции $g_i(x)$ можно представить в виде линейных комбинаций функций $f_i(x)$, где коэффициенты b_{ij} однозначно определяются выражением (4.29). При этом функция $f_j(x)$ оказывается линейной комбинацией функций $f_i(x)$, $i = 1\,(1)\,j-1$, и остаточной функции $f(x)$, которая в силу условия утверждения 4.1 равна нулю во всех точках измерений:

$$f_j(x) = f(x) + \sum_{i=1}^{j-1} \lambda_i f_i(x), \qquad [f^2] = 0. \tag{4.34}$$

Таким образом, достаточность доказана. Если, наоборот, исходить из представления (4.34) и подвергнуть функции $f_i(x)$ ортогональному преобразованию (4.15), то можно получить выражение (4.33), коэффициенты которого $a_{ij}(\lambda_1, \lambda_2, \ldots, \lambda_{j-1})$ однозначно определены. При этом j-я ортогональная функция $g_j(x)$ выражается линейной комбинацией предшествующих, включая остаточную функцию

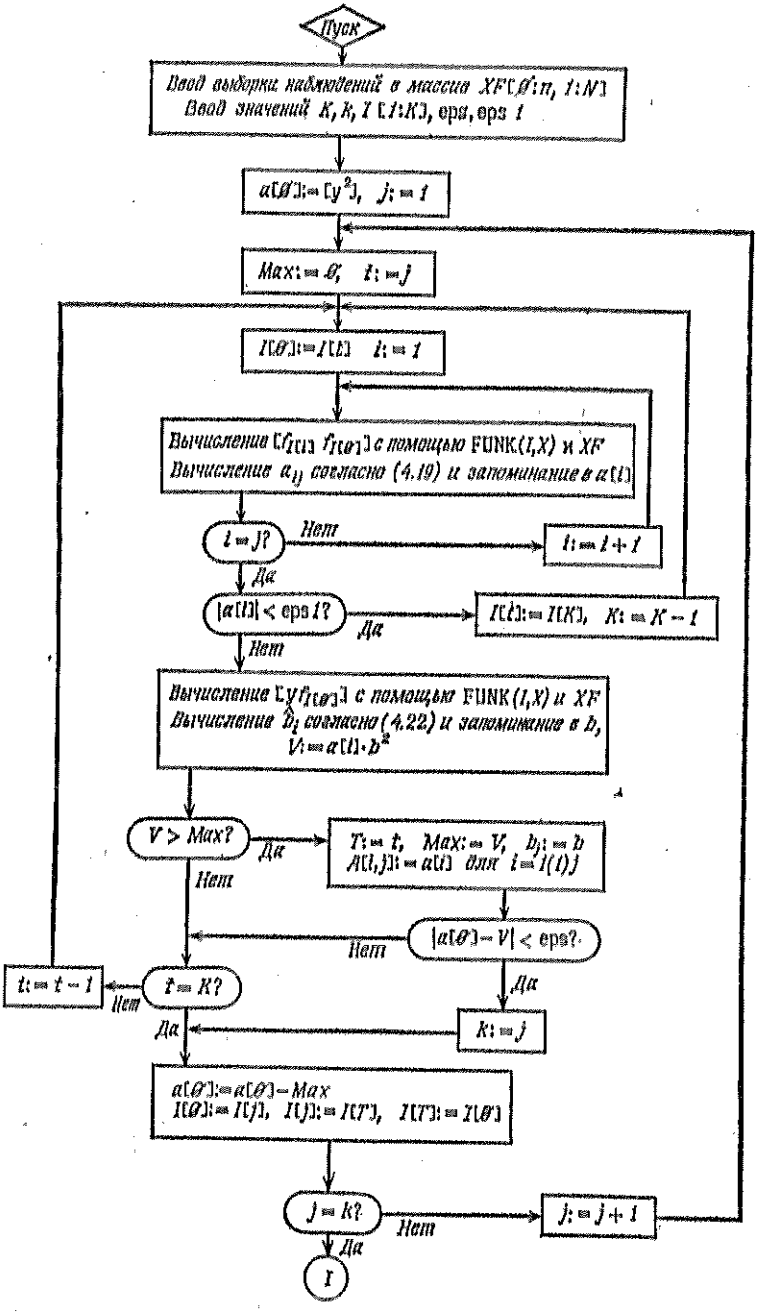
$$g_j(x) = f(x) + \sum_{i=1}^{j-1} \mu_i g_i(x).$$

В соответствии с требованием ортогональности и условием утверждения множитель при $g_j(x)$ равен

$$a_{jj} = [g_j^2] = [f g_j] + \sum_{i=1}^{j-1} \mu_i [g_i g_j] = 0$$

и оказывается независимым от μ_j. Таким образом, равенство $a_{jj} = 0$ является необходимым условием линейной зависимости.

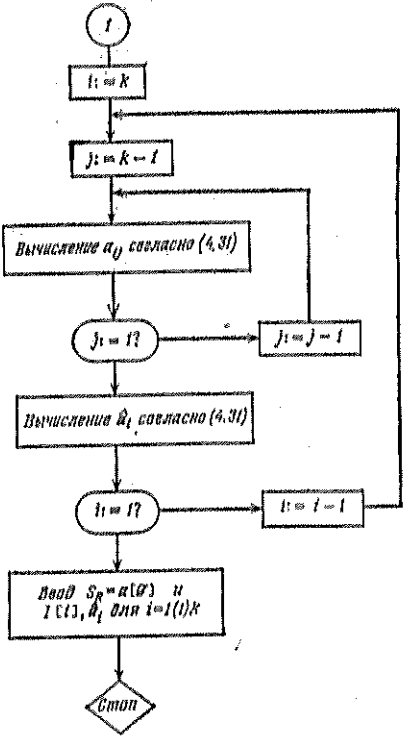
Из приведенного доказательства следует, что если функции $f_i(x)$ линейно зависимы, то линейно зависимы и функции $g_i(x)$, и наоборот.



Фиг. 4.1. Блок-схема программы ORFU.

4.2.6. Указания к программной реализации метода

Реализация принципа отбора (разд. 4.2.4) требует учета ряда соображений, касающихся вычислительных аспектов. Различные виды функций $f_l(x)$ должны быть заранее введены в подпрограмму FUNK(l, X), где l означает текущий номер функции, а X—вектор независимых переменных. При обращении к программе ORFU построения модели необходимо установить множество K функций, приемлемых для построения данной модели, которые вводятся в виде списка индексов $I[1:K]$ функций подпрограммы FUNK. Программа ORFU с учетом этого списка индексов и с помощью принципа отбора находит k функций $f_l(x)$, которые должны включаться в уравнение модели. Так как расчет всех коэффициентов производится последовательно, то после каждого шага индексный список упорядочивается таким образом, чтобы уже отобранные функции попадали в его начало. Выборка наблюдений $\{\tilde{y}^\nu, x^\nu, \nu = 1(1)N\}$ в программе располагается в массиве $XF[0:n, 1:N]$, где n—размерность вектора независимых переменных.



Фиг. 4.1 (продолжение).

На фиг. 4.1 представлена блок-схема программы ORFU определения оценок коэффициентов $\hat{a}_l$ модели (4.25) на основе описанного принципа отбора функций. Требуемый объем памяти для реализации программы определяется числом $(n+1)N$ элементов выборки и числом k^2 элементов матрицы коэффициентов A. Если K не слишком велико, то с целью экономии времени вычислений можно в одном рабочем цикле рассчитать величины $[f_l f_j]$, записать их на место матрицы $A[1:K, 1:K]$, а затем последовательно, по мере вычислений, заменять записи в матрице A на значения коэффициентов a_{lj}.

4.2.7. Пример

Для демонстрации метода рассмотрим пример, в котором из $K=4$ функций двух переменных требуется отобрать $k=3$

функций для построения модели по заданному множеству наблюдений, содержащему $N=4$ опыта.

Исходные данные для примера:

$$f_1(x_1, x_2) = 1,$$
$$f_2(x_1, x_2) = \frac{1}{x_1 + 1},$$
$$f_3(x_1, x_2) = \frac{1}{x_2 + 2},$$
$$f_4(x_1, x_2) = \frac{1}{x_1 + x_2 + 1},$$

$\tilde{y}$	x_1	x_2
0	0	0
1	0	1
1	1	0
1	1	1

(4.35)

$$s = 10^{-2}, \quad \varphi_e = 4.$$

Прежде всего определим первую функцию, которую в соответствии с принципом отбора следует подвергнуть ортогонализации. Согласно выражению (4.24), имеем $S_0 = 3{,}000$. Пользуясь уравнениями (4.18), (4.19), (4.22) и (4.27), получаем для $f = f_1, f_2, f_3, f_4$ [1]

f	a_{11}	$[\tilde{y}f]$	$\hat{b}_1$	V_1	
f_1	4,000	3,000	0,750	2,250*	
f_2	2,500	2,000	0,800	1,600	$S_1 = 0{,}750$
f_3	0,722	1,167	1,615	1,885	
f_4	1,611	1,333	0,816	1,105	

Величина V_1 достигает максимального значения при выборе функции $f_1(x_1, x_2)$. Согласно табл. 15.5, при уровне значимости $\alpha = 0{,}01$ и числах степеней свободы $\varphi_1 = 1$, $\varphi_2 = \varphi_e = 4$ критическое значение F-распределения $F_{кр} = 21{,}2$. Так как $F_1 = V_1/s^2 = 2{,}25 \cdot 10^4 > 21{,}2 = F_{кр}$, то функция $f_1(x_1, x_2)$ принимается для построения ортогонального базиса [2]. Для остаточной суммы квадратов, согласно (4.24), получаем $S_1 = 0{,}750$.

На втором шаге, применяя принцип отбора, находим следующую функцию из множества $f = f_2, f_3, f_4$:

f	$[f_1 f]$	a_{12}	a_{22}	$\hat{b}_2$	V_2	
f_2	3,000	0,750	0,250	—1,000	0,250	
f_3	1,667	0,417	0,028	—3,000	0,250	$S_2 = 0{,}056$
f_4	2,333	0,583	0,250	—1,667	0,694*	

[1] Звездочкой здесь отмечена строка, соответствующая функции с максимальным значением V_1. — *Прим. ред.*

[2] Использование F-распределения в этом разделе предполагает, что ошибка наблюдений распределена по нормальному закону. — *Прим. ред.*

Так как $F_2 = V_2/s^2 = 0{,}694 \cdot 10^4 > 21{,}2 = F_{кр}$, то в качестве второй функции для построения ортогонального базиса принимается функция $f_4(x_1, x_2)$.

Наконец, на третьем шаге, пользуясь принципом отбора, выбираем функцию из множества $f = f_2, f_3$:

f	$[f_1 f]$	$[f_4 f]$	a_{13}	a_{23}	a_{33}	$\hat{b}_3$	V_3	
f_2	3,000	1,915	0,750	0,667	0,139	0,200	5,56	10^{-3} *
f_3	1,667	1,029	0,417	0,222	0,015	0,600	5,56	10^{-3} *

Отсюда видно, что максимум V_3 определяется не однозначно, т. е. для ортогонального базиса можно принять либо функцию f_2, либо функцию f_3. Возможность принятия этих функций не отвергается, так как $F_3 = V_3/s^2 = 55{,}6 > 21{,}2 = F_{кр}$.

С помощью уравнения (4.31) вычисляются элементы матрицы A и оценки $\hat{a}_i$. В случае, если на третьем шаге выбирается функция f_2, получаем

$$f = \begin{bmatrix} f_1 \\ f_4 \\ f_2 \end{bmatrix}, \quad A = \begin{bmatrix} 4{,}000 & 0{,}583 & 0{,}750 \\ -0{,}583 & 0{,}250 & 0{,}667 \\ -0{,}361 & 0{,}667 & 0{,}139 \end{bmatrix}, \quad \hat{b} = \begin{bmatrix} 0{,}750 \\ -1{,}667 \\ 0{,}200 \end{bmatrix}, \quad \hat{a} = \begin{bmatrix} 1{,}650 \\ -1{,}800 \\ 0{,}200 \end{bmatrix}.$$

В случае, если на третьем шаге выбрана функция f_3, имеем

$$f = \begin{bmatrix} f_1 \\ f_4 \\ f_3 \end{bmatrix}, \quad A = \begin{bmatrix} 4{,}000 & 0{,}583 & 0{,}417 \\ -0{,}583 & 0{,}250 & 0{,}222 \\ -0{,}287 & -0{,}222 & 0{,}015 \end{bmatrix}, \quad \hat{b} = \begin{bmatrix} 0{,}750 \\ -1{,}667 \\ 0{,}600 \end{bmatrix}, \quad \hat{a} = \begin{bmatrix} 1{,}550 \\ -1{,}800 \\ 0{,}600 \end{bmatrix}.$$

Математическая запись модели в виде уравнения (4.25) $y = a'f$ выглядит так:

$$\hat{y} = 1{,}65 - \frac{1{,}8}{x_1 + x_2 + 1} + \frac{0{,}2}{x_1 + 1},$$

или (при выборе функции f_3)

$$\hat{y} = 1{,}55 - \frac{1{,}8}{x_1 + x_2 + 1} + \frac{0{,}6}{x_2 + 2}.$$

При этом для выборки (4.35) $S_R = 0{,}05$. При вычислении оценок $\hat{a}$ ортогональные функции $g_i(x)$ в явном виде не фигурируют. Однако при известной матрице A функции $g_i(x)$ легко определить с помощью уравнения (4.33). Так, для случая, когда

на третьем шаге выбрана функция f_2, имеем

$$g_1(x_1, x_2) = 1,$$

$$g_4(x_1, x_2) = \frac{1}{x_1 + x_2 + 1} - 0{,}583,$$

$$g_2(x_1, x_2) = \frac{1}{x_1 + 1} - 0{,}361 - \frac{0{,}667}{x_1 + x_2 + 1}.$$

Случай выбора функций f_3 отличается от рассмотренного выше только третьей ортогональной функцией

$$g_3(x_1, x_2) = \frac{1}{x_2 + 2} - 0{,}287 - \frac{0{,}222}{x_1 + x_2 + 1}.$$

Модель, построенная на основе ортогональных функций $g_i(x)$ в виде уравнения (4.20) $y = \hat{b}'g$, выглядит так:

$$\hat{y} = 0{,}75g_1 - 1{,}667g_4 + 0{,}2g_2,$$

или, для случая выбора f_3,

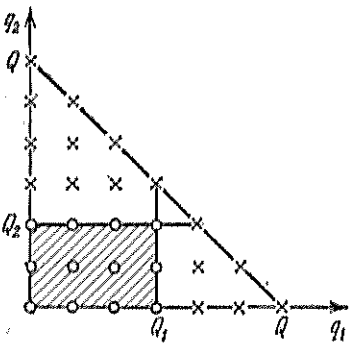
$$\hat{y} = 0{,}75g_1 - 1{,}667g_4 + 0{,}6g_3.$$

Рассмотренный пример показывает, что при не слишком больших K и N можно проводить расчеты без использования средств вычислительной техники.

4.3. Построение модели в случае произвольного плана эксперимента

4.3.1. Постановка задачи

В этом разделе метод, описанный в разд. 4.2, применяется для случая, когда в качестве функций для построения модели выбраны полиномы определенного вида. С одной стороны, это ограничивает область применения метода, с другой же стороны, это дает ряд преимуществ с вычислительной точки зрения. При произвольных функциях довольно сложно указать приемлемый для всех случаев порядок их расположения в подпрограмме FUNK (I, X). Поэтому, как правило, для всякого конкретного случая программа FUNK пишется заново. При использовании полиномов подпрограмму, содержащую набор функций, можно вообще не включить. Для выбора требуемых полиномов необходимо только наложить ограничения на степень отдельных переменных и на суммарную степень переменных в полиноме (см. разд. 4.3.2). Для проведения вычислений с учетом принципа отбора функций применима описанная в разд. 4.2.6 программа ORFU, которая должна быть изменена лишь в части, относящейся к вычислению полиномов (см. разд. 4.3.3).



Фиг. 4.2. Определение числа узлов решетки.

4.3.2. Выбор полиномов

При определении типа полиномов мы будем исходить из требования по возможности малого числа параметров для задания полиномов. Поэтому для метода ортогонализации имеет смысл применить нормальные полиномы, которые однозначно определяются заданием показателей степеней переменных:

$$f_l(x) = f_{q_1 q_2 \ldots q_n}(x) = \prod_{j=1}^{n} (x_j)^{q_j}. \tag{4.36}$$

Показатели степени q_j, $j = 1\,(1)\,n$, являются неотрицательными целыми числами, ограниченными величинами Q_j:

$$0 \leqslant q_j \leqslant Q_j, \qquad j = 1\,(1)\,n. \tag{4.37}$$

Если величины Q_j заданы, то число K допустимых полиномов, из которых осуществляется отбор при построении модели, составит

$$K = \prod_{j=1}^{n} (Q_j + 1). \tag{4.38}$$

Это число может оказаться весьма большим. Так, при $Q_j = 1$ и $n = 20$ получаем $K = 2^{20} \approx 10^6$. Поэтому путем введения дальнейших ограничений число K необходимо привести в соответствие с возможностями вычислительных средств. В качестве приемлемого условия напрашивается следующее ограничение на суммарную степень допустимых полиномов:

$$\sum_{j=1}^{n} q_j \leqslant Q. \tag{4.39}$$

Обозначим число полиномов вида (4.36), удовлетворяющих условиям (4.39) и (4.37), через $Z_n(Q)$, т. е. в виде некоторой

функции от n, Q и Q_j. Значение этой функции при $n=2$ определяется числом узлов решетки по формуле (фиг. 4.2)

$$Z_2(Q)=\binom{Q+2}{2}-\binom{Q-Q_1+1}{2}-\binom{Q-Q_2+1}{2}+\binom{Q-Q_1-Q_2}{2}.$$

В приложениях удобно задавать величину K, а Q определять так, чтобы выполнялось условие

$$Z_n(Q)\leqslant K\leqslant Z_n(Q+1). \tag{4.40}$$

Функция $Z_n(Q)$ удовлетворяет следующему рекуррентному уравнению:

$$Z_i(j)=Z_i(j-1)+Z_{i-1}(j)-Z_{i-1}(j-1-Q_i),\quad i=1(1)n,\quad j=1(1)q, \tag{4.41}$$

$$Z_i(j)=\begin{cases}1 & \text{для } i=1(1)n, \quad j=0,\\ 1 & \text{для } i=0, \quad j=0(1)q,\\ 0 & \text{для } i=1(1)n, \quad j<0.\end{cases} \tag{4.42}$$

Обозначим $q=\sum_{j=1}^{n}Q_j$. Тогда для $Q\geqslant q$ имеем $Z_n(Q)=\prod_{i=1}^{n}(Q_i+1)$ и, следовательно, изменение Q в этом случае не приводит к уменьшению числа полиномов. Расчеты по рекуррентному выражению (4.42) легко реализуются с помощью средств вычислительной техники. При заданных n, K и Q_i, $i=1(1)n$, ограничение Q определяется по следующей программе:

```
ZN[0]:= 1;
for j:= 1 step 1 until q do ZA[j]:= 1;
for i:= 1 step 1 until n do
begin Q:= QQ[i];
      for j:= 1 step 1 until q do
      begin Z:= j - 1 - Q;
            if Z < 0 then Z:= 0 else Z:= ZA[Z];
            ZN[j]:= ZN[j - 1] + ZA[j] - Z;
      end
      for j:= 1 step 1 until q do ZA[j]:= ZN[j];
end
for j:= 1 step 1 until q do
if ZN[j] > K then
begin Q:= j - 1;
      goto FIN;
end
FIN:
```

Множество допустимых полиномов для использования в модели определяются путем задания границ Q_j для показателей степени и максимального числа K полиномов в допустимом множестве. Затем с помощью представленной подпрограммы определяют границу Q (4.39).

Выбор для модели нормальных полиномов (4.36) не связан

с предположением, что при произвольно выбираемых полиномах $P_j(x_1, x_2, \ldots, x_n)$ со степенями переменных x_j, не превосходящими Q_j, $j=1\,(1)\,n$, и при выборе полинома P_j, степень которого не выше Q, должен получиться один и тот же результат.

4.3.3. Указания к программной реализации метода

Рассмотренную в разделе 4.2.6 программу можно изменить в части, касающейся подпрограммы вычисления функций. При реализации принципа отбора может быть использована следующая последовательная стратегия. Пусть уже отобрано $j-1<k$ нормальных полиномов для включения в ортогональный базис. Показатели степеней $(q_n, q_{n-1}, \ldots, q_2, q_1)$ этих функций содержатся в списке показателей, представляющем собой массив целых чисел размерности $n\times k$. Теперь следует выбрать j-й полином из множества допустимых для включения в ортогональный базис. При этом для полиномов, отсутствующих в списке, вычисляют коэффициенты a_{lj}, $\hat{b}_j$ и величины V_j. Среди допустимых полиномов находят полином, которому соответствует максимальная величина V_j. Показатели степеней этого полинома заносятся в список показателей степеней под j-м номером. Если $j=k$, вычисления прекращают.

Опишем алгоритм вычисления допустимых полиномов. Вектор показателей степеней $(q_n, \ldots, q_1)$ изменяется аналогично некоторому десятичному числу, причем перенос десятков заменяется Q_l-переносом и на каждом шаге вычисления проверяется условие $\sum_{l=1}^{n} q_l = q_0 \leqslant Q$. Если всем q_l, $l=1\,(1)\,n$, и q_0 присвоены определенные значения, то соответствующий этим значениям полином вычисляется по следующей программе:

```
A: u := 1;
   for i := 1 step 1 until n do
   begin q := qq[i];
         q0 := q0 − q;
         if u = 1 then
         begin if q = Q[i] then q := 0 else
               begin q := q + 1;
                     u := 0;
               end
         end
         qq[i] := q;
         q0 := q0 + q;
         if u = 0 then goto E;
   end
E: if q0 > Q then goto A;
```

Исходными данными для работы программы ORFU1 являются число переменных n, число точек выборки N, выборка наблюдений $\{x_1^\nu, \ldots, x_n^\nu, \tilde{y}^\nu, \nu = 1(1)N\}$, граничные значения Q_i, $i = 1(1)n$, для степеней переменных, входящих в полиномы, максимальное число K допустимых полиномов и число k, равное числу полиномов, отбираемых для включения в математическую модель.

Программа осуществляет автоматическую проверку данных; в частности, проверяется величина Q, полученная согласно алгоритму (4.41). При этом, если условия $k \leqslant Z_n(Q)$ или $k \leqslant N$ не удовлетворяются, то величины Q или k соответственно изменяются. Выходными результатами работы программы являются значения оценок коэффициентов модели $\hat{a}_i$, $i = 1(1)k$ для отобранных полиномов, а также упорядоченное множество соответствующих этим полиномам показателей степеней $(q_n, \ldots, q_1)_i$.

4.3.4. Пример

Продемонстрируем рассмотренный метод на примере. Из $K = 5$ полиномов $n = 2$ переменных x_1 и x_2 требуется отобрать $k = 3$ полинома с граничными значениями показателей степени $Q_1 = 1$ и $Q_2 = 2$. Так как $Z_2(2) = 5$, то граничное значение для степени полиномов $Q = 2$. Для выбора множества допустимых полиномов имеем

$$\begin{aligned} f_1 &= 1, \\ f_2 &= x_1, \\ f_3 &= x_2, \\ f_4 &= x_1 x_2, \\ f_5 &= x_2^2, \end{aligned} \qquad \begin{array}{c|cc} \tilde{y} & x_1 & x_2 \\ \hline 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 2 \end{array} \qquad (4.43)$$

$$s = 0{,}02, \quad \varphi_e = 4.$$

Выборка наблюдений приведена в (4.43). На первом шаге рассчитываем величины a_{11}, $\hat{b}_1$, V_1 для всех пяти функций и определяем функцию с максимальным значением V_1:

$$\begin{aligned} a_{11} &= [f^2], \\ \hat{b}_1 &= a_{11}^{-1}[\tilde{y}f], \\ V_1 &= a_{11}(\hat{b}_1)^2, \\ S_0 &= [\tilde{y}^2] = 14, \\ S_1 &= S_0 - V_1. \end{aligned}$$

f	a_{11}	$\hat{b}_1$	V_1	
f_1	4	1,500	9,000*	
f_2	2	1,500	4,500	
f_3	5	0,400	0,800	$S_1 = 5$
f_4	4	0,000	0,000	
f_5	17	0,118	0,236	

Критическое значение F-распределения для уровня значимости $\alpha = 0{,}01$ и чисел степеней свободы $\varphi_1 = 1$ и $\varphi_2 = \varphi_e = 4$, согласно

табл. 15.5, равно $F_{кр}=21{,}2$. Пользуясь рассчитанными значениями V_1 и s, находим [1])

$$F_1=\frac{V_1}{s^2}=\frac{9}{4}\cdot 10^4>21{,}2=F_{кр}.$$

Таким образом, в ортогональный базис первой включается функция $f_1=1$.

На втором шаге выбирается лучшая функция из $f=f_2, f_3, f_4, f_5$:

$$a_{12}=a_{11}^{-1}[f_1f],$$
$$a_{22}=[f^2]-a_{11}a_{12}^2,$$
$$\hat{b}_2=a_{22}^{-1}([\tilde{y}f]-\hat{b}_1a_{11}b_{12}),$$
$$V_2=a_{22}(\hat{b}_2)^2,\quad S_1=S_2-V_2,$$

f	$[f_1f]$	a_{12}	a_{22}	$[\tilde{y}f]$	$\hat{b}_2$	V_2	
f_2	2	0,500	1,000	3	0,000	0,000	
f_3	3	0,750	2,750	2	—0,999	2,268	$S_2=2$
f_4	2	0,500	3,000	0	—1,000	3,000*	
f_5	5	1,250	10,750	2	—0,512	2,812	

Второй функцией в ортогональный базис включается функция f_3. Применение F-критерия дает и в этом случае положительный результат, так как

$$F_2=\frac{V_2}{s^2}=\frac{3}{4}\cdot 10^4>21{,}2=F_{кр}.$$

На последнем шаге выбираем лучшую функцию из $f=f_2, f_3, f_5$:

$$a_{13}=a_{11}^{-1}[f_1f],$$
$$a_{23}=a_{22}^{-1}([f_4f]-a_{11}a_{12}a_{13}),$$
$$a_{33}=[f^2]-a_{11}a_{13}^2-a_{22}a_{23}^2,$$
$$\hat{b}_3=a_{33}^{-1}([\tilde{y}f]-\hat{b}_1a_{11}a_{13}-\hat{b}_2a_{22}a_{23}),$$
$$V_3=a_{33}(\hat{b}_3)^2,\quad S_3=S_2-V_3.$$

f	a_{13}	$[f_4f]$	a_{23}	a_{33}	$\hat{b}_3$	V_3	
f_2	0,500	2	0,333	0,667	1,500	1,500	*
f_3	0,750	4	0,834	0,667	0,000	0,000	$S_3=0{,}5$
f_5	1,250	8	1,834	0,667	0,000	0,000	

[1]) См. примечания на стр. 94. — *Прим. ред.*

Так как $F_3 = V_3/s^3 = (1{,}5/4)\cdot 10^4 = 3750 > 21{,}2$, то в ортогональный базис включается функция f_3. Таким образом, для включения в модель отобраны функции f_1, f_4 и f_3.

Расчет коэффициентов проводится по уравнениям

$$\begin{aligned} a_{31} &= -a_{13}, & \hat{a}_3 &= \hat{b}_3, \\ a_{21} &= -(a_{13} + a_{23}a_{31}), & \hat{a}_2 &= \hat{b}_2 + \hat{b}_3 a_{23}, \\ a_{33} &= -a_{33}, & \hat{a}_1 &= \hat{b}_1 + \hat{b}_2 a_{21} + \hat{b}_3 a_{31}. \end{aligned}$$

Подставляя численные значения, получаем

$$f = \begin{bmatrix} 1 \\ x_1x_2 \\ x_1 \end{bmatrix}, \quad A = \begin{bmatrix} 4{,}00 & 0{,}50 & 0{,}50 \\ -0{,}50 & 3{,}00 & 0{,}33 \\ -0{,}33 & -0{,}33 & 0{,}67 \end{bmatrix}, \quad \hat{b} = \begin{bmatrix} 1{,}5 \\ -1{,}0 \\ 1{,}5 \end{bmatrix}, \quad \hat{a} = \begin{bmatrix} 1{,}5 \\ -1{,}5 \\ 1{,}5 \end{bmatrix}.$$

В результате получаем следующее выражение для модели вида $y = a'f$:

$$\hat{y} = \frac{3}{2}(1 + x_1 - x_1x_2).$$

Остаточная сумма квадратов $S_R = 0{,}5$. Ясно, что линейная модель, содержащая функции $f' = (1, x_1, x_2)$, в данном случае не является наилучшей. Как и в примере разд. 4.2.7, здесь не требуется знания ортогональных полиномов.

4.4. Построение полиномиальной модели на прямоугольной решетке

4.4.1. Постановка задачи

Построение модели существенно упрощается, если, во-первых, использовать в качестве ее составляющих полиномы, а во-вторых, применять специальные планы.

Пусть план состоит из множества не обязательно равноотстоящих друг от друга узлов прямоугольной решетки; область Ω_x изменения вектора независимых переменных представляет собой при этом n-мерный прямоугольный параллелепипед. Множество точек измерения можно описать следующим образом:

$$\{x_1^{k_1}, x_2^{k_2}, \ldots, x_n^{k_n}, \quad k_l = 0\,(1)\,K_l, \quad l = 1\,(1)\,n\}. \qquad (4.44)$$

Нижний индекс l у независимых переменных x означает номер переменной, верхний индекс k_l является номером гиперплоскости, на которой лежит узел, а x_l^k — фиксированная координата, определяющая эту гиперплоскость. Значения выходной перемен-

ной y в узлах решетки $(k_1, k_2, \ldots, k_n)$ целесообразно обозначить через

$$\{\tilde{y}^{k_1 k_2 \ldots k_n}, \quad k_l = 0\,(1)\,K_l, \quad l = 1\,(1)\,n\}. \tag{4.45}$$

Процесс построения модели в этом случае упрощается за счет того, что используются ортогональные полиномы только одной переменной. Эти полиномы задаются двухступенчатым рекуррентным уравнением (см. разд. 4.4.3). Кроме того, оказываются легко выполнимыми процедуры вычисления коэффициентов, которые достаточно изящно реализуются с помощью средств вычислительной техники (см. разд. 4.4.2 и 4.4.5). Этот метод является расширением метода, предложенного Форсайтом [3] для одномерной функции, на случай произвольного числа переменных n. Метод рекомендуется применять в задачах, где сложную и измеримую в области Ω_x функцию нескольких переменных нужно описать с помощью полиномиальной модели.

4.4.2. Вычисление коэффициентов

Представим модель в следующем виде:

$$y = P(x_1, \ldots, x_n) = \sum_{\substack{q_l = 0\,(1)\,Q_l \\ l = 1\,(1)\,n}} a_{q_1 q_2 \ldots q_n} \prod_{l=1}^{n} P_{q_l}(x_l). \tag{4.46}$$

Здесь $P_{q_l}(x_l)$ — полином степени q_l от одной переменной x_l. Простейший способ построения таких полиномов состоит в выборе их в виде $P_{q_l}(x_l) = x_l^{q_l}$, однако лучше сделать эти полиномы ортогональными. Степень полинома q_l ограничена:

$$0 \leqslant q_l \leqslant Q_l, \qquad l = 1\,(1)\,n. \tag{4.47}$$

Число полиномов Q, входящих в (4.46), равно числу определяемых коэффициентов модели $a_{q_1 \ldots q_n}$

$$Q = \prod_{l=1}^{n} (Q_l + 1). \tag{4.48}$$

Так как число K имеющихся в нашем распоряжении значений $\tilde{y}^{k_1 \ldots k_n}$ равно числу узлов решетки

$$K = \prod_{l=1}^{n} (K_l + 1), \tag{4.49}$$

то для получения оценок $\hat{a}_{q_1 \ldots q_n}$ необходимо, чтобы выполнялось неравенство

$$Q_l \leqslant K_l, \qquad l = 1\,(1)\,n. \tag{4.50}$$

Для нахождения оценок $\hat{a}_{q_1,\ldots,q_n}$ методом наименьших квадратов имеем [1]

$$\sum_{q_l=0\,(1)\,Q_l \atop l=1\,(1)\,n} \hat{a}_{q_1,\ldots,q_n} \prod_{l=1}^{n} \sum_{k_l=1}^{K_l} P_{q_l}(x_l^{k_l}) P_{r_l}(x_l^{k_l}) = \left\{ \sum_{k_l=0}^{K_l} P_{r_l}(x_l^{k_l}) \right\} \tilde{y}^{k_1,\ldots,k_n}_{l=1\,(1)\,n}, \tag{4.51}$$

$$r_l = 0\,(1)\,Q_l, \quad l = 1\,(1)\,n.$$

Если потребовать, чтобы полиномы для каждой переменной были ортогональны, т. е.

$$\sum_{k_l=0}^{K_l} P_r(x_l^{k_l}) P_s(x_l^{k_l}) = 0 \text{ для } r \neq s, \quad l = 1\,(1)\,n, \tag{4.52}$$

то матрица коэффициентов системы уравнений (4.51) становится диагональной и оценки определяются по формуле

$$\hat{a}_{q_1,\ldots,q_n} = \frac{\left\{ \sum_{k_l=0}^{K_l} P_{q_l}(x_l^{k_l}) \right\} \tilde{y}^{k_1,\ldots,k_n}_{l=1\,(1)\,n}}{\prod_{l=1}^{n} \sum_{k_l=0}^{K_l} P_{q_l}^2(x_l^{k_l})}. \tag{4.53}$$

Для $n = 1$ получаем результат, известный из работы [3]. Расчеты коэффициента $\hat{a}_{q_1,\ldots,q_n}$ можно проводить по рекуррентной формуле с использованием средств вычислительной техники. Обозначим

$$a_{|\,k_1 k_2 \ldots k_n} = \tilde{y}^{k_1 k_2 \ldots k_n} \tag{4.54}$$

и введем коэффициенты $a_{q_1,\ldots,q_l\,|\,k_{l+1}\ldots k_n}$. Тогда можно записать следующее рекуррентное выражение для последовательного вычисления коэффициентов $a_{q_1\,|\,k_2\ldots k_n}$, $a_{q_1 q_2\ldots|\,k_2\ldots k_n}$, $\ldots$, $a_{q_1\ldots q_n\,|}$:

$$a_{q_1,\ldots,q_l\,|\,k_{l+1}\ldots k_n} = \frac{\sum_{k_l=0}^{K_l} P_{q_l}(x_l^{k_l})\, a_{q_1,\ldots,q_{l-1}\,|\,k_l\ldots k_n}}{\sum_{k_l=0}^{K_l} P_{q_l}^2(x_l^{k_l})}. \tag{4.55}$$

[1] Уравнение (4.51) является условной записью системы нормальных уравнений. В частности, правая часть уравнения (4.51) означает следующее выражение:

$$\sum_{k_1=0}^{K_1} \sum_{k_2=0}^{K_2} \cdots \sum_{k_n=0}^{K_n} \prod_{l=1}^{n} P_{r_l}(x_l^{k_l})\, \tilde{y}^{k_1 k_2 \ldots k_n}.$$

— *Прим. ред.*

Очевидно, что n-кратное применение выражения (4.55), исходя из заданных величин $\hat{a}_{|k_1 \ldots k_n}$, позволяет получить искомые оценки $\hat{a}_{q_1 \ldots q_n} = a_{q_1 \ldots q_n}$, причем число коэффициентов на каждом шаге уменьшается [1]. Существенным в данном методе является то, что на каждом шаге необходимо вычислять ортогональные полиномы одной переменной x_l.

4.4.3. Ортогональные полиномы

Особенностью рассмотренного в предыдущем разделе подхода является то, что фигурирующие в нем полиномы зависят только от одной переменной. Пусть x — некоторая переменная. Тогда ортогональные полиномы $P_l(x)$ могут быть найдены с помощью описанной в [3] двухступенчатой рекуррентной процедуры:

$$P_{l+1}(x) = (x - c_l) P_l(x) - d_l P_{l-1}(x), \qquad l = 0, 1, 2, \ldots, \tag{4.56}$$

при начальных значениях

$$P_{-1}(x) = 0, \qquad P_0(x) = 1. \tag{4.57}$$

Коэффициенты c_l и d_l определяют из условия ортогональности (4.52). Значения коэффициентов вычисляются по формулам

$$c_l = \frac{\sum_{k=0}^{K} x^k P_l^2(x^k)}{\sum_{k=0}^{K} P_l^2(x^k)}, \quad d_l = \frac{\sum_{k=0}^{K} x^k P_l(x^k) P_{l-1}(x^k)}{\sum_{k=0}^{K} P_{l-1}^2(x^k)}, \tag{4.58}$$

где индекс k_l заменен индексом k.

Для $l = 0$ в силу (4.56) имеем

$$c_0 = \frac{1}{K+1} \sum_{k=0}^{K} x^k, \qquad d_0 = 0. \tag{4.59}$$

С помощью уравнений (4.56) — (4.59) можно последовательно рассчитать величины c_l и d_l, начиная с $l = 1$. Если на l-м шаге известны коэффициенты c_l и d_l и значения полиномов P_l и P_{l-1} во всех точках x^k, $k = 0(1)K$, то с помощью уравнения (4.56) можно рассчитать значения полинома P_{l+1} в тех же точках. Вычисляя одновременно суммы произведений полиномов во всех точках, можно с помощью (4.58) найти c_{l+1} и d_{l+1}.

---

[1] Величины $a_{q_1 \ldots q_l | k_{l+1} \ldots k_n}$ определяют множество значений коэффициентов, каждый из которых соответствует одному из возможных сочетаний значений $k_{l+1}, \ldots, k_n$. Осуществляя последовательное суммирование по всем значениям одной из переменных, мы на каждом шаге уменьшаем число этих коэффициентов, получая в конце концов искомую оценку $\hat{a}_{q_1 \ldots q_n}$. — *Прим. ред.*

Для представления ортогональных полиномов целесообразно использовать нормальную форму [1]:

$$P_l(x) = \sum_{j=0}^{l} p_{lj}(x)^j. \tag{4.60}$$

Пользуясь уравнением (4.56), легко получить рекуррентное выражение для нахождения коэффициентов p_{lj}, удобное при проведении численных расчетов:

$$\begin{gathered} p_{l+1,\,j} = p_{l,\,j-1} - c_l p_{lj} - d_{l-1} p_{l-1,\,j}, \\ p_{00} = 1, \quad p_{lj} = 0 \quad \text{при} \quad j > l. \end{gathered} \tag{4.61}$$

В частности, для $j = l+1$ из (4.61) получаем $p_{l+1,\,l+1} = p_{ll}$, что при условии $p_{00} = 1$ дает $p_{jj} = 1$ для всех j. Очевидно, что величины c_l, d_l, p_{lj} и полиномы $P_l(x)$ зависят от выбора узлов решетки x^k, $k = 0(1)K$.

Методом полной индукции легко доказать, что все построенные таким образом полиномы ортогональны.

4.4.4. Принцип отбора

Сформулированный в разд. 4.2.4 принцип отбора функций можно применить и для плана эксперимента частного вида, разумеется в подходящей для этого случая форме. Пусть требуется установить, для какой переменной x_l повышение степени полинома от $q_l - 1$ до q_l приводит к максимальному уменьшению остаточной суммы квадратов S_R. Для $S_R(q_1, \ldots, q_n)$ после некоторых преобразований получаем

$$S_R = \sum_{\substack{k_l = 0\,(1)\,K_l \\ l = 1\,(1)\,n}} (\tilde{y}^{k_1 \ldots k_n})^2 - \sum_{\substack{t_j = 0\,(1)\,q_j \\ j = 1\,(1)\,n}} \hat{a}^2_{t_1 \ldots t_n} \prod_{r=1}^{n} \sum_{k_r=0}^{K_r} P^2_{t_r}(x^{k_r}_r). \tag{4.62}$$

Таким образом, задача сводится к нахождению максимума суммы неотрицательных величин:

$$\max_{l = 1\,(1)\,n} \sum_{\substack{t_l = q_l \\ t_j = 0\,(1)\,q_j - 1 \\ j \neq l}} \hat{a}^2_{t_1 \ldots t_n} \prod_{r=1}^{n} N(P_{t_r}) = V.$$

Другие возможные формулировки принципа выбора, связанные, например, с повышением степени полинома лишь для одной функции, т. е. только для составляющей остаточной суммы

[1] Здесь $(x)^j$ означает x в степени j. — *Прим. ред.*

квадратов

$$\hat{a}^2_{t_1 \ldots t_n} \prod_{r=1}^{n} N(P_{t_r}),$$

оказываются слишком трудоемкими для вычислений.

С помощью F-критерия можно аналогично разд. 4.3 проверить значимость изменения остаточной суммы квадратов при повышении степени переменной x_l до значения q_l. Пусть s^2 — оценка дисперсии ошибок наблюдений; тогда, если $F = V/s^2 > F_{кр}$ при заданном уровне значимости α и числах степеней свободы $\varphi_1 = \prod_{\substack{i=1 \\ i \neq l}}^{n} (1 + q_i)$ и $\varphi_2 = \varphi_e$ [1], следует увеличить c $q_l - 1$ до q_l максимальную степень полиномов от переменной x_l.

4.4.5. Модель в нормальной форме

Уравнение (4.55) позволяет получить оценки $\hat{a}_{q_1 \ldots q_n}$ параметров модели вида (4.46), представленной с помощью полиномов, введенных в разд. 4.4.3. Для полного описания модели необходимо располагать значениями коэффициентов $\hat{a}_{q_1 \ldots q_n}$ и вспомогательными величинами c_l и d_l, т. е. иметь всего

$$\prod_{l=1}^{n} (Q_l + 1) + 2 \sum_{l=1}^{n} (Q_l + 1)$$

значений.

При желании модель может быть представлена с помощью нормальных полиномов

$$P(x_1, \ldots, x_n) = \sum_{\substack{t_l = 0\,(1)\,Q_l \\ l = 1\,(1)\,n}} \hat{b}_{t_1 t_2 \ldots t_n} (x_1)^{t_1} (x_2)^{t_2} \ldots x_n)^{t_n}. \quad (4.63)$$

Для этого оценки $\hat{a}_{q_1 \ldots q_n}$ следует пересчитать на $\hat{b}_{t_1 \ldots t_n}$. Полином (4.63) в этом случае однозначно определяется $\prod_{l=1}^{n} (Q_l + 1)$ значениями $\hat{b}_{t_1 \ldots t_n}$. При этом пересчет целесообразно осуществлять рекуррентно. Введем промежуточный полином вида

$$P(x_1, \ldots, x_n) = \sum_{\substack{q_l = 0\,(1)\,Q_l \\ l = 1\,(1)\,r}} \sum_{\substack{t_l = 0\,(1)\,Q_l \\ l = r+1\,(1)\,n}} b_{q_1 \ldots q_r | t_{r+1} \ldots t_n} \times$$

$$\times \prod_{j=1}^{r} P_{q_j}(x_j) \prod_{j=r+1}^{n} (x_j)^{t_j}.$$

[1] Здесь φ_e — число степеней свободы оценки s^2; φ_1 — число функций, добавляемых в модель при увеличении на единицу максимальной степени одной из переменных. — *Прим. ред.*

Из сопоставления коэффициентов следует, что

$$b_{q_1\ldots q_n l} = \hat{a}_{q_1\ldots q_n}, \tag{4.64}$$

$$\sum_{l_r=0\,(1)\,Q_r} b_{q_1\ldots q_{r-1} l l_r\ldots l_n} (x_r)^{l_r} \equiv \sum_{q_r=0\,(1)\,Q_r} b_{q_1\ldots q_r l_{r+1}\ldots l_n} P_{q_r}(x_r).$$

Важно отметить, что последнее тождество содержит только одну переменную x_r. Применяя теперь нормальную форму (4.60) ортогональных полиномов и сопоставляя коэффициенты при x_r, получим формулу пересчета:

$$b_{q_1\ldots q_{r-1} l l_r\ldots l_n} = \sum_{q_r=l_r}^{Q_r} b_{q_1\ldots q_r l_{r+1}\ldots l_n} P_{q_r}(x_r), \tag{4.65}$$

$$r = n\,(-1)\,1, \quad \hat{b}_{l_1\ldots l_n} = b_{l l_1\ldots l_n}.$$

Коэффициенты p_{lj} в ортогональных полиномах для переменных x_r определяем путем вычислений по рекуррентной формуле (4.61). Для получения искомых оценок $\hat{b}_{l_1\ldots l_n}$ коэффициентов модели в нормальной форме (4.63) необходимо для каждого коэффициента провести расчеты по (4.65) n раз, начиная с $r = n$.

4.4.6. Указания к программной реализации метода

Алгоритмы, содержащиеся в разд. 4.4.2, 4.4.3 и 4.4.5, реализованы в программе ORPOL. Исходными данными для вычислений являются:

n — число независимых переменных;

K_i, $i = 1\,(1)\,n$ — число уровней i-й переменной;

$x_i^{k_i}$, $k_i = 0\,(1)\,K_i$, $i = 1\,(1)\,n$ — координаты узлов решетки;

$\tilde{y}^{k_1\ldots k_n}$ — тензор значений выходной переменной;

Q_i, $i = 1\,(1)\,n$ — показатель степени переменной $Q_i \leqslant K_i$.

Основная проблема программной реализации метода состоит в накоплении многократно выбираемых значений тензора $\tilde{y}^{k_1\ldots k_n}$. Так как n — переменная величина, то $\tilde{y}$ необходимо размещать в некотором линейном поле. При поиске значения $\tilde{y}$, соответствующего некоторому узлу решетки, уместно ввести числовую величину

$$Z = \sum_{i=1}^{n} k_i \prod_{j=1}^{i-1} (K_j + 1) = (k_n, k_{n-1}, \ldots, k_1). \tag{4.66}$$

Величина Z изменяется аналогично десятичному числу, только перенос десятков заменяется K_i-переносом. Для заданных n, K_i и k_i значение Z легко определяется по формуле (4.66). Для об-

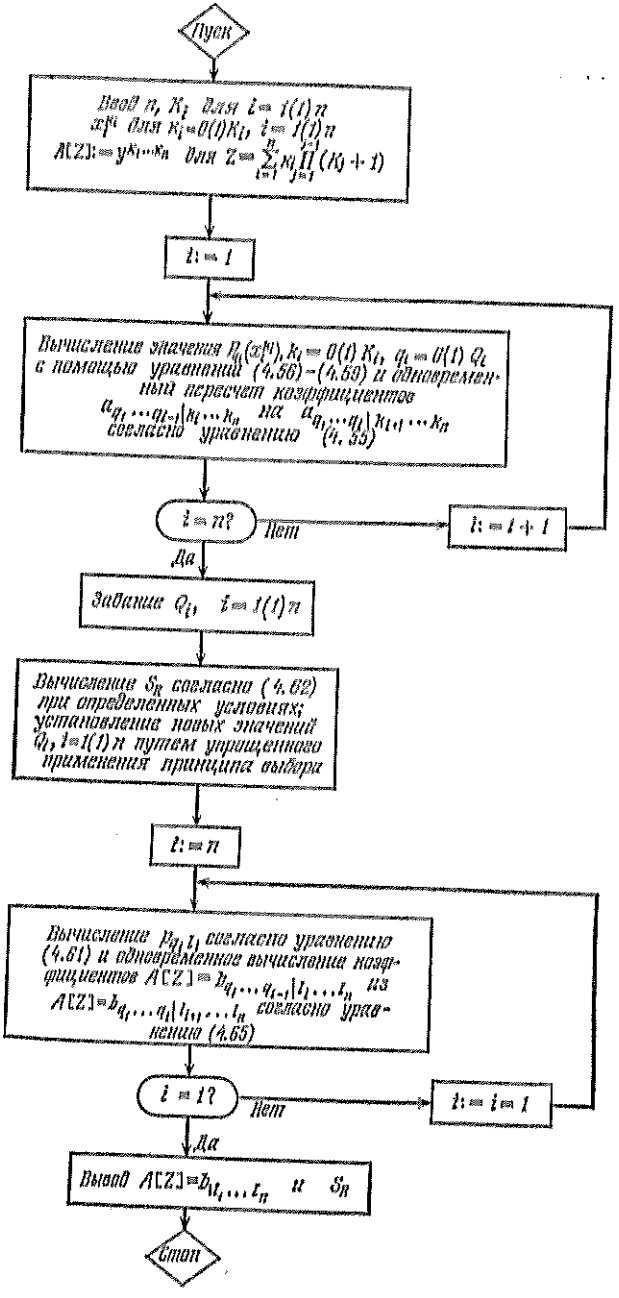
ратного вычисления значений $k_1, \ldots, k_n$ по заданным Z и K_l можно воспользоваться программой

```
kØ := Z;
for l := 1 step 1 until n do
begin h := K[l] + 1;
      k1 := entier (kØ /h);
      k[l] := kØ — h * k1;
      kØ := k1;
end
```

Эту процедуру применяют при вводе или вычислении тензора значений y. Блок-схема программы ORPOL, реализующей расчеты рассмотренным методом, приведена на фиг. 4.3. Важно отметить, что данные промежуточных $2n$ ступеней расчетов коэффициентов a и b размещаются в линейном массиве A, который к началу вычислений содержит значения выходной переменной $\tilde{y}^{k_1 \ldots k_n}$. Как для рекуррентного получения a, согласно уравнению (4.55), так и для получения b по уравнению (4.65) используется только поле массива $A[1:K]$, $K = \prod_{l=1}^{n} (K_l + 1)$. Правильный выбор элементов из массива A для вычисления сумм и последующее помещение данных в массив A на некотором l-м шаге осуществляется с помощью числовой операции, завершающей l-й цикл, который является внутренним циклом в процессе построения сумм:

```
      for j := 1 step 1 until n do q[j] := Ø;
M1: j := 1;
      for q[l] := Ø step 1 until Q[l] do
      Summe;
M2: q[j] := q[j] + 1;
      if q[j] > Q[j] then
      begin q[j] := Ø;
            j := j + 1;
            if j = l then j := j + 1;
            if j > n then goto M3 else goto M2;
      end else goto M1;
M3:
```

Выдача результатов возможна в трех видах. Это могут быть либо коэффициенты полиномиальной модели вида (4.46), либо коэффициенты модели (4.63), использующей нормальные полиномы, либо одновременно оба множества коэффициентов. В первом случае выводятся также вспомогательные коэффициенты c и d ортогональных полиномов.



Фиг. 4.3. Блок-схема программы ORPOL.

4.4.7. Пример

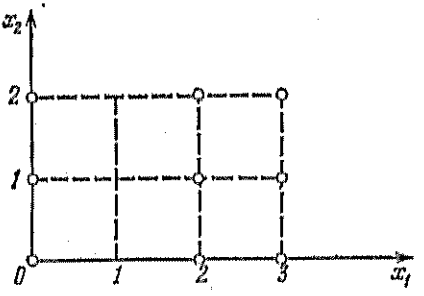
Продемонстрируем на примере изложенный в разд. 4.4 метод построения модели. Пусть задана выборка

$$x_1 = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \tilde{y} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix},$$

т. е. заданы значения выходной переменной y в узлах решетки, показанных на фиг. 4.4. Требуется оценить коэффициенты модели вида

$$y = a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2.$$

В данном случае $K_1 = K_2 = 2$ и $Q_1 = Q_2 = 1$. На первом шаге находим ортогональные полиномы $P_l(x_1)$ и коэффициенты $a_{k_1k_2|} =$



Фиг. 4.4. План эксперимента для примера разд. 4.4.7.

$= \tilde{y}^{k_1k_2}$ пересчитываем на $a_{q_1|k_2}$. Согласно уравнению (4.52),

$$c_0 = \frac{1}{K_1 + 1} \sum_{k_1=0}^{K_1} x_1^{k_1} = \frac{5}{3};$$

при этом

$$P_0(x_1) = 1, \quad P_1(x_1) = x_1 - \frac{5}{3}.$$

Пересчет коэффициентов осуществляем согласно уравнению (4.55):

$$a_{q_1|k_2} = \frac{\sum_{k_1=0}^{K_1} P_{q_1}(x_1^{k_1})\, a_{k_1k_2|}}{\sum_{k_1=0}^{K_1} P_{q_1}^2(x_1^{k_1})}.$$

В силу того что

$$\sum_{l=0}^{3} P_0^2(x_1^l) = 3, \quad \sum_{l=0}^{3} P_1^2(x_1^l) = \frac{14}{3},$$

имеем

$$\left.\begin{aligned} a_{0|j} &= \frac{1}{3}(a_{|0j} + a_{|1j} + a_{|2j}) \\ a_{1|j} &= \frac{1}{14}(-5a_{|0j} + a_{|1j} + 4a_{|2j}) \end{aligned}\right\} \quad j = 0, 1, 2$$

и

$$(a_{q_1|k_2}) = \begin{pmatrix} 4/3 & 1 & 2 \\ 1/14 & 9/14 & 9/14 \end{pmatrix}.$$

На втором шаге подобную процедуру проделываем с ортогональными полиномами от x_2:

$$P_0(x_2) = 1, \quad P_1(x_2) = x_2 - 1,$$

$$a_{q_1 q_2|} = \frac{\sum\limits_{k_2=0}^{K_2} P_{q_2}(x_2^{k_2})\, a_{q_1|k_2}}{\sum\limits_{k_2=0}^{K_2} P_{q_2}^2(x_2^{k_2})}.$$

Так как

$$\sum_{l=0}^{3} P_0^2(x_2^l) = 3, \quad \sum_{l=0}^{3} P_1^2(x_2^l) = 2,$$

имеем

$$\left.\begin{aligned} a_{i0|} &= \frac{1}{3}(a_{i|0} + a_{i|1} + a_{i|2}) \\ a_{i1|} &= \frac{1}{2}(-a_{i|0} + a_{i|2}) \end{aligned}\right\} \quad i = 0, 1.$$

Окончательно получаем

$$(a_{q_1 q_2|}) = (\hat{a}_{q_1 q_2}) = \begin{pmatrix} 13/19 & 1/3 \\ 19/42 & 2/7 \end{pmatrix}.$$

Для S_R находим, согласно (4.62),

$$S_R = 25 - \left(\frac{13}{9}\right)^2 3\times 3 - \left(\frac{1}{3}\right)^2 \times 3 \times 2 - \left(\frac{19}{42}\right)^2 \times \frac{14}{3} \times 3 - \left(\frac{2}{7}\right)^2 \times \frac{14}{3} \times 2,$$

$$S_R = \frac{27}{14} \approx 1{,}929.$$

Полученные результаты позволяют представить модель с помощью ортогональных полиномов следующим образом:

$$\hat{y} = \frac{13}{9} + \frac{19}{42}\left(x_1 - \frac{5}{3}\right) + \frac{1}{3}(x_2 - 1) + \frac{2}{7}\left(x_1 - \frac{5}{3}\right)(x_2 - 1).$$

Представление модели с помощью нормальных полиномов требует пересчета коэффициентов согласно (4.65). Заменим в обозначениях коэффициентов b на a. Тогда для ортогональных

полиномов от переменной x_2 имеем

$$\left.\begin{aligned} a_{i|0} &= a_{i|0} + a_{i|1}p_{10}, \\ a_{i|1} &= a_{i|1} \end{aligned}\right\} \; i = 0, 1$$

при $p_{10} = -1$. Произведя вычисления, получаем

$$a_{q_1|i_2} = \begin{pmatrix} 10/9 & 1/3 \\ 1/6 & 2/7 \end{pmatrix}.$$

Наконец, на втором шаге устраняем ортогональные полиномы от x_1:

$$\left.\begin{aligned} a_{i0|j} &= a_{0i|j} + a_{1i|j}p_{10}, \\ a_{i1|j} &= a_{1i|j} \end{aligned}\right\} \; j = 0, 1$$

при $p_{10} = -5/3$. Искомый результат таков:

$$(a_{q_1q_2}) = \begin{pmatrix} 5/6 & -1/7 \\ 1/6 & 2/7 \end{pmatrix}.$$

Модель, представленная с помощью нормальных полиномов, имеет вид

$$\hat{y} = \frac{5}{6} + \frac{x_1}{6} - \frac{x_2}{7} + \frac{2x_1x_2}{7}.$$

В этом случае число использованных в модели коэффициентов вдвое меньше, чем при представлении модели с помощью ортогональных полиномов. Очевидно, что в алгоритме вычислений параметров модели ортогональные полиномы фигурируют лишь в неявном виде. Их явного представления при построении модели на основе нормальных полиномов не требуется.

ЛИТЕРАТУРА

1. Никифорова Е. С., Метод регрессионного анализа, см. сб. „Планирование эксперимента (алгоритмы на языке Алгол-60)", под ред. Э. К. Лецкого и Н. Н. Челнокова, Труды МЭИ, вып. 76, М., 1970.
2. Efroymson M. A., Multiple regression analysis, см. "Mathematical Methods for Digital Computer", Ed. A. Ralston a. H. S. Wilf, Wiley, New York, 1960.
3. Forsythe G. E., Generation and use of ortogonal polynomials to data-fitting with a digital computer, Y. Soc. Ind. Appl. Math., 1957, S. 74.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Дрейпер Н., Смит Г., Прикладной регрессионный анализ, изд-во „Статистика", М., 1973.

5.

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

5.1. Вводные замечания

Задачи и методы этого раздела непосредственно связаны с описанным в разд. 3 методом наименьших квадратов.

В разд. 3 мы исходили из того, что целевая величина y (выходная величина, зависимая переменная или отклик) следующим образом зависит от вектора независимых переменных $x = (x_1, x_2, \ldots, x_n)$:

$$y(a, x) = a'f(x). \tag{5.1}$$

Для нахождения оценок $\hat{a}$ вектора неизвестных параметров a в определенных точках x^l поставлен эксперимент и получен его результат $\tilde{y}^l$ (см. разд. 3.3.3). Оценки $\hat{a}$ вектора коэффициентов рассчитываются с помощью метода наименьших квадратов на основе выборки x^l, $l = 1, \ldots, N$, путем решения системы нормальных уравнений

$$(F'F)\,\hat{a} = F'\tilde{Y}. \tag{5.2}$$

Матрицы F и $\tilde{Y}$ определяются через (3.78) и (3.81) соответственно. Из (5.2) следует

$$\hat{a} = (F'F)^{-1}F'\tilde{Y} = CF'\tilde{Y}. \tag{5.3}$$

Если

$$\tilde{y}^l = E\{\tilde{y}^l\} + \tilde{e}^l = \overline{y}^l + \tilde{e}^l = \overline{a}'f(x^l) + \tilde{e}^l, \tag{5.4а}$$

то

$$E\{\tilde{Y}\} = F\overline{a}. \tag{5.4б}$$

Здесь $\overline{y}^l$ — истинное значение зависимой переменной в точке x^l; $\tilde{e}^l$ — случайная ошибка в точке x^l. Ошибка e^e считается независимой случайной величиной с нулевым математическим ожиданием и дисперсией σ^2 [см. формулы (3.25), (3.27) и (3.29)]. При этих условиях оценки $\hat{a}$ вектора параметров являются случайными величинами с ковариационной матрицей (3.84):

$$\operatorname{cov}\{\hat{a}\} = (F'F)^{-1}\sigma^2 = C\sigma^2. \tag{5.5}$$

Кроме того, оценки являются несмещенными:

$$E\{\hat{a}\} = \overline{a} \tag{5.6}$$

и, согласно (3.32), обладают минимальной дисперсией среди всех возможных несмещенных линейных оценок для заданной выборки x^i, $i = 1 \div N$. В этом смысле метод наименьших квадратов является оптимальным методом обработки данных. Эта оптимальность имеет место при заданной выборке. Достигаемая точность оценок, как это следует, например, из (5.5), будет зависеть от выбора экспериментальных точек, или, другими словами, от условий проведения опытов. В этом и заключается основная идея планирования эксперимента: добиться требуемых свойств (например, максимальной точности), выбирая условия проведения опытов [1].

Применение методов планирования эксперимента предполагает возможность проведения опытов в заданных исследователем условиях. Такой способ проведения эксперимента называют иногда активным.

В общем случае активный подход к эксперименту в сочетании с методами планирования позволяет получить требуемые результаты, затратив минимальные средства и время на проведение исследования.

5.2. Основные понятия планирования эксперимента

Определение 5.1

Множество всех точек проведения экспериментов

$$x^i = (x_1^i, x_2^i, \ldots, x_n^i), \quad i = 1, 2, \ldots, N, \tag{5.7}$$

представляется с помощью матрицы плана

$$X = (x_j^i) = \begin{bmatrix} x_1^1 & x_2^1 & \ldots & x_n^1 \\ x_1^2 & x_2^2 & \ldots & x_n^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_1^N & x_2^N & \ldots & x_n^N \end{bmatrix} = \begin{bmatrix} (x^1)^1 \\ (x^2)^1 \\ \cdot \\ \cdot \\ \cdot \\ (x^N)^1 \end{bmatrix} \tag{5.8}$$

и называется планом эксперимента. (План эксперимента, заданный с помощью матрицы плана X, будет в дальнейшем обозначаться через X.)

Определение 5.2

Точка

$$x^0 = N^{-1} \sum_{i=1}^{N} x^i \tag{5.9}$$

называется центром плана (центральной точкой плана). Каждая координата x_j^0 вектора x^0 является средним значением i-х коор-

динат всех точек плана

$$x_i^0 = \frac{x_i^1 + x_i^2 + \ldots + x_i^N}{N}. \qquad (5.10)$$

Определение 5.3

План называется центральным, если его центр расположен в начале координат $x = 0$, т. е.

$$x^0 = 0. \qquad (5.11)$$

Очевидно, что оптимальное планирование связано с разработкой планов, представляемых в некоторой стандартной форме. При этом целесообразно рассматривать центральные планы. Всякий план путем переноса начала координат может быть сделан центральным.

Утверждение 5.1

Всякий план Z с точками z^l путем замены

$$x = z - z^0 \qquad (5.12)$$

при

$$z^0 = N^{-1} \sum_{l=1}^{N} z^l$$

может быть преобразован в центральный план X с точками x^l.

Определение 5.4

Область возможных значений независимых переменных называется областью планирования эксперимента. Будем обозначать эту область Ω_x. Все точки x^l плана X должны принадлежать области Ω_x. Это требование мы будем записывать следующим образом:

$$x^l \in \Omega_x \qquad (5.13)$$

или

$$X \in \Omega_x. \qquad (5.14)$$

Независимые переменные x часто называют варьируемыми переменными или факторами. Область планирования эксперимента может быть задана, например, с помощью неравенств

$$-1 \leqslant x_i \leqslant 1, \quad i = 1, 2, \ldots, n. \qquad (5.15)$$

В этом случае говорят, что областью планирования является гиперкуб.

В выражении (5.15) предельные (максимальное и минимальное) значения варьируемых переменных обозначены через $+1$ и -1 соответственно. Эти границы отвечают стандартизованному или нормированному масштабу изменения переменных.

Переход к стандартизованному масштабу может быть осуществлен, например, следующим образом:

$$x_i = \frac{x_i^* - \frac{x_i^*\text{макс} + x_i^*\text{мин}}{2}}{\frac{x_i^*\text{макс} - x_i^*\text{мин}}{2}}. \qquad (5.16)$$

В формуле (5.16) x_i^* — значение i-й переменной в натуральном масштабе измерения.

Ряд необходимых нам в дальнейшем свойств плана X связан с видом модели (5.1), для оценки коэффициентов которой план используется.

Определение 5.5

Матрица $M = F'F$ размера $(k+1)\times(k+1)$ называется информационной матрицей плана X [2][1]. Здесь

$$F = \begin{bmatrix} f_0(x^1) & \dots & f_k(x^1) \\ \vdots & & \\ f_0(x^N) & \dots & f_k(x^N) \end{bmatrix}. \qquad (5.17)$$

Ясно, что информационная матрица плана X зависит от выбора функций $f_0(x), \dots, f_k(x)$.

Определение 5.6

План X называется ортогональным, если информационная матрица диагональная:

$$M = F'F = \begin{bmatrix} \sum_{j=1}^{N} f_0^2(x^j) & 0 \dots & 0 \\ 0 & \sum_{j=1}^{N} f_1^2(x^j) & \dots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \dots & \sum_{j=1}^{N} f_k^2(x^j) \end{bmatrix}. \qquad (5.18)$$

Матрица M есть матрица системы нормальных уравнений (5.3), из решения которой находятся оценки коэффициентов модели. Поэтому для ортогонального плана вычисления оказываются чрезвычайно простыми.

[1] Иногда матрицу $F'F$ называют также матрицей моментов плана, а информационной матрицей — нормированную матрицу моментов, т. е. матрицу $(1/N)\,F'F$, где N — число опытов в плане. — *Прим. ред.*

Определение 5.7

План X называется ротатабельным, если дисперсия оценки $\hat{y}(x)$ зависимой переменной в точке x зависит только от расстояния точки x от центра плана x^0. В соответствии с (3.85) дисперсия $\sigma^2_{\hat{y}(x)}$ оценки зависимой переменной выражается в виде

$$\sigma^2_{\hat{y}(x)} = f'(x)(F'F)^{-1} f(x)\sigma^2, \tag{5.19}$$

где $(F'F)^{-1}$ — дисперсионная матрица плана. Обозначим расстояние точки x от центра плана x^0 через r:

$$r = \sqrt{(x-x^0)'(x-x^0)} = \sqrt{(x_1-x_1^0)^2 + \ldots + (x_n-x_n^0)^2}. \tag{5.20}$$

Тогда условие ротатабельности плана имеет вид

$$f'(x)(F'F)^{-1} f(x) = \text{const} \tag{5.21}$$

при $r = \sqrt{(x-x^0)'(x-x^0)} = \text{const}$. Таким образом, ротатабельные планы обеспечивают одно и то же значение дисперсии оценки целевой величины во всех точках, равноудаленных от центра плана.

Определение 5.8

Зависимость $\xi(r) = 1/[Nf'(x)Cf(x)]$ называется информационным профилем ротатабельного плана. Информационный профиль плана показывает характер изменения дисперсии оценки зависимой переменной при удалении от центра плана.

Пример 5.1

Пусть модель имеет вид

$$y(a, x) = a_0 + a_1x_1 + a_2x_2. \tag{5.22}$$

Область планирования Ω_x определяется неравенствами

$$-1 \leqslant x_1 \leqslant 1,$$
$$-1 \leqslant x_2 \leqslant 1.$$

Рассмотрим план X

$$X = \begin{bmatrix} -1 & -1 \\ +1 & -1 \\ -1 & +1 \\ +1 & +1 \end{bmatrix}. \tag{5.23}$$

Нетрудно убедиться, что центр плана

$$x^0 = 0,$$

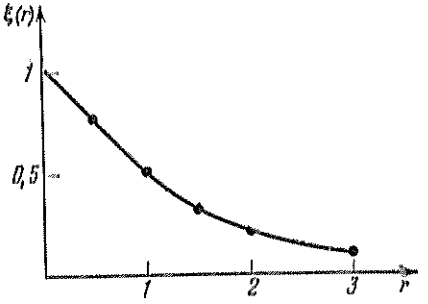
и, следовательно, план X — центральный. Матрица F имеет вид

$$F = \begin{bmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \end{bmatrix}.$$

Информационная матрица M плана X

$$M = F'F = 4I_3$$

диагональная, и, следовательно, план X ортогонален.



Фиг. 5.1. Информационный профиль ротатабельного плана примера 5.1.

Проверим условие (5.21) ротатабельности плана

$$f'(x)(F'F)^{-1}f(x) = \frac{1}{4}(1 + x_1^2 + x_2^2) = \frac{1}{4}(1 + r^2). \qquad (5.24)$$

Дисперсия $\sigma^2_{\hat{y}(x)}$ зависит только от расстояния r точки x до центра плана, и, следовательно, план X ротатабельный.

На фиг. 5.1 показан информационный профиль плана (5.23). Легко убедиться, что свойство ротатабельности плана X зависит от вида модели. Пусть, например, план X, определяемый выражением (5.23), используется для вычисления оценок коэффициентов модели такого вида:

$$y(a, x) = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2.$$

Тогда матрица $F'F$ получается равной

$$F'F = 4I_4.$$

Проверка условия ротатабельности дает

$$f'(x)(F'F)^{-1}f(x) = \frac{1}{4}(1 + x_1^2 + x_2^2 + x_1^2x_2^2) = \frac{1}{4}(1 + r^2 + x_1^2x_2^2). \qquad (5.25)$$

Дисперсия $\sigma^2_{\hat{y}(x)}$ в этом случае зависит не только от расстояния между точкой x и центром плана; следовательно, план X не является ротатабельным.

Определение 5.9

План X называется ненасыщенным, если $N > k+1$, и насыщенным, если $N = k+1$. Здесь N — число точек спектра плана [1], а $(k+1)$ — число оцениваемых коэффициентов.

5.3. Общие требования к плану эксперимента. О критериях планирования эксперимента

В разд. 3 мы при определении оценок параметров модели исходили из того, что результаты наблюдений получены в некотором заданном множестве точек. На практике, однако, часто оказывается возможным свободно выбирать условия проведения опытов в пределах некоторых границ. Выбор числа и условий проведения эксперементов, обеспечивающих получение наилучшего в определенном смысле результата исследования, и составляет цель планирования эксперимента. Разработан ряд критериев оптимальности планов эксперимента, важнейшие из которых будут рассмотрены ниже. Следует отметить, что оптимальный выбор плана эксперимента существенным образом зависит от конкретных особенностей исследуемого объекта, таких, как вид его модели, стоимость отдельных опытов, области варьирования независимых переменных и т. п.

Одной из важнейших характеристик плана, влияющей, с одной стороны, на стоимость и длительность исследования, а с другой, — на точность результатов, является число экспериментов. Заметим, что план с минимально возможным числом экспериментов $N = k+1$ (насыщенный план) не позволяет проверить адекватность модели [2]. Поэтому обычно выбирают $N > k+1$, где $k+1$ — число оцениваемых параметров модели.

К вопросу определения числа наблюдений N, достаточному для достижения требуемой точности модели, мы вернемся в разд. 11. Кроме того, в разд. 6.6 будет рассмотрена задача определения числа наблюдений N с учетом стоимости экспериментов.

---

[1] Спектром плана называется совокупность всех точек плана, отличающихся уровнем хотя бы одного фактора. — *Прим. ред.*

[2] Проверка адекватности модели по способу, описанному в разд. 3.2.5, возможна лишь тогда, когда число различных точек плана (т. е. число точек в спектре плана) превосходит число оцениваемых параметров $(k+1)$. — *Прим. ред.*

Важное значение для оценки качества плана эксперимента имеет вид информационной матрицы плана

$$M = F'F.$$

Матрица M должна быть невырожденной, т. е. $|M| \neq 0$. Только в этом случае система линейных уравнений, к которой приводит критерий наименьших квадратов, имеет единственное решение.

Перечислим теперь некоторые критерии планирования эксперимента, используемые в практических исследованиях.

1. Критерий ортогональности плана

Критерий ортогональности требует такого выбора плана X для оценки коэффициентов модели заданного вида, при котором информационная матрица плана диагональна. Примеры построения некоторых ортогональных планов для моделей, описываемых полиномами первой и второй степени, приводятся в разд. 5.5—5.7.

Использование критерия ортогональности имеет целью упрощение вычислений и получение независимых оценок коэффициентов. Как легко видеть, при ортогональном планировании матрица $(F'F)^{-1}$ является диагональной и, следовательно, ковариации оценок коэффициентов равны нулю. Это значит, например, что замена нулем любого коэффициента в уравнении модели не изменит значений оценок остальных коэффициентов. Такое свойство ортогональных планов оказывается очень полезным, когда точный вид модели неизвестен и исследователь использует экспериментальные данные для отбора переменных, существенно влияющих на выходную величину.

2. Критерий ротатабельности [4]

Этот критерий требует такого расположения экспериментальных точек в области планирования Ω_x, при котором дисперсия $\sigma^2_{\hat{y}}$ оценки значений зависимой переменной в точке x зависит только от расстояния от этой точки до центра плана. Такой критерий хорошо согласуется с требованием равнозначности (с точки зрения точности оценки зависимой переменной) всех направлений от центра плана.

Названные критерии обеспечивают некоторые полезные и удобные свойства оценок коэффициентов, однако они никак не связаны с требованием максимальной точности построения модели. Критерии, приводимые ниже, обеспечивают оптимальность планов эксперимента с точки зрения точности оценки параметров модели или зависимой переменной.

3. Критерий A-оптимальности [2, 5]

Этот критерий требует такого выбора плана X, при котором матрица $C = (F'F)^{-1}$ имеет минимальный след (т. е. сумма диагональных элементов матрицы C минимальна). Так как в соответствии с (3.38) диагональный элемент c_{ii} матрицы C пропорционален дисперсии оценки i-го коэффициента, то критерий A-оптимальности по существу требует минимизации средней дисперсии оценок коэффициентов модели.

4. Критерий D-оптимальности [2, 5]

Этот критерий требует такого расположения точек в области Ω_x, при котором определитель матрицы $C = (F'F)^{-1}$ минимален (или определитель матрицы $M = F'F$ максимален). В разд. 6 мы рассмотрим этот критерий подробнее. Здесь укажем, что D-оптимальный план минимизирует объем эллипсоида рассеяния оценок коэффициентов.

5. Критерий G-оптимальности [6]

Критерий G-оптимальности требует такого расположения точек в области Ω_x, при котором достигается наименьшая величина максимальной дисперсии оценки зависимой переменной в области Ω_x. В отличие от критериев A- и D-оптимальности, связанных с точностью нахождения коэффициентов, критерий G-оптимальности требует максимальной точности оценки зависимой переменной.

В литературе рассматриваются также и другие критерии планирования, связанные с точностью построения модели [5][1]. Заметим, что достижение возможно большей точности модели связано с лучшим использованием области планирования при проведении эксперимента. Поэтому при использовании критериев 3, 4 и 5 вид области планирования является очень важным условием задачи и произвольное изменение ее конфигурации приводит к существенному изменению оптимального плана.

Все перечисленные критерии связаны с предположением, что вид модели известен. Однако на практике часто возникает такая ситуация, когда исследователь не знает истинного вида модели. В таком случае эксперимент обычно сначала планируется исходя из простейшего предположения о линейности модели относительно варьируемых переменных (разд. 3.3). После проведения опытов проверяется адекватность линейной модели. Если линейная модель неадекватна, делается попытка построить квадратич-

[1] См. также работы [1, 7] в списке дополнительной литературы.— *Прим. ред.*

ную модель (разд. 3.4). При этом нужно планировать и эксперимент для квадратичной модели. Оптимальный план для квадратичной модели целесообразно строить таким образом, чтобы он включал точки оптимального плана для линейной модели. Такое построение плана приводит к сокращению числа опытов. Планы для квадратичных моделей, построенные путем добавления точек к плану для линейной модели, называются композиционными планами второго порядка. В этом разделе мы рассмотрим примеры построения композиционных планов.

На практике иногда возникают ситуации, когда для построения модели выбирают полином более низкой степени, чем этого требует действительная зависимость. Причиной этого может быть, например, необходимость в слишком большом числе экспериментов для построения оценок действительной модели. Кроме того, иногда исследователя могут интересовать не все, а лишь часть коэффициентов [например, при реализации крутого восхождения (разд. 12) исследователю достаточно знать лишь коэффициенты при линейных функциях]. В этих случаях эксперимент целесообразно планировать так, чтобы сделать минимальным смещение в оценках коэффициентов, связанное с неадекватным выбором модели (разд. 3.2.4). Можно также потребовать минимизации среднего по области Ω_x квадрата отклонения рассчитанных по модели значений от истинных значений зависимой переменной [3, 7]. Эти требования могут использоваться как критерии выбора плана эксперимента. Список возможных критериев планирования эксперимента можно было бы продолжить.

На практике часто полезно стремиться к тому, чтобы один и тот же план удовлетворял одновременно ряду критериев, например был бы D-оптимальным и ортогональным и т. п. В общем случае такого сочетания свойств не наблюдается, однако иногда в одном плане сочетается ряд полезных свойств. Например, легко убедиться в справедливости следующего утверждения.

Утверждение 5.2

Ортогональный план для линейной модели (3.3) одновременно является и ротатабельным [2], если $(F'F)^{-1} = C = gI$, g — константа, I — единичная матрица.

В самом деле, согласно (5.19), дисперсия оценки зависимой переменной в точке x для таких планов равна

$$\sigma^2_{\hat{y}(x)} = f'(x)(F'F)^{-1} f(x)\sigma^2 = g\sigma^2\left(1 + \sum_{i=1}^{n} x_i^2\right).$$

Если центр плана находится в точке $x^0 = 0$, то величина $\sigma^2_{\hat{y}(x)}$ зависит лишь от расстояния точки x до центра плана.

В разд. 6 мы, кроме того, рассмотрим ряд условий, при выполнении которых план является одновременно D-, A- и G-оптимальным.

Выбор критерия оптимальности плана осуществляется исходя из конкретного содержания решаемой задачи. Часто бывает нецелесообразным отказываться от таких полезных свойств планов, как ортогональность или ротатабельность. Действительно, потеря ортогональности приводит к существенным усложнениям вычислений, в то время как повышение точности за счет перехода к неортогональному плану может оказаться незначительным. Если истинный вид модели бывает не известен исследователю, то при планировании более существенным может оказаться возможность проверки адекватности модели, нежели строго оптимальный с точки зрения точности построения модели предполагаемого вида выбор экспериментальных точек. Естественно, что встречаются и иные задачи, где вид модели известен, а интерес представляет получение наивысшей точности оценок параметров модели.

Во всяком случае, всегда будет полезным анализ используемых планов с точки зрения различных критериев. Ниже мы в этом разделе рассмотрим ортогональные и ротатабельные планы для моделей, описываемых полиномами первой и второй степени относительно независимых переменных. В разд. 6 мы рассмотрим свойства этих планов с точки зрения A-, D- и G-оптимальности.

5.4. Планы для моделей, описываемых полиномами первого порядка

5.4.1. Вид модели

Здесь будут рассмотрены планы, предназначенные для построения линейных моделей процессов вида

$$y(x, a) = a_0 + a_1 x_1 + \ldots + a_n x_n. \tag{5.26}$$

Обозначив через F матрицу

$$F = (l, X), \tag{5.27}$$

где

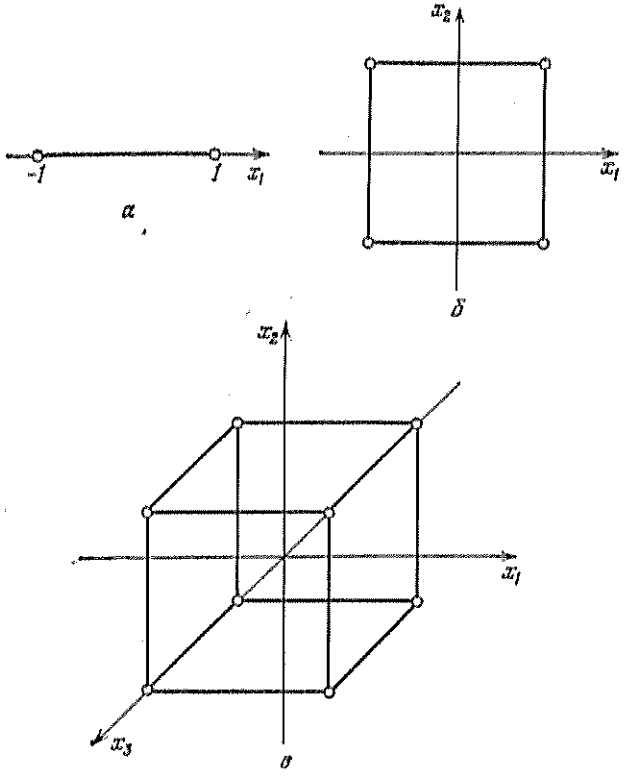
$$l = (1, 1, \ldots, 1)', \tag{5.28}$$

получим для информационной матрицы плана выражение

$$M = F'F = (m_{ij}) = \begin{pmatrix} N & l'X \\ X'l & X'X \end{pmatrix}. \tag{5.29}$$

5.4.2. Полные факторные планы

Мы здесь ограничимся рассмотрением планов, в которых каждый фактор x_l принимает значения только на двух уровнях. Без ограничения общности можно считать, что эти значения



Фиг. 5.2. Полные факторные планы для $n=1$ (*а*), $n=2$ (*б*) и $n=3$ (*в*).

суть $+1$ и -1. В дальнейшем мы будем обозначать эти величины соответственно знаком „+“ или „−“.

Определение 5.10

Множество всех точек в n-мерном пространстве, координаты которых являются $+1$ (+) или -1 (−), называется полным факторным планом типа 2^n. Число точек в этом плане

$$N=2^n. \tag{5.30}$$

Пример 5.2

При n, равном 1, 2 и 3, матрицы планирования X для факторных планов 2^n имеют вид

$$X_1 = \overset{n=1}{\begin{bmatrix} + \\ - \end{bmatrix}}, \quad X_2 = \overset{n=2}{\begin{bmatrix} + & + \\ - & + \\ + & - \\ - & - \end{bmatrix}}, \quad X_3 = \overset{n=3}{\begin{bmatrix} + & + & + \\ - & + & + \\ + & - & + \\ - & - & + \\ + & + & - \\ - & + & - \\ + & - & - \\ - & - & - \end{bmatrix}}.$$

Точки этих планов показаны на фиг. 5.2. Легко получить следующее общее правило построения полных факторных планов.

Утверждение 5.3

Матрица планирования X_{n+1} факторного плана 2^{n+1} может быть получена с помощью матрицы X_n плана 2^n по формуле

$$X_{n+1} = \begin{pmatrix} X_n & l \\ X_n & -l \end{pmatrix}, \tag{5.31}$$

где l определяется уравнением (5.28).

5.4.3. Дробные факторные планы

Число опытов $N = 2^n$ полных факторных планов быстро растет с увеличением размерности факторного пространства n, так что при больших n эти планы оказываются практически неприемлемыми. При этом из множества точек факторных планов 2^n может быть отобрана некоторая часть, представляющая дробный факторный план и содержащая подходящее число опытов.

Правило построения дробных факторных планов

Для построения дробного факторного плана типа 2^{n-p} из множества n отбирают $n-p$ основных факторов, для которых строят полный факторный план с матрицей X_{n-p}. Этот план дополняют затем p столбцами, соответствующими оставшимся факторам. Каждый из этих p столбцов получается как результат поэлементного перемножения не менее двух и не более $n-p$ определенных столбцов, соответствующих основным факторам. Для определения способа образования каждого из p столбцов дробного факторного плана вводится понятие генератора плана.

Генератор представляет собой произведение основных факторов, определяющее значение элементов каждого из дополнительных p столбцов матрицы плана. Очевидно, что в случае плана типа 2^{n-p} должно иметься p генераторов.

Построение дробного факторного плана для $n = 3$

Исходим из факторного плана 2^2 для основных факторов x_1 и x_2 и дополняем этот план столбцом значений третьего фактора, элементы которого являются произведением соответствующих элементов первого и второго столбцов (здесь это единственная возможность определения столбца для фактора x_3):

x_1	x_2	$x_3 = x_1x_2$
−	−	+
+	−	−
−	+	−
+	+	+

Выражение $x_1x_2 = x_3$ является генератором плана. Преимуществом этого плана является меньшее число опытов ($2^{3-1} = 4$) по сравнению с числом опытов полного факторного плана ($2^3 = 8$). Полученный дробный план является полурепликой факторного плана 2^3.

Построение дробных факторных планов для $n = 4, 5, 6, 7$

Для этих размерностей факторного пространства можно построить дробные планы, содержащие восемь опытов. Исходим из полного факторного плана 2^3 для факторов x_1, x_2 и x_3 и дополняем его столбцами, образованными поэлементными произведениями столбцов плана 2^3: x_1x_2, x_1x_3, x_2x_3, $x_1x_2x_3$; эти произведения могут являться генераторами для дробных планов. Используя один из четырех возможных генераторов, можно построить четыре различных дробных плана типа 2^{4-1}:

$$x_4 = \begin{cases} x_1x_2 \\ x_1x_3 \\ x_2x_3 \\ x_1x_2x_3 \end{cases}.$$

По сравнению с $2^4 = 16$ опытами полного факторного плана полученный дробный план состоит из $2^{4-1} = 8$ опытов.

При $n = 5$ для построения дробного плана типа 2^{5-2} имеется возможность выбрать два любых из четырех возможных генераторов для образования столбцов факторов x_4 и x_5. Очевидно, что возможно построение шести различных вариантов плана

типа 2^{5-2}:

$$x_4 = \begin{cases} x_1x_2 \\ x_1x_3 \\ x_1x_2 \\ x_1x_3 \\ x_1x_3 \\ x_2x_3 \end{cases}, \qquad x_5 = \begin{cases} x_1x_2 \\ x_2x_3 \\ x_1x_2x_3 \\ x_2x_3 \\ x_1x_2x_3 \\ x_1x_2x_3 \end{cases}.$$

Еще шесть дробных планов можно получить, если поменять местами столбцы для факторов x_4 и x_5.

Для пяти факторов дробный план типа 2^{5-2} содержит восемь опытов по сравнению с $2^5 = 32$ опытами полного факторного плана.

При $n = 6$ для построения дробного плана 2^{6-3} необходимо выбрать три из четырех возможных генераторов для получения элементов столбцов факторов x_4, x_5 и x_6. Возможны следующие комбинации:

$$x_4 = \begin{cases} x_1x_2 \\ x_1x_2 \\ x_1x_3 \\ x_1x_3 \end{cases}, \quad x_5 = \begin{cases} x_1x_3 \\ x_1x_3 \\ x_2x_3 \\ x_2x_3 \end{cases}, \quad x_6 = \begin{cases} x_2x_3 \\ x_1x_2x_3 \\ x_1x_2x_3 \\ x_1x_2x_3 \end{cases}.$$

Рассматривая теперь всевозможные перестановки трех полученных столбцов, получим 24 варианта дробных планов типа 2^{6-3}. Дробный план типа 2^{7-4} для $n = 7$ приведен в табл. 5.1.

Дробные факторные планы для n от 5 до 15

Исходим из факторного плана 2^4 и дополняем его всевозможными поэлементными произведениями столбцов плана 2^4: x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, $x_1x_2x_3$, $x_1x_3x_4$, $x_1x_2x_4$, $x_2x_3x_4$, $x_1x_2x_3x_4$.

Таблица 5.1

x_1	x_2	x_3	$x_4 = x_1x_2$	$x_5 = x_1x_3$	$x_6 = x_2x_3$	$x_7 = x_1x_2x_3$
+	+	+	+	+	+	+
−	+	+	−	−	+	−
+	−	+	−	+	−	−
−	−	+	+	−	−	+
+	+	−	+	−	−	−
−	+	−	−	+	−	+
+	−	−	−	−	+	+
−	−	−	+	+	+	−

Выбирая необходимое число генераторов для оставшихся факторов, строим дробные планы типа 2^{5-1}, 2^{6-2}, 2^{7-3}, 2^{8-4}, ..., 2^{15-11}, каждый из которых содержит 16 опытов. Заметим, что полный факторный план 2^{15} требует постановки 32 768 экспериментов. Преимущества дробных планов с точки зрения числа опытов очевидны.

5.4.4. Формулы для вычислений и свойства полных и дробных факторных планов для линейных моделей

Матрица F для плана типа 2^{n-p} (при $p=0$ имеем полный факторный план) и линейной модели вида (5.26) содержит $n+1$ столбцов и $N=2^{n-p}$ строк. Например, для плана 2^{3-1} имеем

$$F=\begin{bmatrix}1 & -1 & -1\\ 1 & +1 & -1\\ 1 & -1 & +1\\ 1 & +1 & +1\end{bmatrix}. \tag{5.32}$$

Первый столбец соответствует фиктивной переменной x_0 при свободном члене уравнения модели, которая во всех опытах принимает значение, равное единице. Легко убедиться в том, что информационная матрица плана 2^{n-p} для модели (5.26) имеет вид

$$M=F'F=2^{n-p}I_{n+1}=NI_{n+1}, \tag{5.33}$$

где I_{n+1}—единичная матрица размера $n+1$.

Для дисперсионной матрицы получаем

$$C=M^{-1}=\frac{1}{N}I_{n+1}. \tag{5.34}$$

Из (5.34) следуют простые формулы для оценок коэффициентов

$$\hat{a}_i=\frac{1}{N}\sum_{l=1}^{N}\tilde{y}^l x_i^l, \quad i=0, 1, \ldots, n. \tag{5.35}$$

Для дисперсий оценок коэффициентов s_i^2 получаем

$$s_i^2=\frac{s^2}{N}. \tag{5.36}$$

Здесь s^2—оценка дисперсии $\tilde{y}$ (дисперсии ошибки наблюдений). Из (5.36) следует, что оценки всех коэффициентов имеют одну и ту же дисперсию. Вид выражений (5.33) и (5.34) определяет ортогональность плана 2^{n-p} для модели (5.26). Легко убедиться, что для линейной модели план 2^{n-p} является также и ротатабельным (см. утверждение 5.2).

В разд. 6.4 будет показано, что полные и дробные факторные планы для линейных моделей вида (5.26), в случае когда область планирования — гиперкуб с координатами вершин ± 1, являются также D-, A- и G-оптимальными планами (см. утверждение 6.5 и пример 6.5). Это означает, что для данного частного вида области планирования и числа опытов $N = 2^{n-p}$ рассмотренные планы обеспечивают максимально возможную точность оценок коэффициентов и всей модели в целом.

Перечисленные свойства факторных планов объясняют, почему эти планы находят широкое применение при построении линейных моделей.

Пример 5.3

При изучении возможностей повышения выхода одной из производных пиперазина рассматривались факторы, приведенные в табл. 5.2. Постулировался линейный вид зависимости между

Таблица 5.2

Факторы и их интервалы варьирования

Факторы	Значения факторов для примера 5.3 и их интервалы варьирования			
	−1	0	+1	Интервал варьирования
x_1^* — отношение NaOH/ Исходный продукт 1	1:1	1,25:1	1,5:1	0,25
x_2^* — отношение Исходный продукт 1/Исходный продукт 2	1:1	1,25:1	1,5:1	0,25
x_3^* — длительность реакции, ч	3	4	5	1
x_4^* — температура, °С	20	25	30	5
x_5^* — момент добавления исходного продукта 1, мин	20	40	60	20

выходом и указанными факторами. Был использован дробный факторный план типа 2^{5-2}, т. е. четверть реплики полного факторного плана 2^5, содержащий всего восемь опытов. Полный факторный план 2^5 содержит 32 опыта. План эксперимента и его результаты приведены в табл. 5.3. Для построения плана использованы генераторы $x_1x_2x_3$ и $-x_1x_2$, определяющие характер изменения в плане факторов x_4 и x_5 соответственно. Как ясно из табл. 5.3, опыты плана 2^{5-2} реализовывались в случайном порядке. Оценки коэффициентов, рассчитанные на основе результатов экспериментов, приведены в последней строке табл. 5.3.

Таблица 5.3

План 2^{5-2} и результаты опытов (пример 5.3)

Номер опыта	x_0	x_1	x_2	x_3	x_4	x_5	$\tilde{\tilde{y}}$ (среднее двух параллельных наблюдений)
1	+	−	−	−	−	−	50,0
2	+	+	+	−	−	−	57,2
3	+	−	−	+	+	−	48,1
4	+	+	+	+	−	−	46,0
5	+	−	−	−	+	+	64,8
6	+	+	+	−	+	+	45,3
7	+	−	−	+	+	+	54,8
8	+	+	+	+	+	−	53,0
$\hat{a}_l$	52,3	−1,755	5,050	0,575	−2,10	0,325	

Для проверки адекватности полученной модели рассчитаем, согласно (3.93), сумму квадратов S_e:

$$S_e = 51{,}2$$

с числом степеней свободы $\varphi_2 = N(\nu - 1) = 8$. По формуле (3.96) найдем сумму квадратов S_D:

$$S_D = 47{,}216$$

с числом степеней свободы $\varphi_1 = N - k - 1 = 8 - 5 - 1 = 2$. Отсюда следует

$$F = \frac{S_D/\varphi_1}{S_e/\varphi_2} = \frac{47{,}216/2}{51{,}2/8} = 3{,}69.$$

В табл. 15.4 находим для $1 - P = 0{,}05$

$$F_{кр} = 4{,}46.$$

Следовательно, полученная модель процесса является адекватной. По формуле (3.95) рассчитываем

$$s^2 = \frac{S_e}{\nu\varphi_2} = \frac{51{,}2}{2 \times 8} = 3{,}2$$

и $s = 1{,}79$. Для дисперсий оценок коэффициентов получаем

$$s_l^2 = \frac{s^2}{N} = \frac{3{,}2}{8} = 0{,}4,$$

$$s_l = 0{,}63.$$

Проверка значимости коэффициентов при $1-P=0{,}05$, $\varphi=8$ и соответственно $t_{кр}=2{,}31$ показывает, что коэффициенты a_3 и a_5 незначимы и, следовательно, могут быть исключены из модели. Окончательно получаем

$$\hat{y}(x)=52{,}3-1{,}755x_1+5{,}05x_2-2{,}1x_4.$$

5.5. Планы для моделей, содержащих линейные члены и взаимодействия различного порядка

5.5.1. Вид модели

При построении модели часто недостаточно принимать во внимание только линейные эффекты факторов, ибо влияние на целевую величину могут оказывать также взаимодействия факторов. В этих случаях в модель необходимо вводить взаимодействия различных порядков. Модель принимает вид

$$y(a,x)=a_0+\sum_{l=1}^{n}a_lx_l+\sum_{l=1}^{n}\sum_{k=l+1}^{n}a_{lk}x_lx_k+$$
$$+\sum_{l=1}^{n}\sum_{k=l+1}^{n}\sum_{t=k+1}^{n}a_{lkt}x_lx_kx_t+\ldots+a_{12\ldots n}x_1x_2\ldots x_n. \qquad (5.37)$$

Коэффициент a_{lk} является мерой парного взаимодействия факторов (взаимодействия первого порядка), коэффициент a_{lkt} отражает воздействие тройного взаимодействия (взаимодействия второго порядка). Количество возможных взаимодействий для числа факторов от 2 до 10 приведено в табл. 5.4. Для шести факторов существует, например, 15 взаимодействий первого, 20—второго, 15—третьего, 6—четвертого и 1—пятого порядка.

5.5.2. Применение полных факторных планов для моделей типа (5.37)

Для получения оценок коэффициентов модели типа (5.37) в принципе можно использовать полные факторные планы. Однако, как правило, модель включает не все, а лишь некоторые взаимодействия первого порядка (парные взаимодействия), иногда взаимодействия второго порядка и почти никогда не содержит взаимодействий выше третьего порядка. Поэтому число степеней свободы для проверки адекватности модели с ростом числа факторов быстро увеличивается. Так, например, для случая, когда имеются только взаимодействия первого порядка, при $n=2$ отсутствуют степени свободы, а при $n=6$ имеется уже 42 степени свободы. Так как наша цель состоит в том, чтобы, пользуясь по возможности малым числом опытов, извлечь необ-

Таблица 5.4

Количество взаимодействий для числа факторов от 2 до 10

n	$N=2^n$	Число линейных эффектов	Порядок взаимодействия								
			1	2	3	4	5	6	7	8	9
2	4	2	1								
3	8	3	3	1							
4	16	4	6	4	1						
5	32	5	10	10	5	1					
6	64	6	15	20	15	6	1				
7	128	7	21	35	35	21	7	1			
8	256	8	28	56	70	56	28	8	1		
9	512	9	36	84	126	126	84	36	92	1	
10	1024	10	45	120	210	252	210	120	45	10	1

ходимую информацию об исследуемом объекте, оказывается целесообразным для построения модели типа (5.37) применить рассмотренные в разд. 5.4 дробные факторные планы.

5.5.3. Применение дробных факторных планов для модели типа (5.37) и порядок смешивания оценок коэффициентов

Дробные планы типа 2^{n-p} для моделей, содержащих линейные члены и взаимодействия, строятся так же, как и для линейных моделей, т. е. матрицу полного факторного плана для $(n-p)$ основных факторов дополняют столбцами, элементы которых представляют собой произведения элементов определенных столбцов основных факторов. Обычно предполагается, что только некоторые парные взаимодействия и взаимодействия высших порядков являются значимыми. При этом значимые взаимодействия рассматриваются как самостоятельные факторы, а незначимые приравниваются к факторам, не вошедшим в число основных. При использовании дробных планов для моделей с взаимодействиями можно, таким образом, включать в рассмотрение лишь столько дополнительных факторов, сколько существует незначимых взаимодействий.

При наличии в модели взаимодействий оценки $\hat{a}_i$ коэффициентов при линейных членах остаются независимыми друг от друга, однако они могут быть смешаны со взаимодействиями высших порядков. Часть оценок $\hat{a}_{ik}$ коэффициентов при парных взаимодействиях также оказывается смешанной друг с другом.

Рассмотрим матрицу F для полной модели вида (5.37), содержащей все возможные взаимодействия, в случае, когда для оценки коэффициентов используется дробный факторный план

2^{n-p}. Очевидно, что некоторые столбцы матрицы F окажутся одинаковыми. Так, например, столбцы для фактора x_4 и взаимодействия $x_1x_2x_3$ одинаковы, если в качестве генератора выбрано соотношение $x_4 = x_1x_2x_3$. Это означает, что план не позволяет получить раздельные оценки для коэффициентов $\bar{a}_4$ и $\bar{a}_{123}$ модели. С помощью данного плана можно получить лишь оценку $\hat{a}$, которая характеризует суммарное воздействие фактора x_4 и взаимодействия $x_1x_2x_3$. Оценки подобного рода называют смешанными. Заметим, что если $\bar{a}_{123} = 0$, то величина $\hat{a}$ является несмешанной оценкой коэффициента $\bar{a}_4$.

Для получения правила смешивания, с помощью которого можно было бы определить, совокупность каких линейных эффектов и эффектов взаимодействия оценивается всяким найденным на основе данного плана коэффициентом, введем понятие контраста плана. Под контрастом понимается соотношение между элементами матрицы F, задающее элемент первого столбца матрицы F. Элементы первого столбца, всегда равные единице, обозначим символически через I. Для дробного плана 2^{3-1}, например, имеем следующий контраст[1]:

$$I = x_1x_2x_3.$$

Для дробного плана 2^{4-1} (если генератор $x_4 = x_1x_2x_3$) контраст выражается соотношением

$$I = x_1x_2x_3x_4.$$

Чтобы определить, с какими факторами смешана оценка некоторого данного фактора, умножим обе части контраста на этот фактор, считая, что $x_i^2 = 1$. При этом мы получим порядок смешивания оценок коэффициентов при использовании данного плана.

Пример 5.4

Для дробного факторного плана 2^{3-1} с контрастом $I = x_1x_2x_3$ получаем следующий порядок смешивания для факторов x_1, x_2 и x_3:

$$x_1 = x_1^2x_2x_3 = x_2x_3,$$
$$x_2 = x_1x_2^2x_3 = x_1x_3,$$
$$x_3 = x_1x_2x_3^2 = x_1x_2.$$

Для оценок имеем соответственно

$$\hat{a}_1 \longrightarrow \bar{a}_1 + \bar{a}_{23},$$
$$\hat{a}_2 \longrightarrow \bar{a}_2 + \bar{a}_{13},$$
$$\hat{a}_3 \longrightarrow \bar{a}_3 + \bar{a}_{12}.$$

[1] Предполагается, что $x_ix_i = 1$. Для плана 2^{3-1} генератор $x_3 = x_1x_2$. Поэтому $x_1x_2x_3 = 1$. — *Прим. ред.*

Для дробного факторного плана 2^{4-1} с контрастом $I = x_1x_2x_3x_4$ получаем аналогично

$$\begin{aligned}
&\hat{a}_1 \to \bar{a}_1 + \bar{a}_{234}, \quad \hat{a}_{12} \to \bar{a}_{12} + \bar{a}_{34},\\
&\hat{a}_2 \to \bar{a}_2 + \bar{a}_{134}, \quad \hat{a}_{13} \to \bar{a}_{13} + \bar{a}_{24},\\
&\hat{a}_3 \to \bar{a}_3 + \bar{a}_{124}, \quad \hat{a}_{14} \to \bar{a}_{14} + \bar{a}_{23}.\\
&\hat{a}_4 \to \bar{a}_4 + \bar{a}_{123},
\end{aligned}$$

В зависимости от выбора генераторов получаются дробные факторные планы, обладающие различной разрешающей способностью. В соответствии с порядком контраста (числом элементов в контрасте) говорят о планах с разрешающей способностью III, если контраст состоит из трех элементов, и IV, если контраст состоит из четырех элементов. Если для дробного плана 2^{4-1} в качестве генератора выбрано соотношение $x_1x_2x_3$ и контраст соответственно представляется выражением $I = x_1x_2x_3x_4$, то этот план имеет разрешающую способность IV и обозначается через 2^{4-1}_{IV}. Если, например, в качестве генератора выбрано соотношение x_1x_2, то план имеет контраст $I = x_1x_2x_4$, его разрешающая способность III и обозначение 2^{4-1}_{III}.

Дробные факторные планы с наибольшей разрешающей способностью называют главными. Этим планам следует отдавать предпочтение при использовании.

Если имеется несколько незначимых взаимодействий, можно ввести в план несколько дополнительных факторов. При этом число различных генераторов и контрастов, определенных так же, как и выше, будет таким же, как число дополнительных факторов. Чтобы определить порядок смешивания для этого случая, введем обобщающий контраст, который строится из отдельных контрастов, а также произведений отдельных контрастов во всевозможных сочетаниях по 2, 3, ..., p. При этом все произведения контрастов, так же как и сами контрасты, задают элементы первого столбца матрицы F.

Пример 5.5

Для дробного факторного плана 2^{5-2} в качестве генераторов выбраны соотношения $x_4 = x_1x_2$ и $x_5 = x_1x_2x_3$. Контрасты плана $I = x_1x_2x_4$ и $I = x_1x_2x_3x_5$.

Для получения обобщающего контраста перемножим вышеуказанные контрасты и получим еще один контраст $I = x_2x_4x_5$. Обобщающий контраст при этом

$$I = x_1x_3x_4 = x_2x_4x_5 = x_1x_2x_3x_5.$$

Умножая все составляющие обобщающего контраста на фак-

торы, находим совпадающие столбцы матрицы F:

$$x_1 = x_3x_4 = x_1x_2x_4x_5 = x_2x_3x_5,$$
$$x_2 = x_1x_3x_4 = x_4x_5 = x_1x_3x_5,$$
$$x_3 = x_1x_4 = x_2x_3x_4x_5 = x_1x_2x_5,$$
$$x_4 = x_1x_3 = x_2x_5 = x_1x_2x_3x_4x_5,$$
$$x_5 = x_1x_3x_4x_5 = x_2x_4 = x_1x_2x_3,$$
$$x_1x_2 = x_2x_3x_4 = x_1x_4x_5 = x_3x_5,$$
$$x_2x_3 = x_1x_3x_4 = x_3x_4x_5 = x_1x_5.$$

Отсюда легко получить порядок смешивания оценок. Например, для $\hat{a}_1$ имеем

$$\hat{a}_1 \to \bar{a}_1 + \bar{a}_{34} + \bar{a}_{1345} + \bar{a}_{235}.$$

Если в модель входят функции вида x_i^2 ($i = 1, 2, \ldots, n$), то столбцы матрицы F, соответствующие этим функциям, будут состоять из единиц. Эти столбцы, таким образом, совпадают со столбцом для x_0 (т. е. для свободного члена уравнения). А это означает, что $\hat{a}_0$ является смешанной оценкой для свободного члена $\bar{a}_0$ и всех коэффициентов $\bar{a}_{ii}$, $i = 1, 2, \ldots, n$:

$$\hat{a}_0 \to \bar{a}_0 + \bar{a}_{11} + \bar{a}_{22} + \ldots + \bar{a}_{nn}.$$

Пользуясь этим свойством, можно получить правило для проверки значимости квадратичных эффектов. Для этого проведем в центре плана n_0 опытов и найдем значение $\tilde{y}^0$:

$$\tilde{y}^0 = \frac{1}{n_0}\sum_{j=1}^{n_0} \tilde{y}^{0j}. \qquad (5.38)$$

Если квадратичные эффекты отсутствуют, т. е. $\bar{a}_{11} = \bar{a}_{22} = \ldots = \bar{a}_{nn} = 0$, то

$$E\{\tilde{y}^0 - \hat{a}_0\} = 0.$$

Отсюда следует, что, в случае когда гипотеза H: $E\{\hat{a}_0\} = E\{\tilde{y}^0\}$ отвергается, в модель необходимо включать функции вида x_i^2.

Пусть полный (или дробный) факторный план содержит $N = 2^{n-p}$ точек, причем в каждой точке реализовано ν экспериментов. На основе параллельных опытов в каждой точке найдена оценка s^2 дисперсии ошибок наблюдений с числом степеней свободы φ_e [см. формулы (3.68) и (3.69)]. Для проверки гипотезы можно воспользоваться величиной

$$t = \frac{\tilde{y}^0 - \hat{a}_0}{s\sqrt{\dfrac{n_0 + \nu N}{n_0 N}}},$$

которая при условии нормальности распределения и независимости ошибок наблюдений подчинена закону распределения Стьюдента (t-распределению) с числом степеней свободы φ_e [15]. Гипотеза H: $E\{\hat{a}_0\} = E\{\tilde{y}^0\}$ отклоняется, если

$$|\tilde{y}^0 - \hat{a}_0| > t_{\text{кр}} s \sqrt{\frac{n_0 + \nu N}{n_0 N}}, \tag{5.39}$$

где $t_{\text{кр}}$—критическое значение распределения Стьюдента при выбранном уровне значимости α и числе степеней свободы [1] φ_e. Если гипотеза отвергается, в модель следует ввести квадраты факторов.

5.5.4. Вычислительные формулы и свойства планов 2^{n-p}

Легко убедиться в том, что информационная матрица планов 2^{n-p} для моделей, содержащих $(k+1)$ подлежащий оценке коэффициент, в случае когда оценки всех коэффициентов не смешаны (т. е. матрица F не имеет совпадающих столбцов), имеет следующий вид:

$$M = F'F = 2^{n-p} I_{k+1} = N I_{k+1}. \tag{5.40}$$

Для C получаем

$$C = \frac{1}{N} I_{k+1}. \tag{5.41}$$

Планы типа 2^{n-p} являются, таким образом, ортогональными для моделей вида (5.37). Для вычисления оценок коэффициентов получаем формулы

$$\hat{a}_i = \frac{1}{N} \sum_{j=1}^{N} x_i^j \tilde{y}^j, \quad i = 0, 1, \ldots, n, \tag{5.42}$$

$$\hat{a}_{i \ldots \mu} = \frac{1}{N} \sum_{j=1}^{N} x_i^j \ldots x_\mu^j \tilde{y}^j, \quad i, \mu = 1, 2, \ldots, n, \tag{5.43}$$

$$s_i^2 = \frac{s^2}{N}, \quad i = 0, 1, \ldots, k. \tag{5.44}$$

Здесь $(k+1)$—общее число коэффициентов модели (5.37). Из (5.44) следует, что все коэффициенты оцениваются с одинаковой точностью. Отметим, что планы типа 2^n и 2^{n-p} для моделей, содержащих взаимодействия, не являются ротатабельными (см. пример 5.1).

---

[1] Оценка s_e^2 дисперсии результатов единичного эксперимента связана с оценкой s^2 дисперсии ошибок наблюдений соотношением $s_e^2 = \nu s^2$. — *Прим. ред.*

Как и в случае линейных моделей, планы 2^{n-p} для моделей вида (5.37) являются A-, D- и G-оптимальными, если областью планирования эксперимента является гиперкуб с координатами вершин, принимающими значения ± 1 (см. утверждение 6.2 и пример 6.5). Эти свойства, разумеется, имеют место только в тех случаях, когда возможно получение несмешанных оценок всех коэффициентов модели.

Пример 5.6

При предварительных исследованиях было установлено, что на удельную теплопроводность возгона, получаемого при хлорировании титановых шлаков, влияют следующие факторы:

x_1 — плотность засыпки;
x_2 — содержание хлора в возгоне;
x_3 — отношение концентраций SiO_2 и TiO_2 в возгоне;
x_4 — температура.

Требуется экспериментальным путем найти интерполяционную формулу, с помощью которой можно было бы описать зависимость удельной теплопроводности от перечисленных четырех факторов и их парных взаимодействий:

$$y(a, x) = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_{12}x_1x_2 + a_{23}x_2x_3 + a_{34}x_3x_4.$$

В табл. 5.5 представлены интервалы варьирования факторов. Анализ механизма реакции показывает, что на теплопроводность возгона могут оказывать влияние следующие два взаимодействия: x_2x_3 и x_3x_4 (механизм проводимости).

Таблица 5.5

Факторы (пример 5.6)

	x_1	x_2	x_3	x_4
Основной уровень ($x_j = 0$)	0,87	40,0	1,00	250
Интервал варьирования	0,15	5,0	0,25	50
Верхний уровень ($x_j = +1$)	1,02	45,0	1,25	300
Нижний уровень ($x_j = -1$)	0,72	35,0	0,75	200

Базируясь на этих соображениях для построения модели был выбран дробный факторный план типа 2^{4-1} с генератором $x_4 = x_1x_2$. Контрастом для этого плана является соотношение $I =$

$= x_1x_2x_4$. Оценки смешаны при этом следующим образом:

$$\hat{a}_1 \longrightarrow \bar{a}_1 + \bar{a}_{24}, \quad \hat{a}_{34} \longrightarrow \bar{a}_{34},$$
$$\hat{a}_2 \longrightarrow \bar{a}_2 + \bar{a}_{14}, \quad \hat{a}_{13} \longrightarrow \bar{a}_{13},$$
$$\hat{a}_3 \longrightarrow \bar{a}_3, \quad \hat{a}_{23} \longrightarrow \bar{a}_{23},$$
$$\hat{a}_4 \longrightarrow \bar{a}_4 + \bar{a}_{12},$$

Из приведенных соотношений ясно, что оценки интересующих нас взаимодействий смешаны с другими эффектами.

Таблица 5.6

План и результаты эксперимента (пример 5.6)

Номер опыта	x_0	x_1	x_2	x_3	x_4	x_{13}	x_{23}	x_{34}	Целевая величина $\tilde{y}$ (среднее по двум наблюдениям)
1	+	+	+	+	+	+	+	+	296
2	+	+	−	+	−	+	−	−	122
3	+	−	−	+	+	−	−	+	239
4	+	−	+	+	−	−	+	−	586
5	+	+	+	−	+	−	−	−	232
6	+	+	−	−	−	−	+	+	292
7	+	−	−	−	+	+	+	−	539
8	+	−	+	−	−	+	−	+	383
$\hat{a}_l$	336,12	−100,62	38,12	−25,38	−9,62	−1,12	92,12	−33,62	

Ошибки наблюдений: $s = 20{,}00$; $s_l = 7{,}10$, $\varphi_e = 8$.

План эксперимента и его результаты представлены в табл. 5.6. В нижней части таблицы приведены рассчитанные оценки коэффициентов. Очевидно, что взаимодействия x_2x_3 и x_3x_4 действительно оказывают существенное влияние на целевую величину. Интерполяционная формула имеет вид

$$\hat{y}(x) = 336{,}12 - 100{,}62x_1 + 38{,}12x_2 - 25{,}38x_3 - 9{,}62x_4 - \\ - 1{,}12x_1x_3 + 92{,}12x_2x_3 - 33{,}62x_3x_4.$$

Значимость коэффициентов

Пользуясь (3.102), проверим значимость коэффициентов, принимая $\alpha = 0{,}05$. Значение $t_{кр}$ при восьми степенях свободы и $\alpha = 0{,}05$ составляет $t_{кр} = 2{,}31$ (см. табл. 15.3). При этом

$$t_{кр}s_l = 2{,}31 \times 7{,}10 = 16{,}4.$$

Оценки $\hat{a}_4$ и $\hat{a}_{13}$ оказываются незначимыми, и, следовательно, соответствующие функции необходимо из модели исключить. В результате получаем

$$\hat{y}(x) = 336{,}12 - 100{,}62x_1 + 38{,}12x_2 - 25{,}38x_3 + 92{,}12x_2x_3 - 33{,}62x_3x_4.$$

Проверка значимости квадратичных эффектов

Проверим гипотезу H: $E\{\tilde{y}^0 - \hat{a}_0\} = 0$ для примера 5.6. Имеем

$$\nu = 2,$$
$$s_e^2 = \nu s^2 = 2 \times 400 = 800,$$
$$n_0 = 2,$$
$$N = 8,$$
$$\varphi_e = 8, \quad \hat{a}_0 = 336{,}12.$$

Согласно выражению (5.38), вычисляем $\tilde{y}^0$:

$$\tilde{y}^0 = 350;$$

$t_{кр}$ находим из табл. 15.3 при $\alpha = 0{,}05$ и $\varphi = 8$:

$$t_{кр} = 2{,}31.$$

Проверяем условие (5.39):

$$|\tilde{y}^0 - \hat{a}_0| = |350 - 336{,}12| = 13{,}88 < t_{кр} s \sqrt{\frac{n_0 + \nu N}{n_0 N}} =$$
$$= 2{,}31 \sqrt{\frac{400\,(2+16)}{2 \times 8}} \approx 49.$$

Отсюда следует, что разность $|\tilde{y}^0 - \hat{a}_0|$ незначимо отличается от нуля и квадратичные члены в модель можно не вводить.

5.6. Планы для квадратичных моделей

5.6.1. Вводные замечания

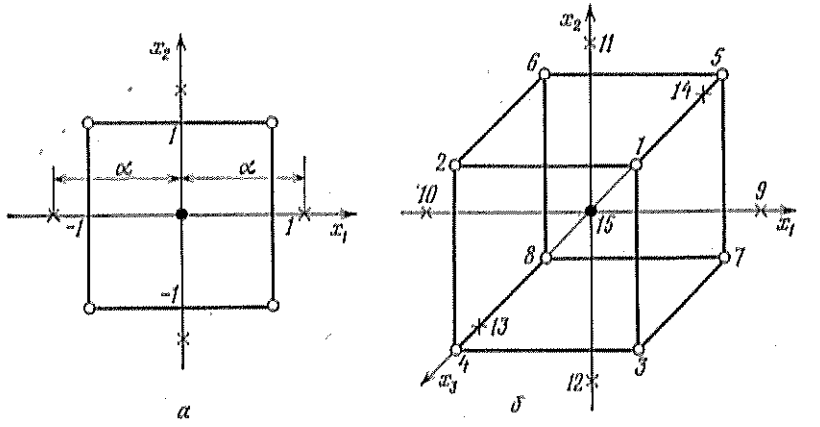
В этом разделе рассматриваются планы для моделей, имеющих вид

$$y(a, x) = a_0 + a_1x_1 + \ldots + a_nx_n + a_{n+1}x_1^2 + \ldots + a_{2n}x_n^2 + a_{2n+1}x_1x_2 + \ldots + a_kx_{n-1}x_n. \qquad (5.45)$$

Общее число неизвестных коэффициентов в модели (5.45) равно

$$(k+1) = \binom{n+2}{2} = \frac{(n+2)(n+1)}{2}. \qquad (5.46)$$

Для оценки коэффициентов модели вида (5.45), содержащей функции независимых переменных типа x_l^3, независимая переменная в плане должна принимать по крайней мере три различных значения. Композиционный план для квадратичных моделей может быть получен путем добавления некоторого количества специальных точек к «ядру», образованному планом для линейной модели. В качестве «ядра» могут быть использованы планы типа 2^n или 2^{n-p}. Если к ядру добавить точку в центре плана с координатами 0, ..., 0 и $2n$ так называемых «звездных»



Фиг. 5.3. Композиционные планы второго порядка для $n=2$ (*а*) и $n=3$ (*б*).

точек с координатами $(\pm\alpha, 0, \ldots, 0), \ldots, (0, \ldots, 0, \pm\alpha)$, то получается центральный композиционный план, предложенный Боксом [14]. На фиг. 5.3 показаны точки композиционных планов для $n=2$ (*а*) и $n=3$ (*б*). Здесь в качестве ядра использованы точки полного факторного эксперимента (обозначены кружками). Крестиками обозначены звездные точки, расположенные на координатных осях на расстоянии α от центра плана.

Матрица планирования X для плана *б* (фиг. 5.3) приведена в табл. 5.7. Номера точек на фиг. 5.3 соответствуют номерам, указанным в табл. 5.7. Выбором величины плеча α композиционного плана и числа n_0 точек в центре могут быть обеспечены различные свойства получаемого плана. Ниже мы опишем ортогональные и ротатабельные центральные композиционные планы.

5.6.2. Ортогональные центральные композиционные планы

При построении этих планов величина α (плечо звездных точек) выбирается так, чтобы обеспечить ортогональность получаемого плана. Число точек в центре плана обычно принимается

Таблица 5.7

Композиционный план второго порядка для $n=3$

l	x_1	x_2	x_3	
1	+1	+1	+1	Полный факторный план (2^n)
2	−1	+1	+1	
3	+1	−1	+1	
4	−1	−1	+1	
5	+1	+1	−1	
6	−1	+1	−1	
7	+1	−1	−1	
8	−1	−1	−1	
9	$+\alpha$	0	0	Звездные точки ($2n$)
10	$-\alpha$	0	0	
11	0	$+\alpha$	0	
12	0	$-\alpha$	0	
13	0	0	$+\alpha$	
14	0	0	$-\alpha$	
15	0	0	0	Центр плана (n_0 точек)

равным единице. Для обеспечения ортогональности оказывается необходимым преобразовать модель (5.45) следующим образом:

$$y(a, x) = b_0 + a_1x_1 + \ldots + a_nx_n + a_{n+1}(x_1^2 - \beta) + \\ + \ldots + a_{2n}(x_n^2 - \beta) + a_{2n+1}x_1x_2 + \ldots + a_kx_{n-1}x_n. \qquad (5.47)$$

Здесь

$$\beta = \frac{\sum_{l=1}^{N}(x_l')^2}{N} = \frac{2^{n-p} + \alpha^2}{N}. \qquad (5.48)$$

В выражении (5.48) N — общее число точек в плане, 2^{n-p} — число точек ядра композиционного плана. От модели (5.47) легко перейти к модели (5.45), определяя a_0 в (5.45) следующим образом:

$$a_0 = b_0 - \beta\sum_{l=1}^{n} a_{n+l}. \qquad (5.49)$$

В общем случае матрица F функций независимых переменных для ортогонального центрального композиционного плана имеет вид, показанный в табл. 5.8. Через x_0 в табл. 5.8 обозначена фиктивная переменная при коэффициенте a_0. Общее число

Таблица 5.8

Матрица F для ортогонального центрального композиционного плана второго порядка

	Номер опыта j	Матрица F										
		x_0	матрица X плана				$x_1^2-\beta$	...	$x_n^2-\beta$	x_1x_2	...	$x_{n-1}x_n$
			x_1	x_2	...	x_n						
Ядро плана	1	1	$+1$	$+1$	...	$+1$	$1-\beta$	...	$1-\beta$	$+1$	...	1
	2	1	-1	$+1$		$+1$	$1-\beta$		$1-\beta$	-1		1
	3	1	$+1$	-1		$+1$	$1-\beta$		$1-\beta$	-1		.
	4	1	-1	-1		$+1$	.		.	$+1$		.
	5	1	$+1$	$+1$		$+1$	.		.			.
	.	.	.	.		.	.		.			.
	.	.	.	.		.	.		.			.
	.	.	.	.		.	.		.			.
	.	.	.	.		.	.		.			.
	2^{n-p}	1	.	.	...	.	$1-\beta$	...	$1-\beta$	...	...	. . .
Звездные точки	$2^{n-p}+1$	1	$+\alpha$	0	...	0	$\alpha^2-\beta$	...	$-\beta$	0	...	0
	$2^{n-p}+2$	1	$-\alpha$	0		0	$\alpha^2-\beta$		$-\beta$	0		0
	.	1	0	$+\alpha$		0	$-\beta$		.	.		
	.	1	0	$-\alpha$		0	$-\beta$		.	.		
	.	.	.	.		.	.		.	.		
	.	.	.	.		.	.		.	.		
	.	.	.	.		.	.		.	.		
	.	.	.	.		.	.		.	.		
	.	1	0	0		$+\alpha$	$-\beta$		$\alpha^2-\beta$	0		0
	$2^{n-p}+2n$	1	0	0	...	$-\alpha$	$-\beta$	...	$\alpha^2-\beta$	0	...	0
Центр плана	$N=2^{n-p}+2n+1$	1	0	0	...	0	$-\beta$	...	$-\beta$	0	...	0

точек плана

$$N = 2^{n-r} + 2n + 1.$$

Рассмотрим вопрос о том, в каких случаях в качестве ядра плана можно использовать дробные факторные планы.

Предположим, что модель имеет вид (5.47) (т. е. модель представлена полным квадратичным полиномом) и требуется получить раздельные (несмешанные) оценки для всех коэффициентов. Тогда необходимо потребовать, чтобы в матрице F размера $N \times \left(\frac{n+2}{2}\right)$ не было одинаковых столбцов. Из табл. 5.8 ясно, что одинаковые столбцы в матрице могут быть только у парных произведений факторов. Таким образом, в качестве ядра в композиционном плане могут быть использованы только такие дробные планы, которые позволяют получить несмешанные друг с другом оценки коэффициентов при парных произведениях факторов. Поскольку выбор величины α не влияет на ортогональность столбцов матрицы F, соответствующих факторам x_i и парным произведениям $x_i x_j$, то для ортогональности композиционного плана мы должны еще дополнительно потребовать, чтобы парные произведения в ядре плана не были смешаны с линейными членами. Нетрудно убедиться в том, что для $n \leqslant 4$ дробных планов, удовлетворяющих указанным требованиям, построить нельзя. Действительно, выбирая, например, при $n = 4$ дробный план с генератором $x_4 = x_1 x_2 x_3$, получаем матрицу композиционного плана, в которой столбцы, соответствующие функциям $x_1 x_4$ и $x_2 x_3$, $x_2 x_4$ и $x_1 x_3$, $x_3 x_4$ и $x_1 x_2$, полностью совпадают. А это означает (см. разд. 5.5.3), что коэффициенты при этих функциях нельзя оценить раздельно. Если же при $n = 4$ мы используем дробный план с генератором $x_4 = x_1 x_2$, то в ядре окажутся смешанными оценки коэффициентов при линейных членах и парных произведениях факторов, и выбором величины α композиционный план не удается сделать ортогональным. Таким образом, при $n \leqslant 4$ в качестве ядра ортогонального центрального композиционного плана может быть использован только полный факторный план. Легко убедиться в том, что для $n = 5$ в качестве ядра можно использовать дробный план 2^{5-1} с генератором $x_5 = x_1 x_2 x_3 x_4$; для $n = 6$ или 7 также возможно применение только планов 2^{n-1} (например, с таким генератором $x_6 = x_7 = x_1 x_2 x_3 x_4 x_5$). Только для $n = 8$ возможно в качестве ядра применить план 2^{8-2}, например, с генераторами $x_7 = x_1 x_2 x_3 x_4$ и $x_8 = x_1 x_2 x_5 x_6$.

Перейдем теперь к выбору плеча ортогональных планов. Из табл. 5.8 легко видеть, что скалярные произведения любых двух столбцов матрицы F равны нулю при любом выборе α. Исключение составляют столбцы при квадратичных функциях вида

Таблица 5.9

Параметры ортогональных центральных композиционных планов

Размерность	Ядро плана	N	α	β	Элементы матрицы C			
					c_0	c_1	c_2	c_3
2	2^2	9	1	0,6667	0,1111	0,1667	0,5	0,25
3	2^3	15	1,215	0,73	0,0667	0,0913	0,2298	0,1250
4	2^4	25	1,414	0,8	0,04	0,05	0,125	0,0625
5	2^{5-1}	27	1,547	0,77	0,03704	0,0481	0,0871	0,0625
6	2^{6-1}	45	1,722	0,843	0,0222	0,0264	0,0564	0,03125
7	2^{7-1}	79	1,885	0,9	0,0127	0,0141	0,0389	0,0156
8	2^{8-2}	81	2,001	0,8889	0,0123	0,0139	0,0312	0,0156

$(x_i^2-\beta)$: их попарные скалярные произведения в общем случае не равны нулю, но зато зависят от α. Приравнивая нулю сумму произведений элементов двух столбцов при функциях вида $(x_i^2-\beta)$, получаем условие для выбора значения α, обеспечивающего ортогональность плана:

$$2^{n-p}\left(1-\frac{2^{n-p}+2\alpha^2}{N}\right)^2-4\frac{2^{n-p}+2\alpha^2}{N}\left(\alpha^2-\frac{2^{n-p}+2\alpha^2}{N}\right)+$$
$$+(2n-3)\left(\frac{2^{n-p}+2\alpha^2}{N}\right)^2=0. \qquad (5.50)$$

Из (5.50) следует

$$\alpha=\sqrt{2^{\frac{n-p}{2}-1}\left(\sqrt{N}-2^{\frac{n-p}{2}}\right)}. \qquad (5.51)$$

Значения α для различных n приведены в табл. 5.9.

Информационная матрица плана имеет вид

$$M=\begin{bmatrix} m_0 & 0 & 0 & 0 \\ 0 & m_1 I_n & 0 & 0 \\ 0 & 0 & m_2 I_n & 0 \\ 0 & 0 & 0 & m_3 I_{\binom{n}{2}} \end{bmatrix}. \qquad (5.52)$$

Здесь

$$m_0=N=2^{n-p}+2n+1, \qquad (5.53)$$
$$m_1=2^{n-p}+2\alpha^2, \qquad (5.54)$$
$$m_2=2^{n-p}(1-\beta)^2+2(\alpha^2-\beta)^2+(2n-1)\beta^2, \qquad (5.55)$$
$$m_3=2^{n-p}, \qquad (5.56)$$

I_n — единичная матрица размера n; $\binom{n}{2}$ — число сочетаний из n по 2.

Отсюда получаем следующее выражение для дисперсионной матрицы плана:

$$C = \begin{bmatrix} c_0 & 0 & 0 & 0 \\ 0 & c_1 I_n & 0 & 0 \\ 0 & 0 & c_2 I_n & 0 \\ 0 & 0 & 0 & c_3 I_{\binom{n}{2}} \end{bmatrix}, \tag{5.57}$$

где

$$c_l = m_l^{-1}. \tag{5.58}$$

Значения элементов дисперсионной матрицы для различных n указаны в табл. 5.9.

Приведем здесь еще формулы для расчета оценок регрессионных коэффициентов:

$$\hat{a}_l = \begin{cases} c_1 \sum\limits_{j=1}^{N} x_i^j \tilde{y}^j, & i = 1, \ldots, n, \\ c_2 \sum\limits_{j=1}^{N} [(x_{i-n}^j)^2 - \beta] \tilde{y}^j, & i = n+1, \ldots, 2n, \\ c_3 \sum\limits_{j=1}^{N} x_\mu^j x_\lambda^j \tilde{y}^j, & \mu, \lambda = 1, 2, \ldots, n, \mu \neq \lambda, i = 2n+1, \ldots, k. \end{cases} \tag{5.59}$$

Оценка $\hat{b}_0$ рассчитывается по формуле

$$\hat{b}_0 = \frac{1}{N} \sum_{j=1}^{N} \tilde{y}^j. \tag{5.60}$$

Для $\hat{a}_0$ в соответствии с (5.49) имеем

$$\hat{a}_0 = \hat{b}_0 - \beta \sum_{i=1}^{n} \hat{a}_{n+i}. \tag{5.61}$$

Оценки дисперсий коэффициентов определяются по формулам

$$s_i^2 = \begin{cases} s^2 c_0, & i = 0, \\ s^2 c_1, & i = 1, \ldots, n, \\ s^2 c_2, & i = n+1, \ldots, 2n, \\ s^2 c_3, & i = 2n+1, \ldots, k. \end{cases} \tag{5.62}$$

Для $\hat{a}_0$ имеем

$$s_{\hat{a}_0}^2 = s^2 (c_0 + n\beta^2 c_2). \tag{5.63}$$

В (5.62) и (5.63) s^2 — оценка дисперсии ошибок наблюдений. С целью получения оценки дисперсии наблюдений опыты во всех точках плана могут проводиться по ν раз (см. разд. 3.3.3 и 3.3.5).

Таблица 5.10

Статистический анализ для ортогональных центральных композиционных планов

n	ν	$1-P$	φ_1	φ_2	h_1	h_2	h_3	$F_{кр}$
2	1	0,1	3	—	0,960	1,663	1,177	
		0,05	3	—	1,299	2,250	1,591	
		0,01	3	—	2,384	4,13	2,921	
	2	0,1	3	9	0,748	1,296	0,917	
		0,05	3	9	0,923	1,599	1,131	3,86
		0,01	3	9	1,327	2,298	1,625	6,99
	3	0,1	3	18	0,708	1,225	0,867	
		0,05	3	18	0,858	1,486	1,051	3,16
		0,01	3	18	1,175	2,035	1,439	5,09
3	1	0,1	5	—	0,609	0,966	0,712	
		0,05	5	—	0,777	1,232	0,909	
		0,01	5	—	1,218	1,933	1,426	
	2	0,1	5	15	0,529	0,840	0,619	
		0,05	5	15	0,644	1,022	0,753	2,90
		0,01	5	15	0,890	1,413	1,042	4,56
	3	0,1	5	30	0,513	0,814	0,600	
		0,05	5	30	0,617	0,979	0,739	2,53
		0,01	5	30	0,831	1,318	0,972	3,70
4	1	0,1	10	—	0,405	0,641	0,453	
		0,05	10	—	0,498	0,788	0,557	
		0,01	10	—	0,709	1,120	0,792	
	2	0,1	10	25	0,382	0,604	0,427	
		0,05	10	25	0,460	0,728	0,515	2,24
		0,01	10	25	0,623	0,985	0,697	3,13
	3	0,1	10	50	0,375	0,593	0,419	
		0,05	10	50	0,449	0,710	0,502	2,03
		0,01	10	50	0,599	0,948	0,673	2,70
5	1	0,1	6	—	0,426	0,573	0,486	
		0,05	6	—	0,537	0,722	0,612	
		0,01	6	—	0,813	1,092	0,928	
	2	0,1	6	27	0,373	0,502	0,426	
		0,05	6	27	0,450	0,606	0,513	2,46
		0,01	6	27	0,608	0,817	0,693	3,56
	3	0,1	6	54	0,367	0,494	0,419	
		0,05	6	54	0,440	0,592	0,501	2,27
		0,01	6	54	0,586	0,770	0,668	3,15
6	1	0,1	17	—	0,283	0,415	0,308	
		0,05	17	—	0,343	0,503	0,373	
		0,01	17	—	0,417	0,691	0,512	

Таблица 5.10 (продолжение)

n	ν	$1-P$	φ_1	φ_2	h_1	h_2	h_3	$F_{кр}$
	2	0,1	17	45	0,273	0,400	0,297	
		0,05	17	45	0,327	0,480	0,356	1,86
		0,01	17	45	0,437	0,642	0,476	2,40
	3	0,1	17	90	0,270	0,396	0,294	
		0,05	17	90	0,323	0,473	0,351	1,74
		0,01	17	90	0,428	0,628	0,465	2,18
7	1	0,1	43	—	0,200	0,332	0,210	
		0,05	43	—	0,239	0,398	0,252	
		0,01	43	—	0,320	0,532	0,337	
	2	0,1	43	79	0,198	0,328	0,208	
		0,05	43	79	0,236	0,393	0,249	1,53
		0,01	43	79	0,315	0,524	0,330	1,83
	3	0,1	43	158	0,196	0,326	0,207	
		0,05	43	158	0,235	0,390	0,247	1,46
		0,01	43	158	0,310	0,514	0,326	1,70
8	1	0,1	36	—	0,199	0,298	0,211	
		0,05	36	—	0,239	0,358	0,254	
		0,01	36	—	0,321	0,481	0,340	
	2	0,1	36	81	0,196	0,294	0,208	
		0,05	36	81	0,235	0,352	0,249	1,56
		0,01	36	81	0,311	0,466	0,330	1,89
	3	0,1	36	162	0,195	0,292	0,207	
		0,05	36	162	0,233	0,349	0,247	1,49
		0,01	36	162	0,307	0,461	0,326	1,75

При проведении статистического анализа результатов с целью проверки адекватности модели и значимости коэффициентов (см. разд. 3.3.7 и 3.3.8) может быть использована табл. 5.10. Здесь приняты следующие обозначения:

n — размерность факторного пространства;

n — число параллельных опытов в каждой точке ортогонального центрального композиционного плана;

$1-P$ — используемый уровень значимости проверки гипотез [1] (адекватность модели, значимость коэффициентов);

φ_1 — число степеней свободы для остаточной дисперсии [см. формулы (3.90) и (3.91)];

---

[1] В табл. 15.2 и 15.3, а также в других разделах книги уровень значимости обозначается через α. Это обозначение использовано и в примерах 5.7 и 5.8. Очевидно, что уровень значимости α не имеет никакого отношения к величине α плеча звездных точек. — *Прим. ред.*

φ_2 — число степеней свободы для оценки дисперсии наблюдений [см. формулы (3.94) и (3.95)];

h_1, h_2, h_3 — величины $t_{кр}\sqrt{c_1}$, $t_{кр}\sqrt{c_2}$ и $t_{кр}\sqrt{c_3}$ соответственно;

$t_{кр}$ — критическое значение распределения Стьюдента при заданных уровне значимости и числе степеней свободы φ_1 (при $\nu = 1$) или φ_2 (при $\nu = 2, 3$).

Величины $h_1, \ldots, h_3$ используются при проверке значимости коэффициентов. При этом вместо (3.102) имеем следующее условие значимости:

$$|\hat{a}_i| > h_i s, \tag{5.64}$$

где s^2 — оценка дисперсии ошибок наблюдений, определяемая по (3.91) при $\nu = 1$ и по (3.95) при $\nu > 1$.

В последней колонке табл. 5.10 приведены значения критерия Фишера для проверки адекватности модели, согласно разд. 3.3.7.

Заметим, что при применении ортогональных планов все коэффициенты оцениваются независимо. Это значит, что изменение оценки любого коэффициента (например, исключение соответствующего члена из уравнения) не приводит к изменению других оценок и их дисперсий.

Пример 5.7

Изучался процесс окисления фосфитов гипохлоритом натрия [13]. В качестве выходной переменной y рассматривался процент окисления фосфитов. Варьируемыми переменными являлись pH анализируемого раствора (x_1), температура (x_2), продолжительность окисления (x_3) и избыток раствора гипохлорита натрия сверх необходимого по стехиометрии. Методом крутого восхождения (разд. 12) были найдены значения факторов, обеспечивающие почти полное окисление фосфитов. Затем в окрестности

Таблица 5.11

Факторы и уровни их варьирования (пример 5.7)

Факторы	Кодовое обозначение	$x_i = -1{,}414$ (звездная точка)	$x_i = -1$ (нижний уровень)	$x_i = 0$ (основной уровень)	$x_i = 1$ (верхний уровень)	$x_i = 1{,}414$ (звездная точка)
pH раствора	x_1	6,29	6,5	7,0	7,5	7,71
Температура, °С	x_2	17,9	20	25	30	32,1
Продолжительность окисления, мин	x_3	1,17	2	4	6	6,83
Избыток гипохлорита натрия, %	x_4	48,6	72,4	129,9	187,4	211,2

Таблица 5.12

План эксперимента, матрица F и результаты наблюдений (пример 5.7)

	Номер опыта	Матрица F														$\tilde{y}$
		x_0	Матрица X плана				$x_1^2-\beta$	$x_2^2-\beta$	$x_3^2-\beta$	$x_4^2-\beta$	x_1x_2	x_1x_3	x_1x_4	x_2x_3	x_2x_4	x_3x_4	
			x_1	x_2	x_3	x_4											
Ядро плана (план 2^n)	1	+1	−1	−1	−1	−1	+0,2	+0,2	+0,2	+0,2	+1	+1	+1	+1	+1	+1	81,42
	2	+1	+1	−1	−1	−1	+0,2	+0,2	+0,2	+0,2	−1	−1	−1	+1	+1	+1	61,31
	3	+1	−1	+1	−1	−1	+0,2	+0,2	+0,2	+0,2	−1	+1	+1	−1	−1	+1	89,21
	4	+1	+1	+1	−1	−1	+0,2	+0,2	+0,2	+0,2	+1	−1	−1	−1	−1	+1	75,31
	5	+1	−1	−1	+1	−1	+0,2	+0,2	+0,2	+0,2	+1	−1	+1	−1	+1	−1	98,35
	6	+1	+1	−1	+1	−1	+0,2	+0,2	+0,2	+0,2	−1	+1	−1	−1	+1	−1	86,09
	7	+1	−1	+1	+1	−1	+0,2	+0,2	+0,2	+0,2	−1	−1	+1	+1	−1	−1	99,83
	8	+1	+1	+1	+1	−1	+0,2	+0,2	+0,2	+0,2	+1	+1	−1	+1	−1	−1	94,65
	9	+1	−1	−1	−1	+1	+0,2	+0,2	+0,2	+0,2	+1	+1	−1	+1	−1	−1	98,99
	10	+1	+1	−1	−1	+1	+0,2	+0,2	+0,2	+0,2	+1	+1	−1	+1	−1	−1	93,71
	11	+1	−1	+1	−1	+1	+0,2	+0,2	+0,2	+0,2	−1	+1	−1	−1	+1	−1	100,15
	12	+1	+1	+1	−1	+1	+0,2	+0,2	+0,2	+0,2	+1	−1	+1	−1	+1	−1	99,07
	13	+1	−1	−1	+1	+1	+0,2	+0,2	+0,2	+0,2	+1	−1	−1	−1	−1	+1	100,88
	14	+1	+1	−1	+1	+1	+0,2	+0,2	+0,2	+0,2	−1	+1	+1	−1	−1	+1	100,31
	15	+1	−1	+1	+1	+1	+0,2	+0,2	+0,2	+0,2	−1	−1	−1	+1	+1	+1	100,67
	16	+1	+1	+1	+1	+1	+0,2	+0,2	+0,2	+0,2	+1	+1	+1	+1	+1	+1	100,10
Звездные точки	17	+1	−1,414	0	0	0	+1,2	−0,8	−0,8	−0,8	0	0	0	0	0	0	99,90
	18	+1	+1,414	0	0	0	+1,2	−0,8	−0,8	−0,8	0	0	0	0	0	0	87,80
	19	+1	0	−1,414	0	0	−0,8	+1,2	−0,8	−0,8	0	0	0	0	0	0	98,98
	20	+1	0	+1,414	0	0	−0,8	+1,2	−0,8	−0,8	0	0	0	0	0	0	100,25
	21	+1	0	0	−1,414	0	−0,8	−0,8	+1,2	−0,8	0	0	0	0	0	0	89,69
	22	+1	0	0	+1,414	0	−0,8	−0,8	+1,2	−0,8	0	0	0	0	0	0	100,20
	23	+1	0	0	0	−1,414	−0,8	−0,8	−0,8	+1,2	0	0	0	0	0	0	87,48
	24	+1	0	0	0	+1,414	−0,8	−0,8	−0,8	+1,2	0	0	0	0	0	0	100,83
Центр плана	25	+1	0	0	0	0	−0,8	−0,8	−0,8	−0,8	0	0	0	0	0	0	99,94

этой точки (основной уровень факторов) был спланирован и реализован эксперимент с целью получения математической модели процесса в виде полинома второго порядка. Использовался ортогональный центральный композиционный план второго порядка. Уровни варьирования факторов указаны в табл. 5.11. Матрица плана приведена в табл. 5.12. В табл. 5.12, кроме того, представлены результаты опытов и приведены элементы матрицы F значений функций независимых переменных. Значения $\alpha = 1{,}414$ и $\beta = 0{,}8$ получены из табл. 5,9 при $n = 4$. В соответствии с табл. 5.9 для элементов дисперсионной матрицы плана имеем

$$c_0 = 0{,}04,$$
$$c_1 = 0{,}05,$$
$$c_2 = 0{,}125,$$
$$c_3 = 0{,}0625.$$

Пользуясь выражениями (5.59) и (5.61), получаем модель процесса в следующем виде:

$$\hat{y}(x) = 99{,}18 - 3{,}80x_1 + 1{,}98x_2 + 4{,}83x_3 + 6{,}30x_4 - 2{,}58x_1^2 + 0{,}30x_2^2 - 2{,}03x_3^2 - 2{,}43x_4^2 + 1{,}09x_1x_2 + 1{,}36x_1x_3 + 2{,}75x_1x_4 - 1{,}17x_2x_3 - 1{,}61x_2x_4 - 3{,}85x_3x_4. \quad (5.65)$$

В (5.65) через x_i обозначены кодированные значения переменных.

Проверим адекватность модели. При этом воспользуемся оценкой дисперсии ошибок наблюдений $s^2 = 1{,}19$ с числом степеней свободы $\varphi_2 = 3$, полученной путем параллельных наблюдений, проведенных в некоторой точке вне рамок данного плана. Остаточная сумма квадратов для полученной модели составляет $S_R = 52{,}2$ при числе степеней свободы $\varphi_1 = 10$. Эта величина в данном случае характеризует степень неадекватности модели. Поскольку известна оценка s^2 дисперсии ошибок наблюдений, не связанная с предположением об адекватности модели, а в случае адекватности $E\{s_R^2\} = E\{S_R/\varphi_1\} = E\{s^2\} = \sigma^2$, то проверку адекватности в данном случае можно осуществить путем проверки гипотезы о равенстве дисперсий, соответствующих оценкам s_R^2 с числом степеней свободы $\varphi_1 = 10$ и s^2 с числом степеней свободы $\varphi_2 = 3$. Согласно разд. 2.4, находим

$$F = \frac{s_R^2}{s^2} = \frac{5{,}22}{1{,}19} = 4{,}38. \quad (5.66)$$

Приняв уровень значимости $1 - P = \alpha = 0{,}05$, найдем с помощью табл. 15.4 при $\varphi_1 = 10$ и $\varphi_2 = 3$ $F_{кр} = 8{,}79$. Так как

$$F = 4{,}38 < F_{кр} = 8{,}79, \quad (5.67)$$

то гипотеза о равенстве дисперсии принимается и модель считается адекватной.

Проверим теперь значимость коэффициентов. При этом используем в качестве оценки s^2 дисперсии ошибок наблюдений остаточную дисперсию $s^2 = s_R^2 = 5{,}22$. Из табл. 5.10 при $n = 4$, $\nu = 1$ и уровне значимости $1 - P = \alpha = 0{,}05$ находим

$$h_1 = 0{,}498,$$
$$h_2 = 0{,}788,$$
$$h_3 = 0{,}557.$$

Сравнение произведения sh_1 с абсолютными значениями оценок коэффициентов при линейных членах, sh_2 — с оценками коэффициентов при квадратах факторов и sh_3 — с оценками при парных взаимодействиях, показывает, что условие (5.64) не выполняется для коэффициентов при следующих функциях: x_2^2, x_1x_3, x_2x_3. Исключая эти функции из уравнения (5.65), получим

$$\hat{y}(x) = 99{,}42 - 3{,}80x_1 + 1{,}98x_2 + 4{,}83x_3 + 6{,}30x_4 - 2{,}58x_1^2 - 2{,}03x_3^2 - 2{,}43x_4^2 + 1{,}36x_1x_2 + 2{,}75x_1x_4 - 1{,}61x_2x_4 - 3{,}85x_3x_4. \quad (5.68)$$

Обратим внимание на то, что все оставшиеся в (5.68) коэффициенты (кроме $\hat{a}_0$) совпадают с их значениями в уравнении (5.65). Этот факт — результат ортогональности плана. Оценка $\hat{a}_0$ связана в соответствии с (5.61) с оценками коэффициентов при квадратах переменных, и поэтому исключение из уравнения функции x_3^2 привело к изменению $\hat{a}_0$.

5.6.3. Ротатабельные центральные композиционные планы

В этом случае в качестве условия для выбора величины плеча звездных точек α используется требование ротатабельности плана.

Полагаем, что модель имеет вид (5.45). Ротатабельный центральный композиционный план [3, 4] строится аналогично ортогональному плану. Так же как и в случае ортогонального плана, в качестве ядра плана может быть использован полный (2^n) или дробный (2^{n-p}) факторный план. Величина плеча α для ротатабельного плана второго порядка вычисляется по формуле

$$\alpha = 2^{\frac{n-p}{4}}. \quad (5.69)$$

Число опытов n_0 в центре плана выбирается из следующих соображений. Выдвигается требование, чтобы информация о значении выходной переменной оставалась неизменной (или почти неизменной) для точек внутри сферы единичного радиуса с центром в центре плана. Иными словами, требуется, чтобы информационный профиль ротатабельного плана мало изменялся при значениях радиуса сферы от 0 до 1. Планы, удовлетворяющие этому условию, называются ротатабельными униформ-планами [3, 8].

Таблица 5.13

Параметры ротатабельных центральных композиционных планов

Размерность	Ядро плана	Число звездных точек	n_0	N	α
2	2^2	4	5	13	1,414
3	2^3	6	6	20	1,682
4	2^4	8	7	31	2,000
5	2^5	10	10	52	2,378
	2^{5-1}	10	6	32	2,000
6	2^6	12	15	91	2,828
	2^{6-1}	12	9	53	2,378
7	2^7	14	21	163	3,333
	2^{7-1}	14	14	92	2,828

Таблица 5.14

Ротатабельный центральный композиционный план и матрица F для $n = 3$

	Матрица F									
	x_0	матрица X			x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3
		x_1	x_2	x_3						
Ядро плана	1	−1	−1	−1	1	1	1	1	+1	+1
	1	+1	−1	−1	1	1	1	−1	−1	+1
	1	−1	+1	−1	1	1	1	−1	+1	−1
	1	+1	+1	−1	1	1	1	1	−1	−1
	1	−1	−1	+1	1	1	1	1	−1	−1
	1	+1	−1	+1	1	1	1	−1	+1	−1
	1	−1	+1	+1	1	1	1	−1	−1	+1
	1	+1	+1	+1	1	1	1	1	+1	+1
Звездные точки	1	+1,682	0	0	2,828	0	0	0	0	0
	1	−1,682	0	0	2,828	0	0	0	0	0
	1	0	+1,682	0	0	2,828	0	0	0	0
	1	0	−1,682	0	0	2,828	0	0	0	0
	1	0	0	+1,682	0	0	2,828	0	0	0
	1	0	0	−1,682	0	0	2,828	0	0	0
Центр плана	1	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0

Оказывается, что униформ-план можно получить, меняя число точек в центре ротатабельного плана. В табл. 5.13 приведены значения плеча α и числа n_0 точек в центре плана, звездных точек и общего числа точек N для ротатабельных униформ-планов второго порядка [9].

В табл. 5.14 приведены матрица X ротатабельного центрального композиционного униформ-плана для $n=3$ и матрица F функций независимых переменных. Обратим внимание на то, что в матрице F центральная точка фигурирует n_0 раз. Это значит, что при вычислении оценок коэффициентов мы будем использовать результат каждого параллельного измерения в центре плана, а не их среднее значение. Такое построение матрицы F вытекает из требования одинаковой величины дисперсии всех величин зависимой переменной, входящих в матрицу $\hat{Y}$ (см. разд. 3.2.2). Заметим, что параллельные опыты в центре плана позволяют рассчитать оценку дисперсии ошибок наблюдений.

В общем случае, при наличии n_0 точек в центре плана и повторении эксперимента ν раз в каждой точке матрицы X, оценка дисперсии единичного эксперимента s_e^2 определяется по формуле

$$s_e^2 = \frac{S_e}{\varphi_e}, \tag{5.70}$$

где

$$S_e = \sum_{j=1}^{2^{n-p}+2n} \sum_{l=1}^{\nu} (\tilde{y}^j - \tilde{y}^{jl})^2 + \sum_{j=1}^{n_0} \sum_{l=1}^{\nu} (\tilde{y}^0 - \tilde{y}^{2^{n-p}+2n+l,\,l})^2. \tag{5.71}$$

Здесь

$$\tilde{y}^0 = \frac{1}{n_0 \nu} \sum_{j=1}^{n_0} \sum_{l=1}^{\nu} \tilde{y}^{2^{n-p}+2+j,\,l}, \tag{5.72}$$

$$\varphi_e = (2^{n-p}+2n)(\nu-1) + \nu n_0 - 1. \tag{5.73}$$

Нетрудно убедиться, что информационная матрица $M = F'F$ рассматриваемых ротатабельных планов близка к диагональной. Действительно, матрица M имеет вид

$$M = \begin{bmatrix} N & 0 & (2^{n-p}+2\alpha^2) l_n' & 0 \\ 0 & (2^{n-p}+2\alpha^2) I_n & 0 & 0 \\ (2^{n-p}+2\alpha^2) l_n & 0 & B_{nn} & 0 \\ 0 & 0 & 0 & 2^{n-p} I_{\binom{n}{u}} \end{bmatrix}. \tag{5.74}$$

Здесь l_n — вектор-столбец размерности n с элементами 1;

$$B_{nn} = \begin{bmatrix} 2^{n-p} & 2\alpha^4 & 2^{n-p} & \dots & 2^{n-p} \\ 2^{n-p} & & 2^{n-p}+2\alpha^4 & \dots & 2^{n-p} \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ 2^{n-p} & & 2^{n-p} & \dots & 2^{n-p}+2\alpha^4 \end{bmatrix}. \qquad (5.75)$$

Из (5.74) и (5.75) следует, что оценки коэффициентов при линейных членах и парных взаимодействиях при использовании ротатабельного плана некоррелированы с оценками остальных коэффициентов, а оценки коэффициентов при квадратичных членах коррелированы между собой и оценкой свободного члена. Этот факт позволяет при статистическом анализе результатов расчета делать выводы о значимости коэффициентов при линейных членах и взаимодействиях независимо от значений остальных коэффициентов. Кроме того, исходя из информационной матрицы (5.74), можно получить расчетные формулы для оценок коэффициентов и их дисперсий, удобные для ручного счета [10]:

$$\hat{a}_i = \begin{cases} \frac{A}{N}\left[2\lambda_1(n+2)(0\tilde{y}) - 2\lambda_2\lambda_1\sum_{j=1}^{n}(jj\tilde{y})\right], & i=0, \\ \frac{\lambda_2}{N}(i\tilde{y}), \quad i=1, 2, \dots, n, \end{cases} \qquad (5.76)$$

$$\tilde{a}_{n+1} = \frac{A}{N}\{\lambda_2^2[(n+2)\lambda_1 - n](ii\tilde{y}) + \lambda_2^2(1-\lambda_1)\sum_{j=1}^{n}(jj\tilde{y}) - 2\lambda_1\lambda_2(0\tilde{y})\},$$

$$i=1, 2, \dots, n,$$

$$\tilde{a}_{2n+j} = \frac{\lambda_2^2}{N\lambda_1}(il\tilde{y}), \quad j=1, 2, \dots, k-2n \text{ (коэффициент при } x_ix_l).$$

Здесь приняты обозначения

$$\lambda_1 = \frac{2^{n-p}N}{(2^{n-p}+2\alpha^2)^2}, \qquad (5.77)$$

$$\lambda_2 = \frac{N}{2^{n-p}+2\alpha^2}, \qquad (5.78)$$

$$A = \frac{1}{2\lambda_1[(n+2)\lambda_1 - n]}. \qquad (5.79)$$

Кроме того, введены обозначения

$$(0\tilde{y}) = \sum_{l=1}^{N} \tilde{y}^l,$$
$$(i\tilde{y}) = \sum_{l=1}^{N} x_i^l \tilde{y}^l, \qquad (5.80)$$
$$(il\tilde{y}) = \sum_{l=1}^{N} x_i^l x_l^l \tilde{y}^l.$$

Соответственно для дисперсий имеем следующие выражения:

$$s_i^2 = \begin{cases} 2\frac{\Lambda}{N}\lambda_1^2 (n+2)\, s^2, & i = 0, \\ \frac{\lambda_2^2}{N} s^2, & i = 1, 2, \ldots, n, \\ \frac{\Lambda}{N}\left[(n+1)\lambda_1 - (n-1)\right]\lambda_2^2 s^2, & i = n+1, \ldots, 2n, \\ \frac{\lambda_2^2}{N\lambda_1} s^2, & i = 2n+1, \ldots, k. \end{cases} \qquad (5.81)$$

В (5.81) s^2 — оценка дисперсии значений $\tilde{y}$ (дисперсии наблюдений) с числом степеней свободы $\varphi = (2^{n-p}+2n)(\nu-1)+\nu n_0-1$, равная

$$s^2 = \frac{s_e^2}{\nu}. \qquad (5.82)$$

Рассмотрим теперь методику проверки адекватности модели для случая, когда используется ротатабельный план [11]. Сумма квадратов S_D, характеризующая неадекватность модели, в данном случае определяется выражением

$$S_D = \nu n_0 (\tilde{y}^0 - \hat{y}^0)^2 + \nu \sum_{l=1}^{2^{n-p}+2n} (\tilde{y}^l - \hat{y}^l)^2. \qquad (5.83)$$

Здесь $\hat{y}^l$ — значения выходной переменной в точке плана x^l, рассчитанные с помощью полученного уравнения [см. (3.87)]; $\tilde{y}^0$ — среднее значение зависимой переменной в центре плана (5.72), ν — число параллельных опытов в точках плана X. С суммой (5.83) связано число степеней свободы

$$\varphi_1 = N - (1+k) - (n_0 - 1) = N - k - n_0. \qquad (5.84)$$

Сумма квадратов S_e рассчитывается по формуле (5.71) с числом степеней свободы (5.73). Модель считается адекватной при выбранном уровне значимости $1-P$, если

$$F = \frac{S_D/\varphi_1}{S_e/\varphi_2} < F_{кр}, \qquad (5.85)$$

где $F_{кр}$ — критическое значение распределения Фишера при уровне значимости $1-P$ и числах степеней свободы φ_1 и φ_2 (см. табл. 15.4 и 15.5).

Пример 5.8

Исследовалось влияние температуры закалки (x_1), температуры старения (x_2) и времени старения (x_3) на время до разрушения одного из жаропрочных никелевых сплавов при темпе-

Таблица 5.15

Уровни варьирования факторов (пример 5.8)

Факторы	Кодовое обозначение	$x_j=-1{,}682$ (звездные точки)	$x_j=-1$ (нижний уровень)	$x_j=0$ (основной уровень)	$x_j=+1$ (верхний уровень)	$x_j=+1{,}682$ (звездные точки)
Температура закалки, °С	x_1	1016	1050	1100	1150	1184
Температура старения, °С	x_2	666	700	750	800	834
Время старения, ч.	x_3	0,6	2	4	6	7,4

ратуре 850°С и напряжении 50 кг/мм² [16]. Уровни варьирования факторов указаны в табл. 5.15. В ходе исследования сначала был реализован полный факторный план 2^3 (см. табл. 5.16) и проведено n_0 параллельных опытов в центре плана. Результаты этих экспериментов позволили убедиться в значимости квадратичных эффектов и, следовательно, в необходимости использования модели в виде полинома второго порядка. Затем план был дополнен звездными точками до ротатабельного плана второго порядка.

Рассмотрим процедуру проверки значимости квадратичных эффектов по результатам, полученным при реализации плана 2^3 и опытов в центральной точке. Используя результаты наблюдений в точках плана 2^3 (см. табл. 5.16), находим оценку

$$\hat{a}_0=\frac{1}{8}\sum_{j=1}^{8}\tilde{y}^j=19{,}525. \qquad (5.86)$$

Среднее наблюдений в центре плана

$$\tilde{y}^0=\frac{1}{6}\sum_{j=15}^{20}\tilde{y}^j=29{,}0. \qquad (5.87)$$

Таблица 5.16

План и результаты эксперимента (пример 5.8)

	Номер опыта	Матрица плана X			$\tilde{y}$
		x_1	x_2	x_0	
Ядра плана (план 2^3)	1	−1	−1	+1	8,0
	2	+1	−1	+1	27,9
	3	−1	+1	+1	5,8
	4	+1	+1	+1	43,0
	5	−1	−1	+1	14,1
	6	+1	−1	+1	7,0
	7	−1	+1	+1	20,2
	8	+1	+1	+1	30,2
Звездные точки	9	−1,682	0	0	12,1
	10	+1,682	0	0	36,0
	11	0	−1,682	0	10,4
	12	0	+1,682	0	25,3
	13	0	0	−1,682	20,0
	14	0	0	+1,682	18,0
Центр плана	15	0	0	0	29,0
	16	0	0	0	28,4
	17	0	0	0	28,6
	18	0	0	0	28,8
	19	0	0	0	28,7
	20	0	0	0	30,5

Оценка $s^2 = \nu s_\theta^2 = s_\theta^2$ дисперсии ошибок наблюдений, рассчитанная по результатам параллельных опытов в центре плана, равна

$$s^2 = \frac{1}{5}\sum_{j=16}^{20}(\tilde{y}^j - \tilde{y}^0)^2 = 0{,}58. \qquad (5.88)$$

Число степеней свободы оценки s^2 равно $\varphi_e = 5$. При уровне значимости $1 - P = \alpha = 0{,}05$ и $\varphi_e = 5$ из табл. 15.3 находим $t_{кр} = 2{,}571$. Подставляя найденные значения в (5.39) при $\nu = 1$, $n_0 = 6$ и $N = 20$, убеждаемся в том, что

$$|\tilde{y}^0 - \hat{a}_0| = |29{,}0 - 19{,}525| = 9{,}475 >$$

$$> t_{кр} s \sqrt{\frac{n_0 + \nu N}{n_0 N}} = 2{,}571\sqrt{0{,}58}\sqrt{\frac{26}{120}} = 0{,}88. \qquad (5.89)$$

Таким образом, эффекты при квадратичных членах значимы, и, следовательно, в модель следует включить квадраты факторов.

Рассчитаем теперь оценки коэффициентов модели второго порядка по результатам наблюдений во всех точках плана табл. 5.16. Отметим, что плечо звездных точек и число n_0 опытов в центре плана соответствуют табл. 5.13. С помощью формул (5.77)—(5.79) находим

$$\lambda_1 = \frac{8\times 20}{(8+2\times 2{,}828)^2} = 0{,}856,$$

$$\lambda_2 = \frac{20}{8+2\times 2{,}828} = 1{,}46,$$

$$A = \frac{1}{2\times 0{,}856\,(5\times 0{,}856-3)} = 0{,}455.$$

Далее, используя формулы (5.76), рассчитываем оценки коэффициентов. В результате получаем следующую модель:

$$\hat{y}(x) = 28{,}98 + 7{,}34x_1 + 4{,}93x_2 - 1{,}21x_3 - 1{,}79x_1^2 - 3{,}99x_2^2 - 3{,}58x_3^2 + 4{,}30x_1x_2 - 6{,}78x_1x_3 + 2{,}05x_2x_3. \quad (5.90)$$

Проверим адекватность модели (5.90). Пользуясь (5.71) и (5.73), рассчитываем величины S_e и φ_2 при $\nu = 1$:

$$S_e = 2{,}9, \quad \varphi_2 = 5.$$

Сумму квадратов S_D определяем согласно (5.83):

$$S_D = 6{,}82.$$

Число степеней свободы для S_D в соответствии с (5.84) равно $\varphi_1 = 20 - 9 - 6 = 5$. Отсюда получаем

$$F = \frac{S_D/\varphi_1}{S_e/\varphi_2} = \frac{1{,}365}{0{,}58} = 2{,}44.$$

При уровне значимости $1 - P = \alpha = 0{,}05$ и при $\varphi_1 = 5$ и $\varphi_2 = 5$ в табл. 15.4 находим

$$F_{кр} = 5{,}05.$$

Так как $F = 2{,}44 < F_{кр} = 0{,}05$, то модель считается адекватной.

Для проверки значимости коэффициентов рассчитаем оценки дисперсий коэффициентов по формулам (5.81). При этом мы воспользуемся оценкой $s^2 = 0{,}58$:

$$s_0^2 = 0{,}0965,$$

$$s_1^2 = s_2^2 = s_3^2 = 0{,}0426,$$

$$s_4^2 = s_5^2 = s_6^2 = 0{,}0403,$$

$$s_7^2 = s_8^2 = s_9^2 = 0{,}0725.$$

При уровне значимости $\alpha = 1 - P = 0{,}01$ и $\varphi_2 = 5$ получаем $t_{кр} = 4{,}032$. Нетрудно убедиться, что для всех коэффициентов

$$|\hat{a}_l| > t_{кр} s_l.$$

Таким образом, с 99%-ной надежностью все коэффициенты значимы.

ЛИТЕРАТУРА

1. Налимов В. В., Теория эксперимента, изд-во „Наука", М., 1971.
2. Новые идеи в планировании эксперимента, под ред. В. В. Налимова, изд-во „Наука", М., 1969.
3. Налимов В. В., Чернова Н. А., Статистические методы планирования экстремальных экспериментов, изд-во „Наука", М., 1965.
4. Box G.E.P., Hunter J. S., Multifactor Experimental Designs for Exploring Response Surfaces, *Ann. Math. Stat.*, 28, № 1, 195 (1957).
5. Федоров В. В., Теория оптимального эксперимента, изд-во „Наука", М., 1971.
6. Kiefer J., Wolfowitz J., The Equivalence to Two Extremum Problems, *Canad. J. Math.*, 12, 363 (1960).
7. Box G.E.P., Draper N.R., The Choice of a Second Order Rotatable Design, *Biometrika*, 50, № 3/4, 335 (1963).
8. Cochran W. G., Cox G. M., Experimental Designs, Wiley, New York, 2-nd ed.; Chapman a. Hall, London, 1957.
9. Experimental Designs in Industry (ed. Chew), Wiley, New York; Chapman a. Hall, London, 1958.
10. Никифорова Е. С., Пеледов Г. В., Метод ротатабельного центрального композиционного планирования, см. сб. „Планирование эксперимента", под ред. Э. К. Лецкого и Н. И. Челнокова, Труды МЭИ, вып. 76, Изд-во МЭИ, М., 1970.
11. Горский В. Г., Адлер Ю. П., О методологии регрессионного и дисперсионного анализа при планировании эксперимента с неравномерным дублированием опытов, *Заводск. лаб.*, № 3 (1971).
12. Адлер Ю. П., Введение в планирование эксперимента, изд-во „Металлургия", М., 1969.
13. Масалович В. М., Агасян П. К., Масалович Н. С., Николаева Е. Р., Применение ортогонального планирования для изучения реакции окисления фосфитов гипохлоритом натрия, *Заводск. лаб.*, № 3 (1969).
14. Box G. E. P., Wilson K. B., On the Experimental Attainment of Optimal Conditions, *J. Roy. Stat. Soc.*, Ser. B, 13, № 1 (1951).
15. Налимов В. В., Применение математической статистики при анализе вещества, Физматгиз, М., 1960.
16. Новик Ф. С., Математические методы планирования эксперимента в металловедении. Раздел II. Планы 2-го порядка, МИСиС, М., 1970.

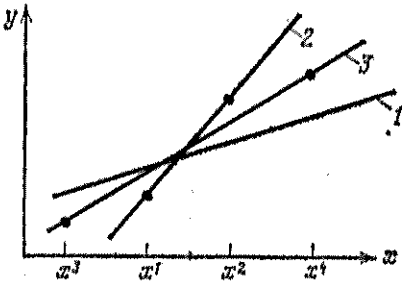
ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Налимов В. В., Голикова Т. И, Логические основания планирования эксперимента, Препринт № 2, Изд-во МГУ, М., 1971.
2. Адлер Ю. П., Маркова Е. В., Грановский Ю. В., Планирование эксперимента при поиске оптимальных условий, изд-во „Наука", М., 1971.
3. Scheffler E., Einführung in die Praxis der statistischen Versuchsplannung, VEB Deutscher Verlag für Grundstoffindustrie, Leipzig.
4. Планирование эксперимента, под ред. Г. К. Круга, изд-во „Наука", М., 1966.
5. Проблемы планирования эксперимента, под ред. Г. К. Круга, изд-во „Наука", М., 1969.
6. Бродский В. З., Многофакторные регулярные планы, Изд-во МГУ, М., 1972.
7. Bandemer H., Bellmann A., Jung W., Richter K., Optimale Versuchsplannung, Akademie-Verlag, Berlin, 1973.
8. Филаретов Г. Ф., Необходимое условие разрешимости задачи построения плана дробного факторного эксперимента, см. сб. „Планирование и автоматизация эксперимента в научных исследованиях", изд-во „Советское радио", М., 1974.

6.
D-ОПТИМАЛЬНЫЕ ПЛАНЫ

6.1. Введение

В этой главе мы вновь будем полагать, что вид модели задан, а обработка экспериментальных данных осуществляется с помощью метода наименьших квадратов. Основное внимание мы уделим здесь вопросам выбора наилучшего плана эксперимента, причем качество плана мы будем теперь рассматривать



Фиг. 6.1. Влияние выбора условий эксперимента на точность модели.
1 — истинная зависимость; *2* и *3* — зависимости, построенные по результатам эксперимента.

с точки зрения достигаемой точности модели (см. разд. 5.3). Подход к планированию эксперимента с точки зрения точности получаемых моделей лежит в основе математической теории планирования эксперимента, современное развитие которой связано с основополагающими работами Кифера [1—3]. В последние годы этот подход привлек интерес широких кругов экспериментаторов и математиков, работающих в области методологии экспериментальных исследований [4—7].

Суть используемого в данном разделе критерия D-оптимальности состоит в требовании выбора таких планов, которые обеспечивают минимальный объем эллипсоида рассеяния оценок коэффициентов. Объем этого эллипсоида характеризует рассеяние случайного вектора относительно его математического ожидания. Таким образом, D-оптимальный план позволяет получить оценки, имеющие минимально возможное рассеяние относительно центра распределения.

С помощью простого примера легко показать, что точность оценок коэффициентов, получаемых из экспериментальных данных, существенно зависит от расположения точек в факторном пространстве. Действительно, пусть ищется оценка коэффициента, связывающего переменные x и y (фиг. 6.1). Истинная зависимость представлена здесь прямой *1*. Пусть в точках x^1 и x^2 поставлены опыты, по результатам которых построена прямая *2*. Если бы те же экспериментальные ошибки имели место при постановке двух опытов в точках x^3 и x^4, то результат получился бы в виде прямой *3*. Сразу видно, что прямая *3* значительно ближе к истинной зависимости, хотя ошибки эксперимента остались теми же. Эффект повышения точности в данном случае явился результатом разнесения экспериментальных точек, или, в общем случае, результатом изменения условий проведения экспериментов.

Естественно, что в практических задачах области варьирования независимых переменных некоторым образом ограничены. Пусть Ω_x — область возможных значений варьируемых переменных. Тогда задача планирования заключается в том, чтобы наилучшим образом расположить экспериментальные точки в области Ω_x. При этом результат решения будет зависеть от вида этой области. Таким образом, при постановке задачи планирования эксперимента мы должны определить область варьирования переменных Ω_x. Естественно, что найти заранее планы для всех возможных областей Ω_x практически невозможно. Поэтому предметом изучения теории являются планы для областей специального вида (например, для гиперкуба), а также способы нахождения планов для областей произвольного вида.

Использование в планировании эксперимента критерия D-оптимальности позволяет охватить единым подходом широкий круг экспериментальных задач, таких, например, как задачи построения нелинейных моделей (оцениваемые параметры нелинейно связаны с выходной переменной) [5, 8], задачи с произвольной областью варьирования независимых переменных, случай неодинаковой точности опытов при различных условиях проведения эксперимента [5], последовательное планирование [8, 9] (см. также разд. 11).

6.2. Критерий D-оптимальности

Займемся теперь задачей построения по экспериментальным данным математической модели некоторого объекта в постановке, описанной в разд. 3.2 и 5.1. Полагаем, что вид модели известен:

$$y(a, x) = a^T f(x). \tag{6.1}$$

Задача состоит в том, чтобы на основе данных, полученных экспериментальным путем, найти оценки вектора коэффициентов модели (6.1). Результат l-го опыта $\tilde{y}^l$ есть

$$\tilde{y}^l = E\{\tilde{y}^l\} + \tilde{e}^l = a'f(x^l) + \tilde{e}^l, \tag{6.2}$$

где e—случайная ошибка, удовлетворяющая условиям (3.25), (3.27) и (3.29) (см. также разд. 5.1).

План эксперимента мы будем обозначать через X, матрицу значений функций независимых переменных в точках плана — через F [см. (3.78)], а вектор измеренных значений выходной переменной—через Y [см. (3.81)]. Применение метода наименьших квадратов приводит к следующему выражению для оценок коэффициентов:

$$\hat{a} = (F'F)^{-1}F'\tilde{Y} = CF'\tilde{Y}. \tag{6.3}$$

Ковариационная матрица оценок коэффициентов (см. разд. 3.2) оказывается равной

$$\operatorname{cov}\{\hat{a}\} = C\sigma^2. \tag{6.4}$$

Здесь C—дисперсионная матрица плана X. Матрица M, обратная матрице C: $M = C^{-1}$, называется информационной матрицей плана X (см. также разд. 5.2). Диагональные элементы матрицы C пропорциональны дисперсиям оценок коэффициентов [см. (3.92)].

Дисперсионная матрица C (а следовательно, и информационная матрица M) зависит от плана эксперимента X, т. е. от положения N опытов в области варьирования независимых переменных.

Критерий D-оптимальности требует такого выбора плана X, содержащего N опытов, при котором определитель дисперсионной матрицы имеет минимальную величину. Естественно, что точки постановки опытов должны выбираться из условия принадлежности их к заданной области Ω_x, где возможны эксперименты. Это требование мы будем обозначать так:

$$X \in \Omega_x, \text{ или } x^l \in \Omega_x, \quad l = 1, 2, \ldots, N. \tag{6.5}$$

Заметим, что требование минимальной величины определителя дисперсионной матрицы C эквивалентно требованию максимальной величины определителя информационной матрицы M. Таким образом, критерий D-оптимальности планирования эксперимента, состоящего из заданного числа N опытов, может быть определен с помощью следующих эквивалентных условий:

$$|C(X^*)| = \min_{X \in \Omega_x} |C(X)| = \min_{X \in \Omega_x} |(F'F)^{-1}|, \tag{6.6а}$$

$$|M(X^*)| = \max_{X \in \Omega_x} |M(X)| = \max_{X \in \Omega_x} |F'F|. \tag{6.6б}$$

Здесь X^* — оптимальный план в смысле критерия D-оптимальности (или D-оптимальный план).

Поясним смысл критерия D-оптимальности. Для этой цели воспользуемся понятием эллипсоида рассеяния случайного вектора. Для случайного вектора a размерности $k+1$, ковариационная матрица которого есть $\operatorname{cov}\{a\}$, уравнение эллипсоида рассеяния имеет вид

$$(a - E\{a\})' \operatorname{cov}^{-1}\{a\} (a - E\{a\}) = k+3. \qquad (6.7)$$

Выражение (6.7) представляет эллипсоид в $(k+1)$-мерном пространстве с центром в точке $E\{a\}$. Этот эллипсоид имеет такие размеры, что ковариационная матрица случайного вектора, равномерно распределенного на нем, совпадает с матрицей $\operatorname{cov}\{a\}$. Поэтому чем больше рассеяние вектора относительно центра распределения, тем большие размеры имеет эллипсоид рассеяния. Таким образом, объем эллипсоида рассеяния является мерой концентрации случайного вектора a около его математического ожидания. В свою очередь объем V_{k+1} эллипсоида рассеяния $(k+1)$-мерного случайного вектора a равен

$$V_{k+1} = \frac{(k+3)^{\frac{k+1}{2}} \pi^{\frac{k+1}{2}}}{\Gamma\left(\frac{k+1}{2}+1\right)} \sqrt{|\operatorname{cov}\{a\}|}, \qquad (6.8а)$$

где $\Gamma[(k+1)/2+1]$ — гамма-функция.

Запишем теперь объем эллипсоида рассеяния для вектора $\hat{a}$ оценок коэффициентов уравнения регрессии. Так как ковариационная матрица вектора $\hat{a}$ определяется выражением (6.4), то, подставляя (6.4) в (6.8а), получим

$$V_{k+1}(\hat{a}) \sim \sqrt{|(F'F)^{-1}|} = \sqrt{|C|}. \qquad (6.8б)$$

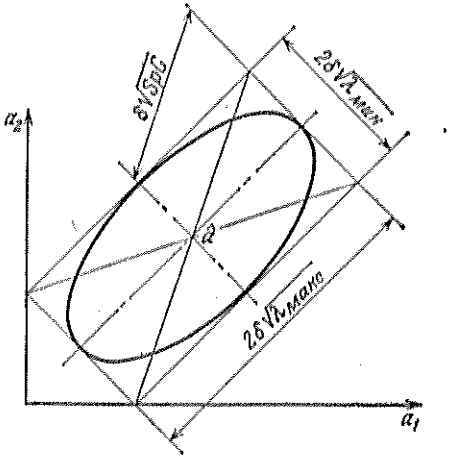
Таким образом, объем эллипсоида рассеяния вектора $\hat{a}$ оценок коэффициентов уравнения регрессии пропорционален корню квадратному из определителя дисперсионной матрицы. Чем меньше величина этого определителя, тем меньше разброс оценок коэффициентов относительно их математических ожиданий, равных истинным значениям коэффициентов.

В случаях когда ошибка e распределена по нормальному закону (случайный вектор $\hat{a}$ также имеет при этом нормальный закон распределения), наглядная интерпретация критерия D-оптимальности может быть получена путем рассмотрения доверительного эллипсоида для истинных значений коэффициентов [10].

Доверительный эллипсоид для вектора коэффициентов размерности $k+1$ описывается следующим уравнением:

$$(a - \hat{a})' F'F (a - \hat{a}) = (k+1) s^2 F_{\text{кр}} = \delta^2. \qquad (6.9)$$

Здесь s^2 — оценка дисперсии ошибок наблюдений с числом степеней свободы φ_e; $F_{кр}$ — критическое значение распределения Фишера при выбранном уровне значимости α и числах степеней свободы $\varphi_1 = k+1$ и $\varphi_2 = \varphi_e$; $\hat{a}$ — вектор размерности $k+1$ оценок коэффициентов уравнения регрессии, найденный по методу наименьших квадратов с помощью плана с информационной матрицей $M = F'F$. Эллипсоид (6.9) с вероятностью $1-\alpha$ накрывает точку $\bar{a}$ истинных значений коэффициентов.



Фиг. 6.2. Эллипсоид рассеяния для оценок коэффициентов модели.

$\hat{a}$ — оценка коэффициентов; $\lambda_{макс}$ и $\lambda_{мин}$ — максимальное и минимальное собственные числа матрицы $C = (F'F)^{-1}$; δ^2 — правая часть уравнения (6.9).

На фиг. 6.2 приведен пример доверительного эллипсоида для двумерной задачи. Заметим, что величины осей доверительного эллипсоида оказываются пропорциональными собственным числам матрицы $C = (F'F)^{-1}$. Так, например, максимальная ось эллипсоида на фиг. 6.2 равна $2\delta\sqrt{\lambda_{макс}}$, где δ^2 — правая часть уравнения эллипсоида, а $\lambda_{макс}$ — максимальное собственное число матрицы $C = (F'F)^{-1}$. Сопоставление выражений (6.7) и (6.9) показывает, что объем доверительного эллипсоида, так же как и эллипсоида рассеяния, пропорционален корню квадратному из величины определителя матрицы C.

Фиг. 6.2 позволяет дать наглядную геометрическую интерпретацию и критерия A-оптимальности. Критерий A-оптимальности требует такого выбора плана эксперимента, при котором след матрицы C [след матрицы (Sp) — сумма диагональных элементов] имеет минимальную величину

$$\operatorname{Sp} C(X^*) = \min_{X \in \Omega_N} \operatorname{Sp} C(X).$$

Заключим доверительный эллипсоид в прямоугольник, как показано на фиг. 6.2. Так как сумма собственных значений матрицы равна сумме диагональных элементов этой матрицы, то диагональ описанного прямоугольника оказывается пропорциональной корню квадратному из следа матрицы C.

Из сказанного выше ясно, что для того, чтобы сформулировать задачу нахождения оптимального плана эксперимента, необходимо задаться видом модели (6.1), определить область Ω_x варьирования переменных x и выбрать число экспериментов N. Далее необходимо осуществить поиск наилучшего в смысле критерия (6.6) плана X^*. В принципе эта задача может быть решена с помощью различных численных методов оптимизации с использованием цифровых вычислительных машин [5].

Этот наиболее общий подход к проблеме планирования эксперимента, однако, далеко не всегда оказывается приемлемым. Другой путь практического использования критерия D-оптимальности для решения сформулированной задачи планирования эксперимента связан с отказом от строгого выполнения условий оптимальности. Этот путь заключается в следующем: каким-нибудь образом находят D-оптимальные планы для различных видов моделей и ряда типовых областей Ω_x, не связанные с заданным числом экспериментов, а лишь определяющие соотношение числа опытов в различных точках области Ω_x. Такие планы называются непрерывными. В следующем разделе мы рассмотрим понятие непрерывного плана более подробно. Множество таких планов может быть сведено в каталог D-оптимальных планов. Далее исследователь, задавшись некоторым N и пользуясь непрерывным планом, определяет конкретный план эксперимента, стремясь соблюсти соотношения частот непрерывного D-оптимального плана. При этом полученный им план может не быть строго D-оптимальным, а лишь близким к нему. Подобный подход к выбору плана эксперимента и имеется в виду в последующем изложении.

6.3. Точные и непрерывные D-оптимальные планы

Приведем некоторые определения.

Определение 6.1

План, заданный с помощью матрицы X размером $N \times n$, где n — число варьируемых переменных, а N — количество экспериментов, называется точным планом [1].

[1] Планы с фиксированным числом опытов называют также дискретными планами.

В плане X не обязательно все точки различные, т. е. в матрице X могут быть повторяющиеся строки. Предположим, что план X сосредоточен в r различных точках x^l, причем каждая из этих точек встретится в плане h_l раз. Тогда план X может быть представлен следующим образом:

$$\begin{bmatrix} x^1 & x^2 & \dots & x^r \\ h_1 & h_2 & \dots & h_r \end{bmatrix}, \tag{6.10}$$

причем

$$\sum_{l=1}^{r} h_l = N.$$

Определение 6.2

Точный план, заданный с помощью совокупности величин

$$\begin{bmatrix} x^1 & x^2 & \dots & x^r \\ l_1 & l_2 & \dots & l_r \end{bmatrix}, \tag{6.11}$$

где $l_l = h_l/N$ и $\sum_{l=1}^{r} l_l = 1$, называется нормированным планом. Очевидно, что l_l — доля наблюдений, приходящаяся на l-ю точку. Величину l_l мы будем называть частотой l-й точки плана.

Определение 6.3

План, заданный с помощью совокупности величин

$$l = \begin{bmatrix} x^1 & x^2 & \dots & x^r \\ l_1 & l_2 & \dots & l_r \end{bmatrix}, \tag{6.12}$$

где $\sum_{l=1}^{r} l_l = 1$, а величины l_l могут принимать любые значения между нулем и единицей, называется непрерывным.

Непрерывные планы мы будем обозначать буквой l. В общем случае непрерывный план может определяться некоторой непрерывной функцией $l(x)$, заданной на области Ω_x, т. е. не содержать конечного числа точек. В дальнейшем, однако, мы будем говорить лишь о непрерывных планах, сосредоточенных в конечном числе точек r, так как (ниже это будет показано) всегда существуют непрерывные D-оптимальные планы с конечным числом точек.

Обратим внимание на следующее замечание: всякому точному плану X соответствует некоторый нормированный, а следовательно, и непрерывный план l. Этот непрерывный план сосредоточен в точках $x^1, x^2, \dots, x^r$ точного плана с частотами $l_1 = h_1/N, \dots, l_r = h_r/N$. Однако не для всякого N можно найти точный план, такой, чтобы соотношение частот было точно таким

же, как у непрерывного плана l. Это можно сделать лишь тогда, когда все произведения Nl_i $(i = 1, 2, \ldots, r)$ являются целыми числами.

Приведем примеры.

Пример 6.1

Пусть задан точный план X:

$$X = \begin{bmatrix} -1 & -1 \\ +1 & -1 \\ -1 & +1 \\ +1 & +1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Обозначим через x^1 точку с координатами $x_1^1 = -1$, $x_2^1 = -1$,
» » x^2 » » » $x_1^2 = +1$, $x_2^2 = -1$,
» » x^3 » » » $x_1^3 = -1$, $x_2^3 = 1$,
» » x^4 » » » $x_1^4 = 1$, $x_2^4 = 1$,
» » x^5 » » » $x_1^5 = 0$, $x_2^5 = 0$.

Плану X соответствует непрерывный план

$$l = \begin{bmatrix} x^1 & x^2 & x^3 & x^4 & x^5 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/3 \end{bmatrix}.$$

Пример 6.2

Пусть задан непрерывный план

$$l = \begin{bmatrix} x^1 & x^2 & x^3 & x^4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}.$$

Этому плану соответствуют точные планы при $N = 4m$, где $m = 1, 2, \ldots,$

$$\begin{bmatrix} x^1 & x^2 & x^3 & x^4 \\ m & m & m & m \end{bmatrix}.$$

Пример 6.3

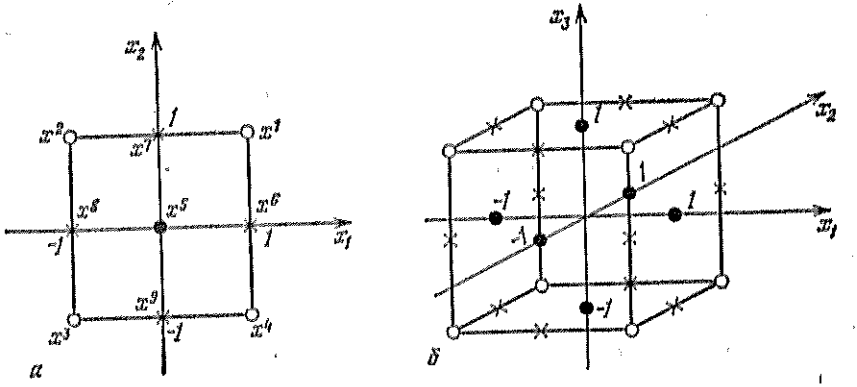
Рассмотрим непрерывный план l, точки сосредоточения которого показаны на фиг. 6.3, *а*. План содержит $r = 9$ точек, в число которых входят вершины квадрата (точки x^1, x^2, x^3, x^4), центр квадрата (x^5) и середины ребер (x^6, x^7, x^8, x^9). Частоты во всех вершинах квадрата одинаковы: $l_1 = l_2 = l_3 = l_4 = 0{,}14805$, частота в центре квадрата $l_5 = 0{,}0962$, а в серединах ребер $l_6 = l_7 = l_8 = l_9 = 0{,}08015$.

Этот план является непрерывным D-оптимальным планом [3] в случае, когда область планирования задана в виде квадрата

с вершинами в точках ± 1, а уравнение модели имеет вид

$$y(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2.$$

Минимальная величина N, при которой все Nl_i — целые числа, есть $N = 2 \cdot 10^4$. Для меньших значений N могут быть найдены



Фиг. 6.3. Точки непрерывного D-оптимального плана Кифера на кубе при $n = 2$ ($а$) и $n = 3$ ($б$) для квадратичных моделей.

точные планы, в которых частоты заданного непрерывного D-оптимального плана соблюдаются лишь приближенно.

Определение 6.4

Нормированной информационной матрицей плана X, состоящего из N опытов, называется матрица L вида

$$L = \frac{1}{N} F'F = \frac{1}{N} \sum_{j=1}^{N} f(x^j) f'(x^j) =$$

$$= \sum_{i=1}^{r} \frac{h_i}{N} f(x^i) f'(x^i) = \sum_{i=1}^{r} l_i f(x^i) f'(x^i). \qquad (6.13)$$

Здесь F — матрица функций независимых переменных, определяемая заданным видом уравнения регрессии. Очевидно, что L — информационная матрица непрерывного плана l, соответствующего X.

Определение 6.5

Нормированной дисперсионной матрицей плана X называется матрица D вида

$$D = L^{-1} = N(F'F)^{-1} = NC = \left(\sum_{i=1}^{r} l_i f(x^i) f'(x^i) \right)^{-1}. \qquad (6.14)$$

Очевидно, что D — дисперсионная матрица непрерывного плана l.

Пример 6.4

Найдем матрицы L и D для плана X примера 6.1 в случае, когда модель ищется в виде

$$y(x, a) = a_0 + a_1 x_1 + a_2 x_2.$$

Матрица F при этом имеет вид

$$F = \begin{bmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Отсюда находим

$$L = \frac{1}{6} F'F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2/3 & 0 \\ 0 & 0 & 2/3 \end{bmatrix},$$

$$D = L^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3/2 & 0 \\ 0 & 0 & 3/2 \end{bmatrix}.$$

Определение 6.6

План X^*, состоящий из N опытов, называется точным D-оптимальным планом, если он на множестве всех точных планов, состоящих из N опытов, размещенных в области Ω_x, минимизирует величину определителя дисперсионной матрицы (или максимизирует определитель информационной матрицы) [см. (6.6)].

Определение 6.7

План l^* называется непрерывным D-оптимальным планом, если он минимизирует на множестве всех непрерывных планов в области Ω_x величину определителя дисперсионной матрицы D (или максимизирует величину определителя информационной матрицы L), т. е.

$$|D(l^*)| = \min_l |D(l)|, \tag{6.15а}$$

$$|L(l^*)| = \max_l |L(l)|. \tag{6.15б}$$

Для нахождения непрерывного D-оптимального плана необходимо определить координаты x^l точек этого плана и частоты l_l повторения наблюдений в этих точках.

Аналогично определениям 6.6 и 6.7 введем теперь понятия точных и непрерывных A- и G-оптимальных планов.

Определение 6.8

План X^*, состоящий из N опытов, называется точным A-оптимальным планом [5], если он на множестве всех точных планов, состоящих из N опытов, размещенных в области Ω_x, минимизирует след дисперсионной матрицы C:

$$\operatorname{Sp} C(X^*) = \min_{x \in \Omega_x} \operatorname{Sp} C(X). \tag{6.16}$$

Определение 6.9

План l^* называется непрерывным A-оптимальным планом, если он на множестве всех непрерывных планов для области Ω_x минимизирует след матрицы D:

$$\operatorname{Sp} D(l^*) = \min_{l} \operatorname{Sp} D(l). \tag{6.17}$$

Определение 6.10

План X^*, состоящий из N опытов, называется точным G-оптимальным планом [5], если он на множестве всех планов, состоящих из N опытов, принадлежащих области Ω_x, минимизирует максимальную величину дисперсии рассчитанной по модели оценки целевой функции для области Ω_x:

$$\max_{x \in \Omega_x} \sigma^2_{\hat{y}}(X^*) = \min_{X \in \Omega_x} \max_{x \in \Omega_x} \sigma^2_{\hat{y}}(X) = \min_{X \in \Omega_x} \max_{x \in \Omega_x} f'(x) C f(x) \sigma^2. \tag{6.18}$$

Обозначим через $d(x, l)$ функцию

$$d(x, l) = f'(x) D f(x). \tag{6.19}$$

Функция $d(x, l)$ следующим образом связана с дисперсией оценки целевой функции в точке x в случае, когда модель построена с помощью плана X:

$$\sigma^2_{\hat{y}(x)} = f'(x)(F'F)^{-1} f(x) \sigma^2 = f'(x) \frac{1}{N} D f(x) \sigma^2 = \frac{\sigma^2}{N} d(x, l). \tag{6.20}$$

Величину $d(x, l)\sigma^2$ мы будем называть дисперсией оценки зависимой переменной в точке x для непрерывного плана l.

Определение 6.11

Непрерывный план l^*, который в области Ω_x минимизирует максимум величины $d(x, l)$, называется G-оптимальным планом. Условие G-оптимальности плана l^* имеет вид

$$\max_{x \in \Omega_x} d(x, l^*) = \min_{l} \max_{x \in \Omega_x} d(x, l) = \min_{l} d_{\text{макс}}(l). \tag{6.21}$$

Утверждение 6.1

Если нормированное плана X приводит к плану l, который является непрерывным D-(или A-, или G-)оптимальным планом, то X есть точный D-(или A-, или G-)оптимальный план.

Обратное утверждение, однако, неверно. Если план X — точный D-(или A-, или G-)оптимальный план, сосредоточенный в r точках с частотами h_1/N, h_2/N, ..., h_r/N, то непрерывный план l, сосредоточенный в тех же точках с частотами $l_1 = h_1/N$, $l_2 = h_2/N$, ..., $l_r = h_r/N$, не всегда является D-(или A-, G-)оптимальным планом.

6.4. Некоторые свойства непрерывных D-оптимальных планов

Приведем здесь без доказательства ряд теорем о непрерывных D-оптимальных планах, которые ниже будут нами использоваться.

Теорема 6.1 [1, 5]

Всегда можно найти непрерывный D-оптимальный план, сосредоточенный в конечном числе точек r, причем

$$k+1 \leqslant r \leqslant \frac{(k+1)(k+2)}{2}, \tag{6.22}$$

где $k+1$ — число оцениваемых коэффициентов модели.

Здесь нужно заметить, что для одной и той же задачи (т. е. для заданного вида уравнения регрессии и заданной области планирования Ω_x) можно найти бесконечное множество непрерывных D-оптимальных планов, отличающихся как числом точек, так и частотами, но имеющих одну и ту же величину определителя дисперсионной матрицы. С точки зрения экспериментатора интерес представляют планы, сосредоточенные в малом числе точек. Поэтому теорема 6.1, устанавливающая границы числа точек D-оптимального плана, оказывается полезной при построении планов.

Приведем теперь формулировку очень важной теоремы [11, 5, 4].

Теорема 6.2

Непрерывный план l тогда и только тогда D-оптимален, когда

$$\max_{x \in \Omega_x} d(x, l) = k+1, \tag{6.23}$$

где $k+1$ — число оцениваемых коэффициентов, а Ω_x — область планирования эксперимента.

Теорема 6.2 дает способ проверки всякого непрерывного плана на близость к D-оптимальному. Для этой цели часто используют показатель

$$\delta = \frac{\max\limits_{x \in \Omega_x} d(x, l) - (k+1)}{k+1}, \qquad (6.24)$$

где l — исследуемый план.

Пользуясь тем, что всякому точному плану соответствует некоторый непрерывный план, из теоремы 6.2 можно получить достаточные условия D-оптимальности некоторого точного плана X.

Правило 6.1

Если для некоторого плана X имеем

$$N \max_{x \in \Omega_x} f'(x)(F'F)^{-1} f(x) = \max_{x \in \Omega_x} f'(x) Df(x) = d_{\text{макс}}(l) = k+1, \qquad (6.25)$$

то план X является D-оптимальным. Здесь D — нормированная дисперсионная матрица плана X.

Приведем теперь два утверждения, устанавливающих связь между свойствами D-, G- и A-оптимальности планов эксперимента.

Утверждение 6.2 [5]

Если l — непрерывный D-оптимальный план, то он является также и G-оптимальным планом. Это означает, в частности, что теорема 6.2 и правило 6.1 устанавливают условия как D-, так и G-оптимальности.

Утверждение 6.3 [5]

Если l — непрерывный D-(или A-)оптимальный план и имеет место

$$\beta D(l) = D^a(l), \qquad (6.26)$$

где β — некоторая постоянная величина, то план l является одновременно и A-(соответственно D-)оптимальным планом. Условие (6.26) имеет место, в частности, для

$$D(l) = I_{k+1}. \qquad (6.27)$$

Отсюда следует

Утверждение 6.4

Если X — точный D-оптимальный план и

$$N(F'F)^{-1} = I_{k+1}, \qquad (6.28)$$

то план X является также и A-оптимальным.

Пример 6.5

Пусть для оценки коэффициентов модели вида

$$y(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2$$

в области Ω_x, заданной неравенствами $-1 \leqslant x_1 \leqslant +1$, $-1 \leqslant x_2 \leqslant 1$, использован полный факторный эксперимент (план типа 2^2):

$$X = \begin{bmatrix} -1 & -1 \\ +1 & -1 \\ -1 & +1 \\ +1 & +1 \end{bmatrix}.$$

Убедимся в том, что план X является D-, G- и A-оптимальным. Матрица F в данном примере

$$F = \begin{bmatrix} 1 & -1 & -1 & +1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 \end{bmatrix}.$$

Нетрудно найти матрицу

$$(F'F)^{-1} = \begin{bmatrix} 1/4 & & 0 & \\ & 1/4 & & \\ 0 & & 1/4 & \\ & & & 1/4 \end{bmatrix}.$$

Очевидно, что $N(F'F)^{-1} = I_4$.

Максимум квадратичной формы (6.19) в области Ω_x достигается в углах квадрата. Очевидно, что

$$\max_{\substack{-1 \leqslant x_1 \leqslant 1 \\ -1 \leqslant x_2 \leqslant 1}} (1 + x_1^2 + x_2^2 + x_1^2x_2^2) = 4.$$

Так как в нашем случае $k + 1 = 4$, то на основании правила 6.1 и утверждения 6.2 план X является точным D- и G-оптимальным. Кроме того, на основании утверждения 6.4, план X является также A-оптимальным. Подобным образом легко можно убедиться в правильности следующего утверждения.

Утверждение 6.5

Планы полного и дробного факторного эксперимента, используемые для оценки коэффициентов линейных моделей (а также и смешанных моделей, содержащих взаимодействия любого порядка, см. разд. 5.4 и 5.5), являются D-, G- и A-оптимальными планами, если область планирования есть гиперкуб

с координатами вершин в точках ± 1, а матрица F плана не содержит одинаковых столбцов (т. е. оценки коэффициентов не смешаны).

Наконец, приведем еще один важный результат [2].

Теорема 6.3

D-оптимальный план инвариантен относительно линейного преобразования пространства x.

Эта теорема позволяет нам рассматривать D-оптимальные планы на некоторых типизированных областях Ω_x, например на гиперкубе с координатами вершин в точках ± 1. Изменение масштаба при переходе в дальнейшем к натуральным единицам измерения не повлияет на расположение и частоты точек D-оптимального плана.

6.5. Непрерывные D-оптимальные планы на отрезке

Рассмотрим случай, когда искомая модель представляется с помощью полинома произвольной степени d в одномерном пространстве, причем областью планирования является отрезок $(-1, +1)$, т. е.

$$-1 \leqslant x \leqslant 1.$$

Модель имеет вид

$$y(a, x) = \sum_{i=0}^{d} a_i x^i. \tag{6.29}$$

Непрерывные D-оптимальные планы для этого случая найдены для любого d [12]. Эти планы сосредоточены в $(d+1)$ точке, причем частоты всех точек одинаковы, т. е.

$$l_1 = l_2 = \ldots = l_{d+1} = \frac{1}{d+1}. \tag{6.30}$$

Координатами точек плана являются корни полинома $G(x)$ степени $d+1$, представляющего собой коэффициент при z^{d+1} в разложении функции $g(z) = \sqrt{1-2xz+z^2}$ в ряд по степеням z в точке $z = 0$.

Найдем в качестве примера координаты точек непрерывного D-оптимального плана при $d = 2$. Модель при этом имеет вид

$$y(x, a) = a_0 + a_1 x + a_2 x^2.$$

Разложим функцию $g(z)$ в ряд по степеням z в точке $z = 0$ и найдем коэффициент разложения при z^3 из

$$\frac{1}{3!}\frac{\partial^3 g(z)}{\partial z^3} = \frac{1}{3!}\frac{\partial^3}{\partial z^3}\sqrt{1-2xz+z^2} =$$

$$= \frac{1}{2}(x-z)(1-x^2)(1-2xz+z^2)^{-5/2}.$$

Таблица 6.1 (продолжение)

$D(l^*)$						$\lvert D(l^*)\rvert$
						$0{,}6749 \cdot 10^{1}$
						$0{,}1953 \cdot 10^{3}$
11,654 0 −83,766 0 76,491						$0{,}2326 \cdot 10^{5}$
14,752 0 −110,22 0 103,34 0	0 111,36 0 −413,82 0 310,34					$0{,}1154 \cdot 10^{8}$
90,23 0 −1375,8 0 3273,2 0 −2000,6	0 139,38 0 −534,52 0 409,26 0	−47,41 0 796,18 0 −2000,6 0 1266,2				$0{,}2211 \cdot 10^{11}$
106,44 0 −1694,5 0 4162,51 0 −2605,5 0	0 1292,28 0 −9274,4 0 17094,6 0 −9143,6	−58,46 0 1013,62 0 −2605,5 0 1677,17 0	0 −626,93 0 4775,05 0 −9143,6 0 5022,33			$0{,}1734 \cdot 10^{15}$
368,91 0 −9768,3 0 43126,1 0 −62897,1 0 29193,7	0 1622,06 0 −11962,6 0 22501,6 0 −12223,0 0	−463,45 0 13493,4 0 −62897,1 0 95061,3 0 −45256,2	0 −815,98 0 6309,52 0 −12223,0 0 6774,56 0	195,50 0 −6036,2 0 29193,7 0 −45256,2 0 21948,3		$0{,}5679 \cdot 10^{19}$

Таблица 6.1

Непрерывные D-оптимальные планы для полиномов на отрезке

	x^l	l^l	$D(l^*)$			
	−1	1/3	3,0	0	−3,0	
	0		0	1,5	0	
	+1		−3,0	0	4,5	
	−1	1/4	3,248	0	−3,748	0
	−0,447		0	15,757	0	−16,257
	+0,447		−3,748	0	6,247	0
	+1		0	−16,257	0	18,756
	−1	1/5	5,0	0	−16,653	0
	−0,655		0	19,284	0	−21,163
	0		−16,653	0	98,541	0
	+0,655		0	−21,163	0	25,541
	+1		11,654	0	−83,766	0
	−1	1/6	5,293	0	−19,671	0
	−0,765		0	59,838	0	−170,82
	−0,285		−19,671	0	124,64	0
	0,285		0	−170,82	0	579,4
	0,765		14,752	0	110,22	0
	+1		0	111,36	0	−413,82
	−1	1/7	7	0	−49,82	0
	−0,830		0	67,75	0	−205,22
	−0,463		−49,82	0	631,52	0
	0		0	−205,22	0	720,80
	0,463		90,23	0	−1375,8	0
	0,830		0	139,58	0	−534,52
	1		−47,41	0	796,18	0
	−1	1/8	7,30	0	−55,58	0
	−0,872		0	149,20	0	−814,90
	−0,592		−55,58	0	744,92	0
	−0,209		0	−814,86	0	5322,63
	+0,209		106,44	0	−1694,50	0
	+0,592		0	1292,28	0	−9274,4
	+0,872		58,46	0	1013,62	0
	+1		0	−626,93	0	4775,05
	−1	1/9	9,009	0	−109,96	0
	−0,89		0	168,19	0	−976,41
	−0,67		−109,96	0	2418,89	0
	−0,36		0	−976,41	0	6652,87
	0		368,91	0	−9768,30	0
	+0,36		0	1622,06	0	−11962,6
	+0,67		−463,45	0	13493,4	0
	+0,89		0	−815,98	0	6309,52
	1		195,50	0	−6036,2	0

Отсюда нетрудно получить уравнение $G(x)=0$ для определения координат плана D-оптимального плана:

$$G(x)=\frac{1}{3!}\frac{\partial^3 g(z)}{\partial z^3}\Big|_{z=0}=\frac{1}{2}x(1-x^2). \tag{6.31}$$

Решение уравнения $G(x)=0$ имеет вид

$$\begin{aligned} x^1&=0,\\ x^2&=+1,\\ x^3&=-1. \end{aligned} \tag{6.32}$$

Таким образом, непрерывный D-оптимальный план сосредоточен в точках (6.32), причем частоты повторения наблюдений одинаковы и равны $l_1=l_2=l_3=1/3$.

Так как частоты всех $(d+1)$ точек плана оказываются одинаковыми, то, используя непрерывный D-оптимальный план, легко получить точный D-оптимальный план при $N=m(d+1)$, где $m=1, 2,\ldots$ Этот план будет сосредоточен в точках непрерывного плана, причем каждая точка будет в матрице X повторяться m раз.

В табл. 6.1 приведены значения координат точек непрерывных D-оптимальных планов для $d=2, 3, \ldots, 8$. Здесь же приведены нормированные дисперсионные матрицы D размером $(d+1)\times(d+1)$, предназначенные для использования при вычислении оценок коэффициентов модели, а также величины определителя нормированной дисперсионной матрицы D.

Матрица D определяется выражением

$$D=N(F'F)^{-1}=(d+1)(F'F)^{-1}. \tag{6.33}$$

Как следует из (6.3), вектор $\hat{a}$ выражается в виде

$$\hat{a}=\frac{1}{d+1}DF'\bar{Y}, \tag{6.34}$$

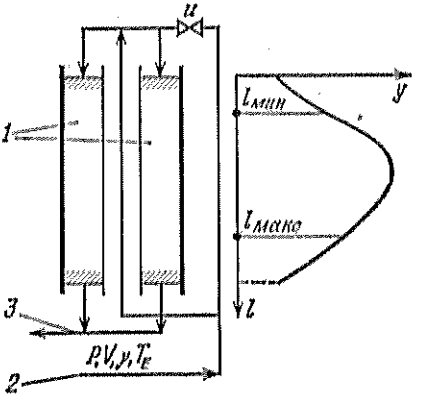
где $\bar{Y}$ — вектор наблюдений размерности $(d+1)$. Матрица F в этом случае имеет вид

$$F=\begin{bmatrix} 1 & x^1 & (x^1)^2 & \ldots (x^1)^d\\ 1 & x^2 & (x^2)^2 & \ldots (x^2)^d\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ 1 & x^{d+1} & (x^{d+1})^2 & \ldots (x^{d+1})^d \end{bmatrix}. \tag{6.35}$$

Если число опытов $N=m(d+1)$, где $m>1$, то вместо вектор-столбца $\bar{Y}$ используется столбец Y, элементами которого являются средние значения переменной y для всякой точки плана.

Пример 6.6

Для управления процессом синтеза аммиака [19] важно знать максимальную температуру в катализаторной зоне колонны (температуру „горячей точки"). Колонна синтеза условно изображена на фиг. 6.4. Здесь *1*—катализаторная зона колонны; *2*—поступающий в колонну контактный газ, характеризуемый отношением водорода к азоту γ, расходом V, давлением P, температурой T_E и рядом других параметров; u—степень открытия байпаса, определяющая соотношение газовых потоков; *3*—выходной газ.



Фиг. 6.4. Схематическое изображение реактора синтеза аммиака (к примеру 6.6).

Справа на фиг. 6.4 показан приблизительный вид кривой изменения температуры по высоте колонны. Положение „горячей точки" и величина максимальной температуры зависят от режима работы колонны, определяемого значениями u, P, V, γ, T_E и другими параметрами. Требуется определить точки по высоте колонны для установки трех термопар, с помощью которых можно было бы с наименьшей ошибкой определять температуру в „горячей точке".

Для решения этой задачи могут быть использованы методы планирования эксперимента. Предположим, что z^1, z^2 и z^3—координаты точек по высоте колонны, где производятся измерения температуры y. Пусть „горячая точка" находится где-то внутри диапазона ($z_{мин}$, $z_{маке}$), причем зависимость температуры y от z для этого диапазона описывается параболой

$$y(b, z) = b_0 + b_1 z + b_2 z^2.$$

Оценки коэффициентов b_0, b_1 и b_2 могут быть найдены по результатам $\tilde{y}^1$, $\tilde{y}^2$, $\tilde{y}^3$ измерений в точках z^1, z^2, z^3 с помощью метода наименьших квадратов, т. е. по формуле (6.3). Тогда для оценки

температуры $\hat{y}_{\text{макс}}$ „горячей точки" имеем

$$\hat{y}_{\text{макс}} = \max_{z_{\text{мин}} \leqslant z \leqslant z_{\text{макс}}} (\hat{b}_0 + \hat{b}_1 z + \hat{b}_2 z^2) = \max_{z_{\text{мин}} \leqslant z \leqslant z_{\text{макс}}} \hat{y}(z).$$

Для выбора точек измерения z^1, z^2, z^3 здесь уместно воспользоваться критерием G-оптимальности, потребовав, чтобы выполнялось условие

$$\max_{z_{\text{мин}} \leqslant z \leqslant z_{\text{макс}}} \sigma^2_{\hat{y}(z)}(z_*^1, z_*^2, z_*^3) = \min_{z^1, z^2, z^3} \max_{z_{\text{мин}} \leqslant z \leqslant z_{\text{макс}}} \sigma^2_{\hat{y}}(z^1, z^2, z^3).$$

G- и D-оптимальные непрерывные планы для этого случая приведены в табл. 6.1. План содержит $r = 3$ точки с равными частотами. Так как число измерений $N = r = 3$, то точные G- и D-оптимальные планы состоят из точек непрерывного плана.

Приведем теперь некоторые числовые расчеты. Положим $z_{\text{мин}} = 1{,}81$ м, $z_{\text{макс}} = 7{,}32$ м. Перейдем к переменной x по формуле

$$x = \frac{z - \frac{z_{\text{макс}} + z_{\text{мин}}}{2}}{\frac{z_{\text{макс}} - z_{\text{мин}}}{2}} = \frac{z - 4{,}565}{2{,}755}.$$

Теперь областью планирования является отрезок

$$-1 \leqslant x \leqslant +1.$$

Модель с новой переменной имеет вид

$$y = a_0 + a_1 x + a_2 x^2.$$

В табл. 6.1 находим точки оптимального плана

$$x^1 = -1,$$
$$x^2 = 0,$$
$$x^3 = 1.$$

Пусть в каждой из выбранных точек проведено по два измерения температуры. Результаты этих измерений приведены в табл. 6.2. Здесь $\tilde{y}^{i1}$ и $\tilde{y}^{i2}$ обозначают результаты параллельных наблюдений в i-й точке плана, а $\bar{y}^i$ — среднее значение наблюдений в этой точке. С помощью этих данных рассчитаем оценку s_e^2 дисперсии ошибки единичного наблюдения:

$$s_e^2 = \frac{1}{3}\sum_{i=1}^{3} (\tilde{y}^{i1} - \bar{y}^i)^2 + (\tilde{y}^{i2} - \bar{y}^i)^2 =$$
$$= \frac{1}{3}[2(1{,}35)^2 + 2(0{,}6)^2 + 2(1{,}55)^2] = 3{,}05.$$

Число степеней свободы оценки s_e^2 равно $\varphi_e = 3$.

Таблица 6.2

t	$\tilde{y}^{(1)}$	$\tilde{y}^{(2)}$	$\tilde{y}^{(3)}$
1	454,3	451,6	452,95
2	522,3	521,1	521,7
3	463,5	466,6	465,05

Матрица F имеет в нашем случае вид

$$F = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Умножая F' на $\tilde{Y}$, найдем

$$F'\tilde{Y} = \begin{bmatrix} 452{,}95 + 521{,}7 + 465{,}05 \\ -452{,}95 + 465{,}05 \\ 452{,}95 + 465{,}05 \end{bmatrix} = \begin{bmatrix} 1433{,}7 \\ 12{,}1 \\ 918 \end{bmatrix}.$$

Пользуясь матрицей D, приведенной в табл. 6.1, и применяя формулу (6.3), получим

$$\hat{a} = \frac{1}{3} DF'\tilde{Y} = \frac{1}{3} \begin{bmatrix} 3 & 0 & -3 \\ 0 & 1{,}5 & 0 \\ -3 & 0 & 4{,}5 \end{bmatrix} \begin{bmatrix} 1433{,}7 \\ 12{,}1 \\ 918 \end{bmatrix} = \begin{bmatrix} 521{,}7 \\ 6{,}05 \\ 112{,}7 \end{bmatrix}.$$

Таким образом, имеем следующее уравнение:

$$\hat{y}(x) = 521{,}7 + 6{,}05x - 112{,}7x^2.$$

Максимальное значение температуры имеет место в точке $x^*_{\text{макс}} = 0{,}027$ и равно

$$\hat{y}_{\text{макс}} = 521{,}9.$$

Нетрудно рассчитать координату $z^*_{\text{макс}}$, соответствующую $x^*_{\text{макс}}$:

$$z^*_{\text{макс}} = x^*_{\text{макс}} \times 2{,}755 + 4{,}565 = 4{,}639.$$

Дисперсия оценки температуры в точке $x^*_{\text{макс}}$ определяется по формуле

$$s^2_{\hat{y}}(x^*_{\text{макс}}) = f'(x^*_{\text{макс}})(F'F)^{-1} f(x^*_{\text{макс}}) s^2 =$$

$$= N^{-1} f'(x^*_{\text{макс}}) Df(x^*_{\text{макс}}) \frac{s^2_\theta}{2}.$$

Здесь s^2 — дисперсия ошибки наблюдений [1], D — дисперсионная матрица непрерывного D-оптимального плана, $(x^*_{\text{макс}})$ — вектор-столбец с координатами 1, $x^*_{\text{макс}}$, $(x^*_{\text{макс}})^2$. Пользуясь матрицей D, приведенной в табл. 6.1, получаем

$$s^2_{\hat{y}}(x^*_{\text{макс}}) = \frac{1}{3}\left(3 - 6x^{*2}_{\text{макс}} + 1{,}5x^{*2}_{\text{макс}} + 4{,}5x^{*4}_{\text{макс}}\right)\frac{s^2_e}{2} \approx \frac{s^2_e}{2} = 1{,}525.$$

6.6. Непрерывные D-оптимальные планы для квадратичной регрессии на гиперкубе

Рассмотрим случай, когда уравнение регрессии включает, помимо линейных членов, парные взаимодействия и квадраты независимых переменных:

$$y(x, a) = a_0 + a_1x_1 + \ldots + a_nx_n + a_{n+1}x_1^2 + \\ + \ldots + a_{2n}x_n^2 + a_{2n+1}x_1x_2 + \ldots + a_kx_{n-1}x_n. \qquad (6.36)$$

Область планирования — гиперкуб в n-мерном пространстве, определяемый неравенствами

$$-1 \leqslant x_i \leqslant 1, \quad i = 1, 2, \ldots, n. \qquad (6.37)$$

Непрерывные D-оптимальные планы для этого случая для размерностей пространства независимых переменных n от 2 до 7 построены аналитически в работах Кифера [3] и Коно [13]. Планы Кифера концентрируются в следующих точках n-мерного гиперкуба (6.37): в вершинах гиперкуба (множество точек плана E_0), в серединах ребер (множество E_1) и в центрах двумерных граней (множество E_2).

Обозначим через α частоту точек множества E_0 (вершин куба), через β — частоту точек множества E_1 (середин ребер) и через γ — частоту точек множества E_2 (центров двумерных граней). В табл. 6.3 приведены значения этих частот для D-оптимальных планов, найденных в [3] (план Кифера). Оказалось, что построить D-оптимальные планы, сосредоточенные в этих множествах точек, можно лишь при размерности $n \leqslant 5$. В табл. 6.3 указано число точек каждого типа и общее число r точек непрерывных D-оптимальных планов Кифера. Очевидно, что число вершин n-мерного куба есть 2^n, число ребер равно $n2^{n-1}$, а двумерных граней $n(n-1)2^{n-3}$. Пример расположения точек непрерывного D-оптимального плана для $n = 2$ приведен на

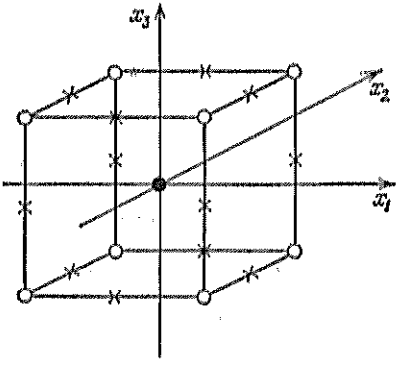
[1] В данном примере при расчетах оценок коэффициентов используются средние значения наблюдений по двум параллельным опытам в каждой точке плана (это ясно из вида матрицы F). При этом оценка s^2 дисперсии ошибок наблюдений следующим образом связана с оценкой s^2_e дисперсии единичного наблюдения: $s^2 = s^2_e/2$. — *Прим. ред.*

Таблица 6.3

Планы Кифера для квадратичной регрессии на гиперкубе

n	Вершины куба E_0		Середины ребер E_1		Центры двумерных граней E_2		Число точек плана r
	α	2^n	β	$n2^{n-1}$	γ	$n(n-1)2^{n-2}$	
1	0,6666	2	0,3333	1	0	0	3
2	0,5832	4	0,3206	4	0,0962	1	9
3	0,5758	8	0,2274	12	0,1968	6	26
4	0,5928	16	0,1228	32	0,2844	24	72
5	0,6170	32	0,02548	80	0,358	80	192

фиг. 6.3, *а*. Здесь точки x^1, x^2, x^3 и x^4 относятся к множеству E_0; x^6, x^7, x^8, x^9 — точки множества E_1; x^5 — точка множества E_2. На фиг. 6.3, *б* показаны точки плана Кифера при $n=3$. Здесь кружочки использованы для обозначения точек множества E_0, крестики — для точек множества E_1 и зачерненные кружочки — для точек множества E_2. Заметим, что частота всех точек плана,



Фиг. 6.5. Точки непрерывного D-оптимального плана Коно для квадратичной модели на кубе ($n=3$).

принадлежащих одному и тому же множеству E_l, одинакова. Например, если x^l — вершина куба, то $l_l=\alpha/2^n$; если x^l — середина ребра, то $l_l=\beta/n2^{n-1}$ и т. д.

Обратим теперь внимание на число точек r, в которых концентрируются планы Кифера. В табл. 6.4 указаны число коэффициентов $(k+1)$ в уравнении регрессии, максимальные числа $r_{\text{макс}}$ точек непрерывного D-оптимального плана, определенные по теореме 6.1, и приведены также количества точек в планах Кифера. Видно, что при $n\geqslant 4$ число точек планов Кифера пре-

Таблица 6.4

Число точек непрерывных D-оптимальных планов для квадратичной регрессии на гиперкубе

n	Число коэффициентов $k+1$	$r_{\text{макс}}=\frac{(k+1)(k+2)}{2}$	Число точек плана	
			план Кифера	план Коно
1	3	5	3	3
2	6	15	9	9
3	10	35	26	21
4	15	70	72	49
5	21	126	196	113
6	28	210	—	257
7	36	330	—	577

восходит $r_{\text{макс}}$, что свидетельствует о возможности построения D-оптимальных планов с меньшим числом точек. Такие планы получены Коно в работе [13] (планы Коно).

Планы Коно концентрируются в следующих точках гиперкуба (6.37): в вершинах гиперкуба (множество E_0), в серединах ребер (множество E_1) и в центре n-мерного куба (множество E_n). Число точек каждого множества и их частоты для n от 1 до 7 приведены в табл. 6.5. Частота точки, расположенной в центре гиперкуба, обозначена через ξ. Остальные обозначения те же, что и в табл. 6.3.

Очевидно, что для $n=2$ планы Коно совпадают с планами Кифера. Точки плана Коно для $n=3$ показаны на фиг. 6.5. Здесь приняты те же обозначения, что и на фиг. 6.3, только

Таблица 6.5

Планы Коно для квадратичной регрессии на гиперкубе

n	Вершины гиперкуба E_0		Середины ребер E_1		Центр куба E_n		Число точек плана r
	α	2^n	β	$n2^{n-1}$	ξ		
1	0,6666	2	0,3333	1	—	—	3
2	0,5832	4	0,3206	4	0,0962	1	9
3	0,5103	8	0,4242	12	0,06547	1	21
4	0,4506	16	0,5021	32	0,04738	1	49
5	0,4021	32	0,5622	80	0,03675	1	113
6	0,3623	64	0,6097	192	0,02710	1	257
7	0,3297	128	0,6486	448	0,02160	1	577

зачерненным кружком здесь обозначена точка в центре n-мерного куба. Заметим, что при $n = 6,7$ число точек планов Коно превосходит $r_{\text{макс}}$ (см. табл. 6.4).

6.7. Численный метод построения непрерывных D-оптимальных планов

Аналитический путь построения непрерывных D-оптимальных планов, использованный для получения планов разд. 6.5 и 6.6, оказывается возможным лишь в простейших случаях. На практике иногда возникают задачи, в которых область планирования Ω_x не имеет правильной геометрической формы, а модель представляется неполным полиномом второй степени или полиномом более высокой степени. В этих случаях для построения оптимальных непрерывных планов целесообразно использовать численные методы, связанные с максимизацией определителя информационной матрицы. На одном из таких методов мы здесь кратко остановимся. Этот метод основан на применении процедуры последовательного планирования эксперимента (см. разд. 11) [14, 5].

Пусть требуется построить непрерывный D-оптимальный план для оценки коэффициентов модели вида (6.1), где $f_j(x)$, $j = 0, \ldots, k$, — заданная система функций, а область планирования есть некоторая произвольная заданная область Ω_x в n-мерном пространстве x. Пусть имеется некоторый план X_0, состоящий из $N \geqslant k+1$ наблюдений из области Ω_x, такой, что определитель информационной матрицы $|F_0'F_0| \neq 0$. На первом шаге находим точку x_*^1 из условия

$$f'(x_*^1)(F_0'F_0)^{-1} f(x_*^1) = \max_{x^1 \in \Omega_x} f'(x^1)(F_0'F_0)^{-1} f(x^1).$$

Точку x_*^1 теперь следует присоединить к матрице X_0, в результате чего получаем план X_1 с информационной матрицей

$$F_1'F_1 = F_0'F_0 + f(x_*^1) f'(x_*^1).$$

Поступая аналогичным образом и далее, мы на l-м шаге должны будем проделать следующие вычисления:

1) определить точку x_*^l из условия

$$f'(x_*^l)(F_{l-1}'F_{l-1})^{-1} f(x_*^l) = \max_{x^l \in \Omega_x} f'(x^l)(F_{l-1}'F_{l-1})^{-1} f(x^l); \quad (6.38)$$

2) найти информационную матрицу плана X_l, включающего точку x_*^l:

$$F_l'F_l = F_{l-1}'F_{l-1} + f(x_*^l) f'(x_*^l). \quad (6.39)$$

Обозначим через l^l непрерывный план, соответствующий плану X_l, полученному после l шагов применения процедур

(6.38) и (6.39). Тогда можно доказать следующее утверждение [5, 14]:

$$\lim_{i \to \infty} |L(l^i)| = |L(l^*)| = \max_{l} |L(l)|, \qquad (6.40)$$

где $L(l^i)$ — информационная матрица плана l^i, а $L(l^*)$ — информационная матрица D-оптимального плана. Таким образом, при достаточно большом i с помощью указанной вычислительной процедуры можно найти непрерывный D-оптимальный план.

В работе [15] описан конкретный алгоритм, построенный на изложенном выше принципе, а в [5] обсуждаются различные модификации алгоритмов построения планов.

После всякого конечного числа шагов i применения описанной выше вычислительной процедуры мы получаем план l^i, который может быть как угодно близок к D-оптимальному плану l^*, но все же отличается от него. Проверка близости плана l^i к непрерывному D-оптимальному осуществляется путем вычисления величины δ по формуле (6.24). Если начальный план X_0 состоял из N наблюдений, то величина δ после i итераций будет равна

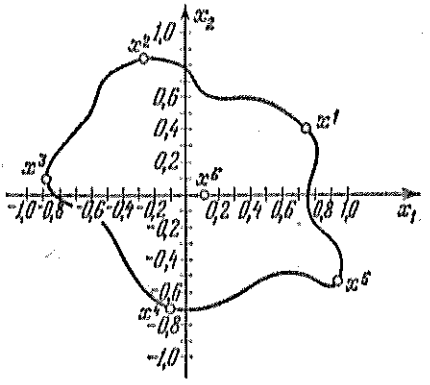
$$\delta = \frac{(N+i)\max\limits_{x \in \Omega_x} [f'(x)(F_i'F_i)^{-1} f(x)] - (k+1)}{k+1} = \frac{\max\limits_{x \in \Omega_x} d(x, l^i) - (k+1)}{k+1}. \qquad (6.41)$$

При вычислении δ требуется находить максимум квадратичной формы $d(x, l^i) = f'(x) D(l^i) f(x)$ по области Ω_x. Если эта задача решается не аналитически, а путем поиска, например, с помощью цифровой вычислительной машины, то полученные значения $d_{\text{макс}}(l^i)$ и значения δ нельзя, строго говоря, считать истинными. В дальнейшем мы такие значения $d_{\text{макс}}$ и δ будем отмечать звездочкой, хотя во всех случаях эти величины были определены после весьма тщательных вычислений.

Величина δ равна 0, когда план l^i является D-оптимальным. Задаваясь некоторой степенью близости δ^0 плана l^i к D-оптимальному, можно прекращать расчеты, связанные с поиском плана, когда $\delta \leqslant \delta^0$.

Заметим, что план l^i, полученный в результате расчетов, может содержать „посторонние" точки с малыми частотами, появившиеся в ходе вычислений в результате неудачного выбора начального плана X_0; кроме того, в плане l^i могут иметься чрезвычайно близкие точки, различие значений координат которых может быть вызвано ограниченной точностью вычислений. Так как планы с большим числом точек нежелательны, то вместо плана l^i в этих случаях целесообразно рассматривать некоторое его приближение $\tilde{l}^i$, где близкие точки объединяются в одну группу, а отдельные точки с малыми частотами, не тяготеющие

ни к одной из групп, отбрасываются. Близость плана $\tilde{\xi}'$ к D-оптимальному должна быть проверена путем вычисления величины δ.



Фиг. 6.6. Область планирования и точки непрерывного D-оптимального плана примера 6.7.

Приведем здесь ряд примеров непрерывных D-оптимальных планов (или планов, близких к D-оптимальным), полученных с помощью описанной вычислительной процедуры.

Пример 6.7

Непрерывный D-оптимальный план для уравнения регрессии вида

$$y(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2 \qquad (6.42)$$

и области планирования Ω_x, показанной на фиг. 6.6, был найден с помощью изложенного метода [15]. Точки плана показаны на фиг. 6.6, а значения координат и частоты приведены в табл. 6.6. План концентрируется в шести точках с равными частотами. Здесь $\delta^* = 0$.

Пример 6.8

Непрерывный D-оптимальный план для модели

$$y(x, a) = a_0 + a_1e^{x_1}\sin 2x_2 + a_2e^{-2x_1}\cos\frac{x_2}{2} + a_3e^{-x_1}\sin 2x_2\cos\frac{x_2}{2} \qquad (6.43)$$

и области планирования, заданной неравенствами

$$-1 \leqslant x_1 \leqslant 1,$$
$$-1 \leqslant x_2 \leqslant 1,$$

Таблица 6.6

D-оптимальный план (пример 6.7)
$\delta^* = 0$

i	x_1^i	x_2^i	l_i
1	0,75	0,4	1/6
2	—0,265	0,81	1/6
3	—0,89	0,1	1/6
4	—0,1	0,7	1/6
5	0,93	—0,55	1/6
6	0,1	0	1/6

Таблица 6.7

D-оптимальный план (пример 6.8)
$\delta^* = 0$

i	x_1^i	x_2^i	l_i
1	—1	$\pi/4$	1/4
2	+1	$\pi/4$	1/4
3	—1	$-\pi/4$	1/4
4	+1	$-\pi/4$	1/4

полученный с помощью изложенного метода [15], имеет вид, показанный в табл. 6.7. И здесь $\delta^* = 0$.

Пример 6.9

Непрерывные D-оптимальные планы для неполных полиномов второго порядка [15].

В табл. 6.8 приведены некоторые планы для неполных квадратичных полиномов при $n = 2$ в случае, когда область планирования — квадрат:

$$-1 \leqslant x_1 \leqslant +1,$$
$$-1 \leqslant x_2 \leqslant +1.$$

Нумерация точек плана показана на фиг. 6.3,*а*.

В табл. 6.9 приведен план при $n = 3$, который близок к D-оптимальному плану для следующих моделей:

$$y(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_1^2 + a_5x_2^2, \qquad \delta^* = 0{,}055; \tag{6.44а}$$

$$y(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_1^2 + a_5x_2^2 + a_6x_1x_2 + a_7x_1x_3 + a_8x_2x_3, \qquad \delta^* = 0{,}1481. \tag{6.44б}$$

Таблица 6.8

Непрерывные D-оптимальные (или близкие к ним) планы для неполных квадратичных полиномов на квадрате

	Виды модели	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	δ
1	$\eta(x, a)=a_0+a_1x_1+a_2x_2+a_3x_1^2+a_4x_1x_2$	1/6	1/6	1/6	1/6	0	0	1/6	0	1/6	0
2	$\eta(x, a)=a_0+a_1x_1+a_2x_2+a_3x_1^2+a_4x_2^2$	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	0
3	$\eta(x, a)=a_0+a_1x_1^2+a_2x_2^2+a_3x_1x_2$	0,2353	0,17647	0	0	0,17647	0,17647	0,17647	0	0	0,00213

Таблица 6.9

Планы на кубе для моделей вида (6.44а) и (6.44б), близкие к D-оптимальным

i	x_1^i	x_2^i	x_3^i	l_i
1	1	1	1	1/16
2	1	1	−1	1/16
3	1	−1	1	1/16
4	1	−1	−1	1/16
5	−1	1	1	1/16
6	−1	1	−1	1/16
7	−1	−1	1	1/16
8	−1	−1	−1	1/16
9	0	1	1	1/16
10	0	1	−1	1/16
11	0	−1	1	1/16
12	0	−1	−1	1/16
13	1	0	1	1/16
14	1	0	−1	1/16
15	−1	0	1	1/16
16	−1	0	−1	1/16

6.8. Непрерывные D-оптимальные планы для кубической регрессии на гиперкубе

В табл. 6.10 приведены некоторые непрерывные D-оптимальные или почти D-оптимальные планы для ряда уравнений, содержащих переменные в третьей степени. Индексы коэффициентов

Таблица 6.10

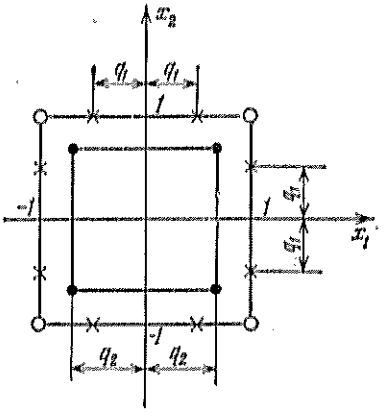
Непрерывные D-оптимальные планы для кубических моделей на гиперкубе

	Вид модели	q_1	q_2	Вершины большого куба		Точки на ребрах		Вершины малого куба		r	δ
				α	r_1	β	r_2	γ	r_3		
1	$y(a,x)=a_0+\sum_{i=1}^{2}a_ix_i+\sum_{i=1}^{2}\sum_{j=i}^{2}a_{ij}x_ix_j+\sum_{i=1}^{2}\sum_{j=i}^{2}\sum_{k=j}^{2}a_{ijk}x_ix_jx_k$	0,354	0,476	0,367	4	0,46104	8	0,17192	4	16	$\delta^*=0{,}00488$
2	$y(a,x)=a_0+\sum_{i=1}^{2}a_ix_i+\sum_{i=1}^{2}a_{ii}x_i^2+\sum_{i=1}^{2}a_{iii}x_i^3$	0,445	0,445	0,25	4	0,5	8	0,25	4	16	$\delta=0$
3	$y(a,x)=a_0+\sum_{i=1}^{3}a_ix_i+\sum_{i=1}^{3}a_{ii}x_i^2+\sum_{i=1}^{3}a_{iii}x_i^3$	0,445	0,445	—	—	0,75	24	0,25	8	32	$\delta=0$
4	$y(a,x)=a_0+\sum_{i=1}^{3}a_ix_i+\sum_{i=1}^{3}\sum_{j=i}^{3}a_{ij}x_ix_j+\sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{k=j}^{3}a_{ijk}x_ix_jx_k$	0,5	0,5	0,2369	8	0,6305	24	0,1336	8	40	$\delta^*=0{,}0065$
5	$y(a,x)=a_0+\sum_{i=1}^{4}a_ix_i+\sum_{i=1}^{4}\sum_{j=i}^{4}a_{ij}x_ix_j+\sum_{i=1}^{4}\sum_{j=i}^{4}\sum_{k=j}^{4}a_{ijk}x_ix_jx_k$	0,5	0,5	0,1666	16	0,6666	64	0,1666	16	96	$\delta^*=0{,}07228$

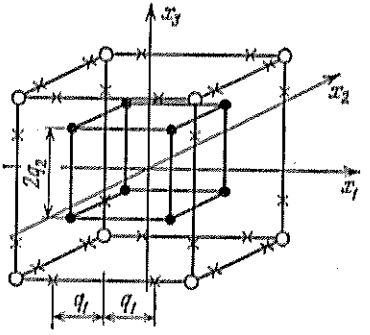
в табл. 6.10 представляют собой перечень номеров независимых переменных, образующих соответствующую функцию, например коэффициент при функции $x_i x_j$ обозначается a_{ij}, коэффициент при $x_i x_j^2$ обозначается a_{ijj} и т. д. Планы, представленные в табл. 6.10, были получены с помощью описанной выше вычислительной процедуры [15, 16]. Область планирования здесь — гиперкуб, ограниченный неравенствами

$$-1 \leqslant x_i \leqslant +1, \quad i = 1, 2, \ldots, n.$$

На фиг. 6.7 и фиг. 6.8 показаны точки планов для двумерного и трехмерного случаев соответственно. Здесь вершины большого куба с координатами, принимающими значения ± 1, от-



Фиг. 6.7. Точки непрерывного D-оптимального плана для кубических моделей на квадрате ($n = 2$).



Фиг. 6.8. Точки непрерывного D-оптимального плана для кубических моделей на кубе ($n = 3$).

мечены кружками; частота множества этих точек в табл. 6.10 обозначена буквой α. Точки, лежащие на ребрах куба на расстоянии q_1 от середины ребра, на рисунках отмечены крестиками; частота этого множества обозначена через β. Наконец, вершины малого куба с координатами $\pm q_2$ отмечены зачерненными кружками; их частота обозначена через γ.

6.9. О применении непрерывных D-оптимальных планов

Как уже выше отмечалось, непрерывный план не связан с определенным количеством наблюдений, а лишь задает соотношение числа опытов в различных точках области Ω_x. Сформулированная же в разд. 6.2 задача планирования эксперимента требует нахождения способа наилучшего размещения заданного числа N экспериментов. В некоторых случаях эта задача может

быть достаточно просто решена с помощью известного непрерывного D-оптимального плана. Возможность получения таким путем хорошего решения зависит прежде всего от соотношения числа опытов N и числа точек непрерывного плана r, а также от соотношения максимальной $l_{\text{макс}}$ и минимальной $l_{\text{мин}}$ частот точек непрерывного плана. Очевидно, что наиболее благоприятные условия имеют место, когда $N \gg r$ и $l_{\text{макс}} = l_{\text{мин}}$.

Рассмотрим несколько типовых ситуаций, возникающих при использовании непрерывных D-оптимальных планов для получения точного плана при заданном N.

а) $N = mr$, $m = 1, 2, \ldots$; $l_{\text{макс}} = l_{\text{мин}}$.

Это наиболее простой случай. Так как $l_{\text{макс}} = l_{\text{мин}}$, то $l_1 = l_2 = \ldots = l_r = 1/r$. Точный план, определяемый при этих условиях с помощью непрерывного D-оптимального плана, является D-оптимальным планом. Количество наблюдений в точке x^i этого плана равно

$$h_i = Nl_i = N\frac{1}{r} = mr\frac{1}{r} = m,$$

где l_i — частота точки x^i непрерывного плана.

Пример 6.10

Пусть требуется найти наилучший с точки зрения критерия D-оптимальности способ расположения $N = 32$ наблюдений в области

$$-1 \leqslant x_i \leqslant +1, \quad i = 1, 2,$$

для нахождения оценок коэффициентов уравнения вида

$$y(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1^3 + a_6x_2^3.$$

Воспользуемся непрерывным D-оптимальным планом, приведенным в табл. 6.10 (план номер 2). Точки этого плана показаны на фиг. 6.7. Из табл. 6.10 следует, что частоты во всех $r = 16$ точках плана одинаковы и равны $l_i = 1/16$, $i = 1 \div 16$. Так как $N = 32 = 2r$, то в каждой из точек плана фиг. 6.7 следует провести по два наблюдения. Полученный таким образом точный план является D-оптимальным планом.

б) $r \ll N$.

В этом случае можно ожидать, что с помощью непрерывного плана можно получить точный план, достаточно близкий к D-оптимальному. Число наблюдений h_i в точке x^i определяется путем округления произведения Nl_i до ближайшего целого числа.

Пример 6.11

Пусть требуется найти наилучший с точки зрения критерия D-оптимальности план эксперимента, содержащего около 100 опы-

тов, для нахождения оценок коэффициентов уравнения вида

$$y(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2.$$

Область планирования — квадрат:

$$-1 \leqslant x_i \leqslant +1, \quad i = 1, 2.$$

Воспользуемся непрерывным D-оптимальным планом для $n = 2$, приведенным в табл. 6.5. Точки этого плана показаны на фиг. 6.3,*а*. Из табл. 6.5 следует, что

$$l_1 = l_2 = l_3 = l_4 = \frac{0{,}5832}{4} = 0{,}14805,$$

$$l_5 = 0{,}0962,$$

$$l_6 = l_7 = l_8 = l_9 = \frac{0{,}3206}{4} = 0{,}08015.$$

Найдем количества наблюдений в точках точного плана

$$h_1 = h_2 = h_3 = h_4 = 100 \times 0{,}14805 \approx 15,$$

$$h_5 = 100 \times 0{,}0962 \approx 10,$$

$$h_6 = h_7 = h_8 = h_9 = 100 \times 0{,}08015 \approx 8.$$

Полученный таким образом точный план содержит 102 опыта. Можно ожидать, что этот план будет близок по своим свойствам к точному D-оптимальному плану при $N = 102$.

в) $k+1 \leqslant N < r$.

Это наиболее неприятный случай, хотя в практических задачах ввиду большого количества r точек непрерывных D-оптимальных планов (см., например, табл. 6.4 и 6.5 при $n \geqslant 4$) подобная ситуация возникает довольно часто. Рассмотрим один из возможных способов построения точного плана в этой ситуации [17, 18]. Предположим, что непрерывный D-оптимальный план концентрируется в точках некоторых множеств $E_1, E_2, \ldots, E_\mu$ с частотами множеств $\alpha_1, \alpha_2, \ldots, \alpha_\mu$, причем $N \geqslant \mu$. Тогда с помощью непрерывного плана можно найти числа h^i наблюдений для каждого множества точек, округлив до ближайшего целого числа величину $N\alpha_i$, $i = 1, 2, \ldots, \mu$. Допустим теперь, что множество E_i состоит из r_i точек, в которых требуется расположить h^i опытов. Если $h^i = mr_i + r_i'$, где $m = 0, 1, 2, \ldots; r_i' < r_i$, то в каждой точке следует поставить по m опытов, а оставшиеся r_i' экспериментов случайным образом распределить среди точек множества E_i. Очевидно, что такой метод приводит к неоднозначному решению, причем получаемые планы могут существенно отличаться по своим свойствам. Можно было бы поставить задачу выбора наилучшего плана с точки зрения величины определителя информационной матрицы, варьируя распределение r_i' опытов, $i = 1, 2, \ldots, \mu$, по точкам соответствующих множеств.

Эта задача, хотя и проще задачи построения численным методом точного D-оптимального плана для заданного числа опытов, однако все-таки достаточно трудоемка и в большинстве случаев может быть решена лишь с помощью цифровых вычислительных машин. В связи с этим практический интерес для планирования эксперимента, в случаях когда $N < r$, представляют близкие по своим свойствам к D-оптимальным точные планы, найденные тем или иным способом для некоторых видов моделей и ряда значений N [18]. Эти планы мы рассмотрим в разд. 6.11.

Пример 6.12

Пусть требуется найти наилучший с точки зрения критерия D-оптимальности план эксперимента, содержащего $N = 18$ опытов, для оценки коэффициентов модели вида

$$y(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_1^2 + a_6x_2^2 + a_7x_3^2 + a_8x_4^2 + a_9x_1x_2 + a_{10}x_1x_3 + a_{11}x_1x_4 + a_{12}x_2x_3 + a_{13}x_2x_4 + a_{14}x_3x_4, \qquad (6.45)$$

если область планирования в пространстве размерности $n = 4$ определяется неравенствами

$$-1 \leqslant x_i \leqslant +1, \quad i = 1, \ldots, 4.$$

Число оцениваемых коэффициентов здесь равно

$$k + 1 = 15.$$

Воспользуемся планом Коно (см. табл. 6.5), который для $n = 4$ концентрируется в $r = 49$ точках. Эти точки распределяются по множествам E_0 (вершины куба), E_1 (середины ребер) и E_0 (центр куба) с частотами $\alpha = 0{,}4506$, $\beta = 0{,}5021$ и $\xi = 0{,}04738$ соответственно. Найдем числа h^l точек каждого множества в точном плане:

$$h^0 = N\alpha = 18 \times 0{,}4506 \approx 8,$$
$$h^1 = N\beta = 18 \times 0{,}5021 \approx 9,$$
$$h^0 = N\xi = 18 \times 0{,}04738 \approx 1.$$

Восемь экспериментов множества E_0 разумно расположить так, чтобы они образовали регулярную полуреплику полного факторного эксперимента. Число регулярных полурепликa для $n = 4$ равно четырем. Девять экспериментов множества E_1 можно расположить в 32 серединах ребер C_{32}^9 различными способами (здесь C_{32}^9 — число сочетаний из 32 по 9). Очевидно, что общее число различных вариантов построения плана здесь превосходит миллион. В табл. 6.11 приведен наилучший план, полученный в работе [18] путем просмотра небольшого числа вариантов.

Таблица 6.11

План эксперимента (пример 6.12)

l	x'_1	x'_2	x'_3	x'_4	*Вершина куба*
1	1	−1	−1	−1	
2	−1	1	−1	−1	
3	−1	−1	1	−1	
4	1	1	1	−1	*Полуреплика полного*
5	1	−1	−1	1	*факторного плана*
6	−1	1	−1	1	
7	−1	−1	1	1	
8	1	1	1	1	
9	0	1	−1	1	
10	−1	0	−1	−1	
11	1	0	−1	1	
12	−1	1	−1	0	
13	−1	0	1	1	*Середины ребер*
14	0	1	1	−1	
15	1	−1	1	0	
16	1	−1	0	1	
17	0	−1	1	−1	
18	0	0	0	0	*Центр плана*

Определитель информационной матрицы нормированного плана равен 10^{-7} [18], что более чем на два порядка меньше определителя непрерывного D-оптимального плана (величина этого определителя $2{,}2 \cdot 10^{-5}$ [18]). Для сравнения можно привести величину определителя нормированной информационной матрицы ортогонального плана при $n = 4$. Эта величина равна $0{,}18 \cdot 10^{-10}$. Таким образом, полученный здесь план с точки зрения критерия D-оптимальности превосходит ортогональный план.

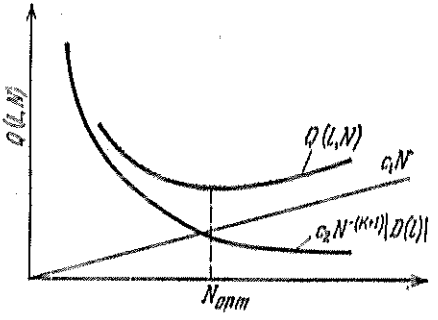
6.10. Определение оптимального объема выборки

До сих пор задача планирования эксперимента в этом разделе рассматривалась как задача наилучшего с точки зрения критерия D-оптимальности расположения заданного числа N опытов в области Ω_x. При такой постановке стоимость каждого эксперимента не влияет на выбор плана исследования. Все потери, связанные с качеством исследования, возникают лишь за счет неточного определения коэффициентов модели, число же N определяется максимально допустимым числом экспериментов. Иногда, однако, стоимость каждого опыта следует включать в критерий планирования эксперимента. Тогда возникает задача выбора оптимального числа опытов N.

Введем следующий критерий планирования:

$$Q(l, N) = c_1 N + c_2 |(F_N' F_N)^{-1}| = c_1 N + c_2 N^{-(k+1)} |D(l)|. \quad (6.46)$$

Здесь c_1 и c_2 — некоторые неотрицательные константы, определяющие соответственно стоимость отдельного опыта и удельные потери от неточного знания коэффициентов. Мы полагаем, что стоимость отдельного опыта не зависит от условий его проведения, а потери от неточного знания коэффициентов пропорциональны величине определителя дисперсионной матрицы, или, иными словами, объему эллипсоида рассеяния. В (6.46) $D(l)$ — норми-



Фиг. 6.9. Оптимальный выбор числа экспериментов по критерию (6.46).

рованная дисперсионная матрица плана X, $k+1$ — число оцениваемых коэффициентов, а F_N — матрица функций независимых переменных размером $N \times (k+1)$. Величина критерия (6.46) зависит от числа экспериментов N и от выбора плана l. Первое слагаемое в (6.46) растет с увеличением N, а второе убывает (фиг. 6.9). Это значит, что при некотором $N = N_{опт}$ будет иметь место минимум величины $Q(l, N)$. Так как при всяком $N = N^*$ величина $Q(l, N^*)$ является строго возрастающей функцией от $|D(l)|$, то при всяком N наилучшим будет план, минимизирующий определитель нормированной дисперсионной матрицы, т. е. D-оптимальный план. Положим, что $N_{опт}$ велико, так что при N, близких к $N_{опт}$, могут быть найдены планы, достаточно близкие к непрерывному D-оптимальному, т. е. имеющие величину определителя $|D(l)| \approx |D(l^*)|$, где $|D(l^*)|$ — определитель дисперсионной матрицы D-оптимального плана. Тогда для величины $Q(l, N)$ можно записать

$$Q(l, N) \approx Q(l^*, N) = c_1 N + c_2 N^{-(k+1)} |D(l^*)|. \quad (6.47)$$

Из (6.47) нетрудно найти выражение для $N_{опт}$, приравнивая нулю частную производную по N:

$$\left.\frac{\partial Q(l^*, N)}{\partial N}\right|_{N_{опт}} = c_1 - c_2 (k+1) N_{опт}^{-(k+2)} |D(l^*)| = 0. \quad (6.48)$$

Отсюда получаем

$$N_{\text{опт}} = \left(\frac{c_2 (k+1) \mid D(l^*) \mid}{c_1} \right)^{1/(k+2)}. \qquad (6.49)$$

Выражение (6.49) будет справедливо лишь при достаточно больших $N_{\text{опт}}$, т. е. когда

$$c_1 \ll c_2 (k+1) \mid D(l^*) \mid. \qquad (6.50)$$

Приведенные здесь выражения могут быть использованы для определения нижней грани числа опытов в общем случае, когда условие (6.50) не имеет места. Действительно, так как определитель дисперсионной матрицы непрерывного D-оптимального плана l^* не превосходит величины определителя нормированной дисперсионной матрицы произвольного плана l, то в общем случае имеем

$$N_{\text{опт}} \geqslant \left(\frac{c_2 (k+1) \mid D(l^*) \mid}{c_1} \right)^{1/(k+2)}. \qquad (6.51)$$

Заметим, что величина определителя дисперсионной матрицы для непрерывных D-оптимальных планов может быть легко подсчитана и часто приводится при описании планов (см. табл. 6.1).

Пример 6.13

Необходимо найти план проведения экспериментов, имеющих целью нахождение оценок коэффициентов уравнения вида

$$y(a, x) = a_0 + a_1 x + a_2 x^2.$$

Область планирования Ω_x ограничена неравенством

$$-1 \leqslant x \leqslant +1.$$

Критерий оптимальности плана

$$Q(l, N) = 0{,}001N + \mid (F_N' F_N)^{-1} \mid.$$

Здесь $c_1 = 0{,}001$ и $c_2 = 1$.

Непрерывный D-оптимальный план для данной задачи (см. табл. 6.1) сосредоточен в точках с координатами $+1$, 0, -1 с равными частотами. Определитель нормированной дисперсионной матрицы (см. табл. 6.1) равен

$$\mid D(l^*) \mid = 6{,}749,$$

а число оцениваемых коэффициентов $k+1 = 3$. Подставляя указанные значения в выражение (6.49) и имея в виду, что в данном случае условие (6.50) имеет место, получаем

$$N_{\text{опт}} = \left(\frac{c_2 (k+1) \mid D(l^*) \mid}{c_1} \right)^{1/(k+2)} = \left(\frac{3 \times 6{,}749}{0{,}001} \right)^{1/4} \approx 12.$$

Оптимальный план X (см. разд. 6.9, п. „а") будет включать точки с координатами -1, 0, $+1$, повторенные по четыре раза.

6.11. Точные планы, близкие к D-оптимальным

Наибольший практический интерес представляют точные D-оптимальные планы с малым числом опытов N. Как уже указывалось выше, в случае когда известный непрерывный план сосредоточен в большом числе точек r, так что $r \geqslant N$, найти точный план, близкий к D-оптимальному, простыми средствами не удается. Решение здесь может быть получено с помощью цифровых вычислительных машин. Результаты исследований в этом направлении изложены в работах [17, 18].

Точный план для заданного числа опытов N в принципе может быть найден с помощью тех или иных численных методов решения экстремальных задач [5] и без знания соответствующего непрерывного D-оптимального плана. Задача, однако, существенно упрощается, если точный план составляется только из точек непрерывного D-оптимального плана, что приводит, естественно, к некоторой потере оптимальности точного плана. Мы здесь приведем ряд точных планов, близких по своим свойствам к D-оптимальным и содержащих точки непрерывных D-оптимальных планов. Эти планы строились с помощью цифровых вычислительных машин для случаев, когда $N < r$, где r — число точек непрерывного плана, для квадратичных моделей вида

$$y(a, x) = a_0 + a_1x_1 + \ldots + a_nx_n + a_{n+1}x_1^2 + \ldots + a_{2n}x_n^2 + a_{2n+1}x_1x_2 + \ldots + a_kx_{n-1}x_n. \quad (6.52)$$

Область планирования — гиперкуб:

$$-1 \leqslant x_i \leqslant +1, \qquad i = 1, 2, \ldots, n.$$

В табл. 6.12 приведены точные планы для $n = 2, 3, 4$, близкие к D-оптимальным при числе экспериментов $N = k + 1$.

Заметим, что точные планы с числом опытов, равным числу оцениваемых коэффициентов, называются насыщенными. Эти планы были получены с помощью цифровых вычислительных машин путем поиска наилучшего подмножества из точек, входящих в непрерывные планы Кено (см. табл. 6.5 и фиг. 6.5). Работа была проведена в Проблемной лаборатории автоматики Московского энергетического института. В табл. 6.13 приведены основные характеристики этих планов — величина определителя нормированной информационной матрицы L и максимальное по области Ω_x значение квадратичной формы:

$$\max_{x \in \Omega_x} f'(x) D f(x) = d_{\text{макс}}, \quad (6.53)$$

где D — дисперсионная матрица нормированного плана. Величины $d_{\text{макс}}$, найденные численными методами, помечены звездоч-

Таблица 6.12

Насыщенные планы, близкие к D-оптимальным, для квадратичных моделей

а	*i*	x_1^i	x_2^i
	1	−1	−1
	2	−1	+1
	3	+1	−1
	4	+1	+1
	5	0	−1
	6	0	0

б	*i*	x_1^i	x_2^i	x_3^i
	1	1	−1	1
	2	1	1	−1
	3	1	−1	−1
	4	1	1	1
	5	−1	1	0
	6	−1	0	1
	7	−1	0	−1
	8	−1	−1	0
	9	0	−1	1
	10	0	0	0

в	*i*	x_1^i	x_2^i	x_3^i	x_4^i
	1	−1	1	1	1
	2	1	−1	1	1
	3	1	1	−1	1
	4	1	1	−1	−1
	5	1	−1	1	−1
	6	−1	−1	−1	1
	7	−1	−1	−1	−1
	8	0	1	1	−1
	9	−1	0	1	−1
	10	−1	1	0	−1
	11	1	−1	−1	0
	12	−1	1	−1	0
	13	−1	−1	1	0
	14	1	1	1	0
	15	0	0	0	0

кой. Для сравнения приведены аналогичные характеристики для непрерывных D-оптимальных планов. Заметим, что различие в указанных показателях для приведенных насыщенных и непрерывных D-оптимальных планов является закономерным, ибо характеристики точного плана всегда хуже соответствующих характеристик непрерывного плана. Однако наблюдаемые здесь

Таблица 6.13

Свойства насыщенных планов (табл. 6.12)

n	План	Число точек плана	$d_{\text{макс}}$		$\lvert L \rvert$	
			насыщенные планы	непрерывные D-оптимальные планы	насыщенные планы	непрерывные D-оптимальные планы
2	Табл. 6.12, а	6	16,77 *	6	$0{,}549 \cdot 10^{-1}$	$0{,}114 \cdot 10^{-1}$
3	Табл. 6.12, б	10	53,9 *	10	$0{,}104 \cdot 10^{-3}$	$0{,}518 \cdot 10^{-3}$
4	Табл. 6.12, в	15	33,15 *	15	$0{,}267 \cdot 10^{-5}$	$0{,}22 \cdot 10^{-4}$

Таблица 6.14

Планы типа B_n

i	x_1^i	x_2^i	x_3^i	x_4^i	
1	+1	+1	+1	+1	
2	−1	+1	+1	+1	
3	+1	−1	+1	+1	
4	−1	−1	+1	+1	
5	+1	+1	−1	+1	
6	−1	+1	−1	+1	
7	+1	−1	−1	+1	
8	−1	−1	−1	+1	*Вершины гиперкуба (полный факторный план)*
9	+1	+1	+1	−1	
10	−1	+1	+1	−1	
11	+1	−1	+1	−1	
12	−1	−1	+1	−1	
13	+1	+1	−1	−1	
14	−1	+1	−1	−1	
15	+1	−1	−1	−1	
16	−1	−1	−1	−1	
17	+1	0	0	0	
18	−1	0	0	0	
19	0	+1	0	0	
20	0	−1	0	0	*Центры (n−1)-мерных граней (звездные точки)*
21	0	0	+1	0	
22	0	0	−1	0	
23	0	0	0	+1	
24	0	0	0	−1	

различия относительно невелики, а это дает основания предполагать, что приведенные насыщенные планы близки к точным D-оптимальным насыщенным планам.

Приведем здесь еще ряд точных планов, близких к D-оптимальным, для квадратичных моделей вида (6.52), построенных на гиперкубе и содержащих относительно небольшое число опытов. Эти планы, называемые планами типа B_n, были исследованы в работах [17, 18, 20] для размерностей $n=4$, 5 и 6.

Таблица 6.15

Свойства планов для квадратичных моделей

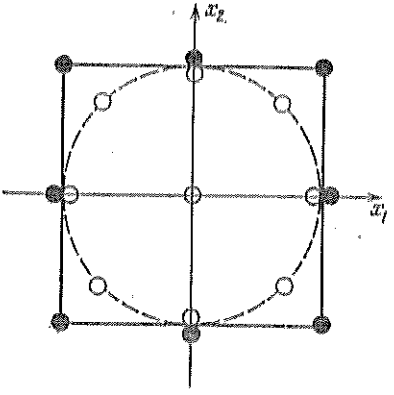
n	Тип плана	Число точек	$d_{\text{макс}}$	$\lvert L\rvert$
4	Непрерывный D-оптимальный план	—	15	$0{,}22\cdot10^{-4}$
	B_4	24	18,5	$0{,}8\cdot10^{-5}$
	Ортогональный план	25	—	$0{,}18\cdot10^{-10}$
	Ротатабельный план Бокса	31	267,2	$0{,}18\cdot10^{-16}$
5	Непрерывный D-оптимальный план	—	21	$0{,}63\cdot10^{-6}$
	B_5	42	34,2	$0{,}68\cdot10^{-7}$
	Ортогональный план (с ядром 2^{5-1})	43	149,2	$0{,}4\cdot10^{-18}$
	Ротатабельный план (с ядром 2^{5-1})	52	692,4	$0{,}28\cdot10^{-28}$
6	Непрерывный D-оптимальный план	—	28	$0{,}15\cdot10^{-7}$
	B_6	76	66,5	$0{,}23\cdot10^{-9}$
	Ортогональный план (с ядром 2^{6-1})	77	312,1	$0{,}22\cdot10^{-28}$
	Ротатабельный план (с ядром 2^{6-1})	91	1726,4	$0{,}59\cdot10^{-45}$

Планы B_n содержат вершины n-мерного куба с координатами ±1. Эти точки образуют полный факторный эксперимент. Число их 2^n. Кроме того, в план B_n входят центры $(n-1)$-мерных граней. Это точки с координатами 0, 0, ..., ±1, 0, ..., 0. По аналогии с ортогональными и ротатабельными планами эти точки можно называть звездными с величиной плеча, равной единице. Число таких точек равно $2n$. Общее число точек плана $N=2^n+2n$. В табл. 6.14 приведен план B_4, содержащий $N=2^4+2\times4=24$ точки.

В табл. 6.15, заимствованной нами из работы [18], указаны основные характеристики планов B_n [число наблюдений N, $d_{\text{макс}}$ в соответствии с (6.53) и $|L|$ для $n=4$, 5, 6]. Для сравнения даны соответствующие характеристики непрерывных D-оптимальных планов, ортогональных планов и ротатабельных планов Бокса (см. разд. 5.6). Видно, что по приведенным показателям планы B_n близки к D-оптимальным и существенно превосходят ортогональные и ротатабельные планы Бокса.

Чтобы объяснить причину подобных расхождений в свойствах планов B_n и ортогональных и ротатабельных планов, обратимся к фиг. 6.10. Здесь изображены для размерности $n=2$ точки плана типа B_2 (зачерненные кружки) и ротатабельного плана Бокса для случая, когда область планирования Ω_x—квадрат:

$$-1 \leqslant x_i \leqslant +1, \qquad i=1,\ 2. \tag{6.54}$$



Фиг. 6.10. К сравнению планов типа B_n и ротатабельных планов Бокса.

Точки ротатабельного плана лежат на окружности, вписанной в квадрат, ограничивающий область планирования. При этом вершины квадрата не попадают в ротатабельный план. Если бы мы поместили опыты в точки с координатами ± 1, то для получения ротатабельного плана следовало бы поставить эксперименты в точках с координатами $\pm 1/\sqrt{2}$; но это невозможно, так как область планирования ограничена неравенствами (6.54). Таким образом, ротатабельный план использует не всю область планирования эксперимента, а лишь часть ее, ограниченную окружностью, вписанной в область Ω_x. Планы типа B_n лучше используют всю область Ω_x, ибо они включают вершины квадрата. Именно за счет этого и наблюдается различие в свойствах рассматриваемых планов.

Приведем здесь еще дисперсионные матрицы C для планов B_n при $n=4$, 5, 6, знание которых может быть полезным при вычислении оценок коэффициентов уравнения модели вида (6.52) [20]:

$$C=(F'F)^{-1}=\begin{bmatrix} a & 0 & -bH'_{n1} & 0 \\ 0 & g_1 I_n & 0 & 0 \\ -bH_{n1} & 0 & cI_n-dH_{nn} & 0 \\ 0 & 0 & 0 & g_2 I_{\binom{n}{2}} \end{bmatrix}. \tag{6.55}$$

Здесь H_{n1} — вектор-столбец размерности n (H'_{n1} — строка размерности n); H_{nn} — квадратная матрица порядка n. Элементами H_{n1} и H_{nn} являются единицы. I_n — единичная матрица порядка n. a, b, c, d, g_1 и g_2 — скаляры, значения которых для $n=4, 5, 6$ приведены в табл. 6.16.

Таблица 6.16

Параметры дисперсионных матриц планов B_4, B_5 и B_6

n	a	b	c	d	g_1	g_2
4	0,2291 (6)	0,0625	0,5	0,1041 (6)	0,0555556	0,0625
5	0,158206	0,0332	0,5	0,0917969	0,0291118	0,03125
6	0,120625	0,020625	0,5	0,079375	0,0151515	0,015625

Пример 6.14

Этот пример имеет целью проиллюстрировать процедуру вычисления коэффициентов уравнения регрессии в случае, когда эксперимент проведен по плану типа B_n.

Положим, что экспериментальным путем требуется построить оценки коэффициентов уравнения, связывающего концентрацию продукта на выходе некоторого химико-технологического процесса со следующими параметрами, определяющими режим процесса: T_1 — температурой на входе, T_2 — температурой на выходе, V — количеством сырья, поступающего на вход процесса в единицу времени, P — давлением. Пусть уравнение регрессии имеет вид

$$y = c_0 + c_1T_1 + c_2T_2 + c_3V + c_4P + c_5T_1^2 + c_6T_2^2 + c_7V^2 + c_8P^2 + c_9T_1T_2 + c_{10}T_1V + c_{11}T_1P + c_{12}T_2V + c_{13}T_2P + c_{14}PV, \quad (6.56)$$

а область изменения переменных T_1, T_2, P и V ограничена неравенствами

$$677\ ^\circ\text{C} \leqslant T_1 \leqslant 697\ ^\circ\text{C},$$
$$788\ ^\circ\text{C} \leqslant T_2 \leqslant 808\ ^\circ\text{C},$$
$$8350\ \text{кг/ч} \leqslant V \leqslant 9150\ \text{кг/ч},$$
$$323\ \text{атм} \leqslant P \leqslant 337\ \text{атм}.$$

Перейдем к безразмерным переменным x_1, x_2, x_3, x_4:

$$x_1 = \frac{T_1 - 687}{10},$$
$$x_2 = \frac{T_2 - 798}{10},$$
$$x_3 = \frac{V - 8750}{400}, \quad (6.57)$$
$$x_4 = \frac{P - 330}{7}.$$

Таблица 6.17

Матрица F и результаты экспериментов (пример 6.14)

j	x_0^j	x_1^j	x_2^j	x_3^j	x_4^j	$(x_1^j)^2$	$(x_2^j)^2$	$(x_3^j)^2$	$(x_4^j)^2$	$x_1^j x_2^j$	$x_1^j x_3^j$	$x_1^j x_4^j$	$x_2^j x_3^j$	$x_2^j x_4^j$	$x_3^j x_4^j$	$\bar{y}^j$
1	1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	0,1976
2	1	−1	+1	+1	+1	+1	+1	+1	+1	−1	−1	−1	+1	+1	+1	0,1979
3	1	+1	−1	+1	+1	+1	+1	+1	+1	−1	+1	+1	−1	−1	+1	0,1980
4	1	−1	−1	+1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	+1	0,1979
5	1	+1	+1	−1	+1	+1	+1	+1	+1	−1	+1	−1	−1	+1	−1	0,2012
6	1	−1	+1	−1	+1	+1	+1	+1	+1	−1	+1	−1	−1	+1	−1	0,2017
7	1	+1	−1	−1	+1	+1	+1	+1	+1	−1	−1	+1	+1	−1	−1	0,2020
8	1	−1	−1	−1	+1	+1	+1	+1	+1	+1	+1	−1	+1	−1	−1	0,2023
9	1	+1	+1	+1	−1	+1	+1	+1	+1	+1	+1	−1	+1	−1	−1	0,1911
10	1	−1	+1	+1	−1	+1	+1	+1	+1	−1	−1	+1	+1	−1	−1	0,1913
11	1	+1	−1	+1	−1	+1	+1	+1	+1	−1	+1	−1	−1	+1	−1	0,1914
12	1	−1	−1	+1	−1	+1	+1	+1	+1	+1	−1	+1	−1	+1	−1	0,1911
13	1	+1	+1	−1	−1	+1	+1	+1	+1	+1	−1	−1	−1	−1	+1	0,1947
14	1	−1	+1	−1	−1	+1	+1	+1	+1	−1	+1	+1	−1	−1	+1	0,1952
15	1	+1	−1	−1	−1	+1	+1	+1	+1	−1	−1	−1	+1	+1	+1	0,1954
16	1	−1	−1	−1	−1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	0,1956
17	1	+1	0	0	0	+1	0	0	0	0	0	0	0	0	0	0,1982
18	1	−1	0	0	0	+1	0	0	0	0	0	0	0	0	0	0,1984
19	1	0	+1	0	0	0	+1	0	0	0	0	0	0	0	0	0,1980
20	1	0	−1	0	0	0	+1	0	0	0	0	0	0	0	0	0,1984
21	1	0	0	+1	0	0	0	+1	0	0	0	0	0	0	0	0,1964
22	1	0	0	−1	0	0	0	+1	0	0	0	0	0	0	0	0,2004
23	1	0	0	0	+1	0	0	0	+1	0	0	0	0	0	0	0,2017
24	1	0	0	0	−1	0	0	0	+1	0	0	0	0	0	0	0,1951

Таблица 6.18

Матрица $(F'F)^{-1}$ (пример 6.14)

0,229167	0	0	0	0	−0,0625	−0,0625	−0,0625	−0,0625	0	0	0	0	0	0
	0,0555556	0	0	0	0	0	0	0	0	0	0	0	0	0
		0,0555556	0	0	0	0	0	0	0	0	0	0	0	0
			0,0555556	0	0	0	0	0	0	0	0	0	0	0
				0,0555556	0	0	0	0	0	0	0	0	0	0
					0,39583	−0,104167	−0,104167	−0,104167	0	0	0	0	0	0
						0,39583	−0,104167	−0,104167	0	0	0	0	0	0
							0,39583	−0,104167	0	0	0	0	0	0
								0,39583	0	0	0	0	0	0
									0,0625	0	0	0	0	0
										0,0625	0	0	0	0
											0,0625	0	0	0
												0,0625	0	0
													0,0625	0
														0,0625

Область изменения переменных x_1, x_2, x_3, x_4 — гиперкуб с вершинами в точках с координатами ± 1. Уравнение регрессии запишем следующим образом:

$$y(x, a) = a_0 = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_1^2 + a_6x_2^2 + a_7x_3^2 + a_8x_4^2 + a_9x_1x_2 + a_{10}x_1x_3 + a_{11}x_1x_4 + a_{12}x_2x_3 + a_{13}x_2x_4 + a_{14}x_3x_4. \quad (6.58)$$

Эксперимент осуществим по плану B_4, показанному в табл. 6.17. Матрица F и результаты измерения выходной концентрации приведены в табл. 6.17. Умножая матрицу F на вектор-столбец $\hat{Y}$, получим

$$(F'\hat{Y})' = (4{,}717;\ -0{,}018;\ -0{,}0034;\ -0{,}0358;\ 0{,}0594;\ 3{,}541;\ 3{,}5408;\ 3{,}5412;\ 3{,}5412;\ -0{,}0014;\ 0{,}0014;\ -0{,}0004;\ 0{,}002;\ -0{,}0006;\ 0{,}0002).$$

Пользуясь выражением (6.55) и табл. 6.16, найдем матрицу $(F'F)^{-1}$ (табл. 6.18). Так как матрица $(F'F)^{-1}$ симметричная, то в таблице приведены лишь элементы, лежащие выше главной диагонали. Умножая матрицу $(F'F)^{-1}$ на $F'\hat{Y}$, найдем вектор-столбец $\hat{a}$ оценок коэффициентов. Полученное уравнение при этом имеет вид

$$\hat{y} = 0{,}1989 - 0{,}1\cdot 10^{-3}x_1 - 0{,}19\cdot 10^{-3}x_2 - 0{,}1998\cdot 10^{-3}x_3 + 0{,}33\cdot 10^{-3}x_4 - 0{,}625\cdot 10^{-3}x_1^2 - 0{,}725\cdot 10^{-3}x_2^2 - 0{,}525\cdot 10^{-3}x_3^2 - 0{,}525\cdot 10^{-3}x_4^2 - 0{,}875\cdot 10^{-4}x_1x_2 + 0{,}875\cdot 10^{-4}x_1x_3 - 0{,}25\cdot 10^{-4}x_1x_4 + 0{,}125\cdot 10^{-3}x_2x_3 - 0{,}375\cdot 10^{-4}x_2x_4 + 0{,}125\cdot 10^{-4}x_3x_4. \quad (6.59)$$

Дальнейший анализ полученных результатов (вычисление остаточной дисперсии, проверка значимости коэффициентов и пр.) может быть осуществлен так, как описано в разд. 3.3. Заметим еще, что, пользуясь полученным уравнением (6.59), нетрудно перейти к исходным переменным T_1, T_2, V и P. Для этого достаточно вместо x_1, x_2, x_3, x_4 в уравнение (6.59) подставить выражения (6.57) и осуществить простейшие преобразования.

ЛИТЕРАТУРА

1. Kiefer J., Optimum experimental designs, *J. Roy. Stat. Soc.*, Ser. B, **21**, 272—319 (1959).
2. Kiefer J., Optimum designs in regression problems, *Ann. Math. Stat.*, **30**, 271—294 (1959).
3. Kiefer J., Optimum designs in regression problems, II, *Ann. Math. Stat.*, **32**, 298—325 (1961).
4. Голикова Т. И., Микешина Н. Г., Свойства D-оптимальных планов и методы их построения, см. сб. „Новые идеи в планировании эксперимента", изд-во „Наука", М., 1969.

5. Федоров В. В., Теория оптимального эксперимента, изд-во „Наука", М., 1971.
6. Налимов В. В., Теория эксперимента, изд-во „Наука", М., 1971.
7. Вучков И. Н., Круг Г. К., D-оптимальные экспериментальные планы, Труды МЭИ, вып. 68, 5—19, 1969.
8. Box G., Hunter W., Sequetial design of experiments for nonlinear models, Proc. IBM Sci. Comput. Symp. on Statistic, Oct. 1963, New York, 1965.
9. Федоров В. В., Последовательные методы планирования эксперимента при изучении механизма явлений, см. сб. „Новые идеи в планировании эксперимента", изд-во „Наука", М., 1969.
10. Wilks S. S., Mathematical Statistics, New York, 1962; русский перевод: Уилкс С., Математическая статистика, изд-во „Наука", М., 1967.
11. Kiefer J., Wolfowitz J., The equivalence of two extremum problems, *Canad. J. Math.*, **12**, 363—366 (1960).
12. Guest P. G., The spacing of observations in polynomial regression, *Ann. Math. Stat.*, **29**, 294—299 (1958).
13. Kono K., Optimum design for quadratic regression on k-cube, *Mem. Fac. Sci. Kyushu Univ.*, Ser A, **16**, 114—122 (1962).
14. Вучков И. Н., Круг Г. К., Применение метода непрерывного планирования для получения D-оптимальных планов, см. сб. „Проблемы планирования эксперимента", изд-во „Наука", М., 1969.
15. Вучков И. Н., Ермуратский П. В., Круг Г. К., Лецкий Э. К., Никифорова Е. С., Об одном алгоритме получения D-оптимальных планов, Труды МЭИ, вып. 68, 1969.
16. Вучков И. Н., Круг Г. К., Лецкий Э. К., Никифорова Е. Н., D-оптимальные планы для кубической регрессии, *Заводск. лаб.*, № 7 (1971).
17. Голикова Т. И., Микешина Н. Г., Налимов В. В., Веселая Г. Н., Построение на кубе планов второго порядка, близких к D-оптимальным, *Заводск. лаб.* **33**, № 7 (1967).
18. Андрукович П. Ф., Голикова Т. И., Костина С. Г., Планы второго порядка на гиперкубе, близкие по свойствам к D-оптимальным, см. сб. „Новые идеи в планировании эксперимента", изд-во „Наука", М., 1969.
19. Атрощенко В. И. и др., Курс технологии связанного азота, изд-во „Химия", М., 1969.
20. Горский В. Г., Бродский В. З., Регрессионный анализ при композиционном планировании второго порядка специального вида, АН СССР, Научный совет по комплексной проблеме „Кибернетика", Информационные материалы, 8 (45), М., 1970.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Дубова И. С., Пицхеладзе Т. В., Федоров В. В., Таблицы оптимальных планов. I (непрерывные D-оптимальные планы), МГУ, вып. 16, 1970.
2. Дубова И. С., Федоров В. В., Таблицы оптимальных планов. II (насыщенные D-оптимальные планы на кубе), МГУ, вып. 40, 1972.
3. Песочинский Л. Л., D-оптимальные и близкие к ним точные планы для квадратичной регрессии на кубе и шаре, МГУ, вып. 42, 1972.
4. Лецкий Э. К., Никифорова Е. С., Об использовании спектров непрерывных D-оптимальных планов при построении точных планов, *Заводск. лаб.*, № 5 (1974).
5. Божанов Е., Вучков И., Статистически методи за моделиране и оптимизиране на многофакторни обекти, „Техника", София, 1973.
6. Успенский А. Б., Федоров В. В., Вычислительные аспекты метода наименьших квадратов при анализе и планировании регрессионных экспериментов, МГУ, 1975.

7.
ВЫДЕЛЕНИЕ СУЩЕСТВЕННЫХ ФАКТОРОВ

7.1. Введение

Для правильного отображения объекта необходимо, чтобы его математическая модель включала все факторы, существенно влияющие на выходную величину. Отсутствие в модели хотя бы одного из существенных факторов может повлечь за собой ошибочную интерпретацию явлений, протекающих в объекте. В свою очередь это может явиться причиной серьезных ошибок в решениях, принимаемых с помощью модели. Например, расчетное положение оптимума окажется значительно удаленным от действительного и т. п. Кроме того, модели, в которые входят не все доминирующие факторы, как правило, являются неадекватными.

При исследовании сложных процессов, на которые воздействует большое число факторов, обычно заранее не известно, какие из этих факторов следует считать доминирующими. В сложных промышленных процессах, особенно в так называемых больших системах, число различных факторов может быть очень большим — от десятков до нескольких сотен. Степень же влияния этих факторов весьма различна; как правило, лишь небольшое число факторов оказывает существенное влияние на выходную величину, а остальные воздействуют на нее незначительно. Задача заключается в том, чтобы с помощью соответствующих методов выделить и идентифицировать существенные, доминирующие факторы на „шумовом фоне" всех остальных факторов.

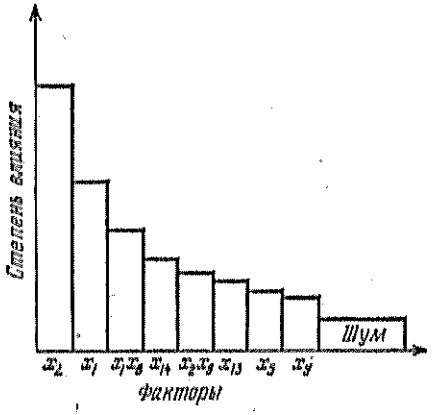
Цель рассмотренных в этом разделе методов состоит в получении грубых оценок воздействия факторов (фиг. 7.1) при возможно меньших затратах (минимальном числе экспериментов).

Обычные методы построения математических моделей, например регрессионный анализ, включающий проверку значимости параметров, при большом числе факторов оказываются неприемлемыми. Требуемые для их применения число экспериментов и затраты вычислительного времени на обработку данных становятся недопустимо большими. Те же трудности возникают при необходимости исследовать не только сами факторы, но и их всевозможные комбинации, а также при случайном поиске важных факторов. Поэтому были разработаны специальные методы, которые при выполнении некоторых предпосылок позволяют выявить существенные факторы с помощью небольшого числа

экспериментов и при небольших затратах вычислительного времени. К числу таких методов относятся, например, следующие:

1. Дисперсионный анализ, в основу которого положено предположение о том, что существенность некоторого (дискретного) фактора характеризуется его вкладом в дисперсию выходной величины.

2. Насыщенные дробные факторные планы, основанные на предположении о наличии лишь линейных эффектов и приводя-



Фиг. 7.1. Распределение факторов по степени их влияния на целевую величину.

щие к оценке существенности факторов по их вкладу в математическое ожидание выходной величины.

3. Насыщенные экспериментальные планы Плакетта—Бермана, для которых также предполагается наличие только линейных эффектов.

4. Метод случайного баланса, применяемый в предположении, что среди рассматриваемых факторов не все являются существенными. Благодаря этому можно применять сверхнасыщенное планирование, в котором число факторов больше числа экспериментов.

5. Опрос экспертов с целью ранжирования факторов по степени их влияния на выходную величину или сочетание опроса экспертов и эксперимента (см. разд. 8).

В данном разделе будут рассмотрены:

— дисперсионный анализ;
— насыщенные дробные факторные планы;
— планы Плакетта—Бермана;
— метод случайного баланса.

Для приводимых ниже методов важное значение имеют понятия числа степеней свободы и степени насыщенности планов

экспериментов. Под числом степеней свободы φ будет пониматься разность числа экспериментов N и числа искомых параметров k[1]:

$$\varphi = N - k.$$

Если $\varphi > 0$ и $N > k$, т. е. число проведенных экспериментов превышает число оцениваемых параметров, планы называют ненасыщенными. В этом случае возможна проверка адекватности модели. При $\varphi = 0$ и $N = k$, т. е. при равенстве числа экспериментов числу искомых параметров, планы называют насыщенными. Для проверки адекватности здесь уже не остается степеней свободы.

При $\varphi < 0$, т. е. при $N < k$, число оцениваемых параметров превышает число экспериментов и планы называют сверхнасыщенными.

7.2. Дисперсионный анализ [1—4]

7.2.1. Область применения и вводные замечания

Дисперсионный анализ, первоначально использовавшийся для обработки сельскохозяйственных экспериментов, в настоящее время является одним из наиболее распространенных и общих методов статистического анализа. Дисперсионный анализ основывается на линейной математической модели. В качестве исходных данных используются n случайных измеренных значений, являющихся реализациями n случайных переменных $y_1, y_2, \ldots, y_n$, математические ожидания которых представляют собой линейную форму от p неизвестных параметров $a_1, a_2, \ldots, a_p$:

$$y_i = a_1 x_{1i} + a_2 x_{2i} + \ldots + a_p x_{pi} + e_i, \quad i = 1\,(1)\,n. \qquad (7.1)$$

Здесь x_{ji} принимает значения 1 или 0 в зависимости от того, входит ли параметр a_j в i-е уравнение (7.1) или нет. Случайная переменная e_i означает ошибку измерений в i-м наблюдении. Относительно e_i предполагается, что их математические ожидания равны нулю и что они некоррелированны:

$$E(e_i) = 0,$$

$$E(e_i e_j) = \begin{cases} \sigma^2 & \text{для } i = j, \\ 0 & \text{для } i \neq j. \end{cases}$$

Необходимо подчеркнуть, что дисперсия σ^2 не зависит от i и, следовательно, должна иметь одно и то же значение для всех

[1] Здесь предполагается, что в каждой точке плана эксперимента проводится одно наблюдение, т. е. число точек плана (в разд. 6 это число обозначено через r) совпадает с числом наблюдений N. Кроме того, в разд. 2 было дано более общее определение числа степеней свободы.— *Прим. ред.*

случайных переменных y_i. Оценки параметров a_j, $j = 1(1)p$, в модели (7.1) рассчитываются методом наименьших квадратов по измеренным значениям $\tilde{y}_i$. Для дальнейшего статистического анализа необходимо сделать определенные допущения о характере распределения случайных переменных y_i. Ниже всюду будет предполагаться, что y_i принадлежат нормально распределенной генеральной совокупности и что наблюдаемые значения являются независимыми. При этих предпосылках могут быть проверены гипотезы о параметрах, например о том, что они равны между собой, или о равенстве их нулю.

Эйзенхардтом предложена классификация задач, связанных с дисперсионным анализом. В задачах первого вида речь идет в основном о сравнении множества средних на основе модели I (модель с постоянными эффектами). В этой модели a_j, $j = 1(1)p$, рассматриваются как неизменные параметры, подлежащие оценке. В задачах второго вида, напротив, представляет интерес доля изменчивости, вносимая отдельными факторами; параметры a_j в модели II рассматриваются как случайные переменные (модель со случайными эффектами).

Если же постановка задачи включает как сравнение средних, так и анализ общей изменчивости, то за основу принимается модель III, представляющая собой сочетание двух ранее названных. В настоящей работе в разд. 7.2.2—7.2.7 будет рассматриваться только модель I.

7.2.2. Однофакторная классификация

Пусть получено p выборок из $p\,(p > 1)$ генеральных совокупностей и необходимо сопоставить p значений выборочных средних. Задачу можно сформулировать иначе: имеется фактор A, принимающий одно из p дискретных значений, и необходимо сравнить p эффектов. Можно говорить о разделении фактора A на p уровней или классов, например:

Фактор A	p уровней
Нефть	p разных сортов, отличающихся содержанием воды
Бензин	p разных сортов, отличающихся октановыми числами
Катализатор	p разных катализаторов, отличающихся активностью

Действие фактора A проявляется в p сравниваемых средних значениях. Мы будем исходить из следующей линейной модели, описывающей наблюдения:

$$\tilde{y}_{ij} = m + a_i + \tilde{e}_{ij}, \quad j = 1(1)n_i, \quad i = 1(1)p, \tag{7.2}$$

где n_i — объем выборки для i-го значения фактора A. Константа m обозначает общее среднее значение, константы a_i — эффекты i-го уровня фактора A, а переменные e_{ij} — независимые и нормально распределенные ошибки наблюдений (с математическим ожиданием 0 и дисперсией σ^2). Оценки параметров нельзя получить однозначно на основании модели (7.2), поскольку можно добавить к m и вычесть из a_i произвольную константу. Поэтому вводят еще условие репараметризации:

$$n_1 a_1 + n_2 a_2 + \ldots + n_p a_p = 0, \tag{7.3}$$

которое означает группировку a_i в виде положительных или отрицательных эффектов около общего среднего m и не ограничивает общности модели. Применяя метод наименьших квадратов, связанный с минимизацией суммы квадратов ошибок

$$S = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\tilde{y}_{ij} - m - a_i)^2 \tag{7.4}$$

при дополнительном условии (7.3), получаем оценки параметров

$$\begin{aligned} \hat{m} &= \tilde{y}_{..} = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \tilde{y}_{ij}, \\ \hat{a}_i &= \tilde{y}_{i.} - \tilde{y}_{..} = \frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{y}_{ij} - \hat{m}, \quad i = 1\,(1)\,p, \end{aligned} \tag{7.5}$$

где

$$n = \sum_{i=1}^{p} n_i.$$

Результаты измерений целесообразно представлять в виде табл. 7.1.

Смысл дисперсионного анализа при однофакторной классификации заключается в проверке гипотезы $H_A: a_1 = a_2 = \ldots = a_p = 0$, означающей равное воздействие всех p уровней фактора. Помимо общего и построчных средних, применяются также и другие выборочные характеристики:

$$\tilde{y}_{..} = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \tilde{y}_{ij} \quad \text{(общее среднее)},$$

$$\tilde{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{y}_{ij} \quad \text{(средние по строкам)},$$

$$S_0 = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\tilde{y}_{ij} - \tilde{y}_{..})^2 \quad \text{(сумма квадратов, связанная с рассеянием относительно общего среднего)}, \tag{7.6}$$

$S_A = \sum_{l=1}^{p} n_l (\tilde{y}_{l\cdot} - \tilde{y}_{\cdot\cdot})^2$ (сумма квадратов, связанная с рассеянием между уровнями фактора A),

$S_R = S_0 - S_A$ (сумма квадратов, связанная с рассеянием внутри отдельных уровней фактора A).

Таблица 7.1

Результаты измерений при однофакторной классификации

l \ j		Число наблюдений по выборкам: 1, 2, ..., n_l	Сумма результатов наблюдений по выборкам	Среднее по выборкам
Фактор A	1	$\tilde{y}_{11}\tilde{y}_{12} \dots \tilde{y}_{1n_1}$	$\sum_{j=1}^{n_1} \tilde{y}_{1j}$	$\tilde{y}_{1\cdot}$
	2	$\tilde{y}_{21}\tilde{y}_{22} \dots \tilde{y}_{2n_2}$	$\sum_{j=1}^{n_2} \tilde{y}_{2j}$	$\tilde{y}_{2\cdot}$
	⋮		⋮	⋮
	p	$\tilde{y}_{p1}\tilde{y}_{p2} \dots \tilde{y}_{pn_p}$	$\sum_{j=1}^{n_p} \tilde{y}_{pj}$	$\tilde{y}_{p\cdot}$
			$\sum_{l=1}^{p} \sum_{j=1}^{n_l} \tilde{y}_{lj}$	$\tilde{y}_{\cdot\cdot}$

Если справедлива гипотеза H_A, то отношение s_A^2/s_R^2, где $s_A^2 = S_A/(p-1)$ и $s_R^2 = S_R/(n-p)$, имеет F-распределение с числами степеней свободы $\varphi_1 = p-1$ и $\varphi_2 = n-p$. Эти значения целесообразно представлять в виде табл. 7.2. Здесь s_R^2 — несмещенная оценка дисперсии экспериментальной ошибки. Гипотеза $H_A: a_1 = a_2 = \dots = a_p = 0$ отвергается при выбранном уровне значимости α, если $F > F_{кр}$, где $F_{кр}$ — критическое значение F-распределения при числах степеней свободы $\varphi_1 = p-1$ и $\varphi_2 = n-p$. В противном случае делается вывод о том, что результаты наблюдений не противоречат гипотезе H_A.

Смысл дальнейшего анализа результатов эксперимента заключается в проверке гипотез о линейных комбинациях эффектов, например $a_1 - 2a_2 + a_3 = 0$. Мы не будем останавливаться здесь на этих проверках. Наиболее благоприятным является случай, когда объемы выборок равны между собой: $n_1 = n_2 = \dots = n_p$.

При этом, например, меньше сказывается нарушение допущения о равенстве дисперсий σ^2 для всех y_{lj}. Критерии для проверки ряда следующих ниже гипотез вообще оказываются применимыми лишь при одинаковых объемах выборок.

В качестве примера рассмотрим задачу исследования катализатора. Катализатор для химической реакции получался четырьмя различными способами, которые соответствовали $p = 4$ уровням фактора А — катализатора. В экспериментах измерялась активность катализатора, причем для каждого уровня было сделано по одинаковому числу $n_l = 5$ наблюдений, где $l = 1(1)4$. Необходимо было проверить равенство средних значений активности, т. е. независимость качества катализатора от способа получения. Результаты измерений приведены в табл. 7.3, а

Таблица 7.2

Дисперсионный анализ при однофакторной классификации

Источник изменчивости	S	φ	s^2	F
Между уровнями фактора А	$S_A = \sum_{l=1}^{p} n_l \tilde{y}_{l\cdot}^2 - n\tilde{y}_{\cdot\cdot}^2$	$p-1$	$s_A^2 = \frac{S_A}{p-1}$	$\frac{s_A^2}{s_R^2}$
Внутри уровней фактора А	$S_R = \sum_l \sum_j \tilde{y}_{lj}^2 - \sum_l n_l \tilde{y}_{l\cdot}^2$	$n-p$	$s_R^2 = \frac{S_R}{n-p} = s^2$	
	$S_0 = \sum_l \sum_j \tilde{y}_{lj}^2 - n\tilde{y}_{\cdot\cdot}^2$	$n-1$		

Таблица 7.3

Измеренные значения для задачи исследования активности катализатора

l / t		Номер наблюдения 1	2	3	4	5	$\sum$	$\tilde{y}_{l\cdot}$
Уровни фактора А	1	56	55	62	59	60	292	58,4
	2	64	61	50	55	56	286	57,2
	3	45	46	45	39	43	218	43,6
	4	42	39	45	43	41	210	42,0
							1006	50,3

Таблица 7.4

Дисперсионный анализ для задачи исследования активности катализатора

Вид изменчивости	S	φ	s^2	F
Междууровневая	1135,0	3	378,3	29,8
Внутриуровневая	203,2	16	12,7	
	1338,2	19		

табл. 7.4 иллюстрирует метод дисперсионного анализа для данной задачи. Найдены следующие оценки параметров: $\hat{m}=50{,}3$; $\hat{a}_1=8{,}1$; $\hat{a}_2=6{,}9$; $\hat{a}_3=-6{,}7$; $\hat{a}_4=-8{,}3$. Для проверки гипотезы $H_A: a_1=a_2=a_3=a_4=0$ находим из таблиц F-распределения (табл. 15.5) при уровне значимости $\alpha=0{,}01$ для $\varphi_1=p-1=3$ и $\varphi_2=\sum_{i=1}^{p} n_i - p=16$ критическое значение $F_{кр}=5{,}29$. Это значение сравнивается с расчетной величиной F. Поскольку $F_{кр}=5{,}29<29{,}8=F$, гипотеза о равенстве средних значений отвергается. Другими словами, на качество катализатора оказывают влияние способы его получения.

7.2.3. Двухсторонняя классификация с однократными наблюдениями

Задача двухсторонней классификации (кроссклассификации, или двухфакторного дисперсионного анализа) возникает при проведении наблюдений в эксперименте, в котором одновременно действуют два фактора A и B, варьируемые соответственно на p и q уровнях. При однократных наблюдениях для каждой комбинации уровней факторов A и B в распоряжении исследователя имеется лишь одно измеренное значение, а их общее число равно pq. Для $n=pq$ выборок y_{ij}, $i=1\,(1)p$, $j=1\,(1)q$, объема 1 предполагается, что наблюдаемые значения представимы в виде

$$\tilde{y}_{ij}=m+a_i+b_j+\tilde{e}_{ij}, \tag{7.7}$$

где m, a_i и b_j — некоторые константы, а e_{ij} — независимые нормально распределенные случайные величины с математическим ожиданием 0 и одной и той же дисперсией σ^2.

Используя выборочные характеристики

$$\tilde{y}_{i\cdot} = \frac{1}{q}\sum_{j=1}^{q}\tilde{y}_{ij} \qquad \text{(средние по строкам)},$$

$$\tilde{y}_{\cdot j} = \frac{1}{p}\sum_{i=1}^{p}\tilde{y}_{ij} \qquad \text{(средние по столбцам)}, \tag{7.8}$$

$$\tilde{y}_{\cdot\cdot} = \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q}\tilde{y}_{ij} \qquad \text{(общее среднее)}$$

и условия репараметризации

$$\sum_{i=1}^{p} a_i = 0, \quad \sum_{j=1}^{q} b_j = 0, \tag{7.9}$$

можно методом наименьших квадратов найти оценки параметров модели (7.7):

$$\begin{aligned} &\hat{m} = \tilde{y}_{\cdot\cdot}, \\ &\hat{a}_i = \tilde{y}_{i\cdot} - \tilde{y}_{\cdot\cdot}, \qquad i = 1\,(1)\,p, \\ &\hat{b}_j = \tilde{y}_{\cdot j} - \tilde{y}_{\cdot\cdot}, \qquad j = 1\,(1)\,q. \end{aligned} \tag{7.10}$$

Результаты измерений и выборочные характеристики целесообразно представить в виде табл. 7.5. Задачей дисперсионного анализа здесь является проверка равенства нулю параметров a_i и b_j. Говоря точнее, необходимо проверить гипотезы H_A: $a_1 = a_2 = \ldots = a_p = 0$ и H_B: $b_1 = b_2 = \ldots = b_p = 0$.

Рассмотрим суммы квадратов

$$\begin{aligned} &S_A = q\sum_{i=1}^{p}(\tilde{y}_{i\cdot} - \tilde{y}_{\cdot\cdot})^2, \quad s_A^2 = \frac{S_A}{p-1}, \\ &S_B = p\sum_{j=1}^{q}(\tilde{y}_{\cdot j} - \tilde{y}_{\cdot\cdot})^2, \quad s_B^2 = \frac{S_B}{q-1}, \\ &S_R = \sum_{i=1}^{p}\sum_{j=1}^{q}(\tilde{y}_{ij} - \tilde{y}_{i\cdot} - \tilde{y}_{\cdot j} + \tilde{y}_{\cdot\cdot})^2, \quad s_R^2 = \frac{S_R}{(p-1)(q-1)}. \end{aligned} \tag{7.11}$$

Тогда для гипотез H_A и H_B отношения s_A^2/s_R^2 и s_B^2/s_R^2 имеют F-распределение со степенями свободы $\varphi_1 = p-1$, $\varphi_2 = (q-1)(p-1)$ и $\varphi_1 = q-1$, $\varphi_2 = (p-1)(q-1)$ соответственно [1]. Рассчитанные числовые значения заносятся в табл. 7.6.

[1] Сумма квадратов S_R в данном случае характеризует рассеяние, вызванное совместным действием факторов A и B. Полагая, что эффект взаимодействия факторов равен нулю, с помощью оценки s_R^2 определяем дисперсию ошибок наблюдений. — *Прим. ред.*

Гипотеза H_A: $a_1 = a_2 = \ldots = a_p = 0$ отвергается, если для заданного уровня значимости α рассчитанное значение $F_A > F_{кр}$, где $F_{кр}$ находится по таблицам F-распределения при числах

Таблица 7.5

Результаты измерений при двухсторонней классификации с однократными наблюдениями

i \ j		Уровни фактора B: 1	2	...	q	Среднее по строкам
Уровни фактора A	1	$\tilde{y}_{11}$	$\tilde{y}_{12}$	...	$\tilde{y}_{1q}$	$\tilde{y}_{1\cdot}$
	2	$\tilde{y}_{21}$	$\tilde{y}_{22}$	...	$\tilde{y}_{2q}$	$\tilde{y}_{2\cdot}$
	⋮	⋮	⋮		⋮	⋮
	p	$\tilde{y}_{p1}$	$\tilde{y}_{p2}$	...	$\tilde{y}_{pq}$	$\tilde{y}_{p\cdot}$
Среднее по столбцам		$\tilde{y}_{\cdot 1}$	$\tilde{y}_{\cdot 2}$	...	$\tilde{y}_{\cdot q}$	$\tilde{y}_{\cdot\cdot}$

Таблица 7.6

Дисперсионный анализ для двухсторонней классификации с однократными наблюдениями

Источник изменчивости	S	φ	s^2	F
Фактор A	$S_A = q\sum_{i=1}^{p} \tilde{y}_{i\cdot}^2 - pq\tilde{y}_{\cdot\cdot}^2$	$p-1$	s_A^2	$\frac{s_A^2}{s_R^2} = F_A$
Фактор B	$S_B = p\sum_{j=1}^{q} \tilde{y}_{\cdot j}^2 - pq\tilde{y}_{\cdot\cdot}^2$	$q-1$	s_B^2	$\frac{s_B^2}{s_R^2} = F_B$
Экспериментальная ошибка	$S_R = \sum_{i=1}^{p}\sum_{j=1}^{q} \tilde{y}_{ij}^2 - q\sum_{i=1}^{p} \tilde{y}_{i\cdot}^2 - p\sum_{j=1}^{q} \tilde{y}_{\cdot j}^2 + pq\tilde{y}_{\cdot\cdot}^2$	$(p-1)\times(q-1)$	$s_R^2 = s^2$	
	$S = S_A + S_B + S_R$	$pq-1$		

степеней свободы $\varphi_1 = p-1$ и $\varphi_2 = (p-1)(q-1)$. Если $F_A < F_{кр}$, то на основании данного критерия гипотезу следует признать приемлемой. Аналогичным образом проверяется гипотеза H_B. Необходимо отметить, что альтернативой этим гипотезам является то, что по крайней мере один эффект a_i или соответственно b_j значимым образом отличается от 0.

В качестве примера рассмотрим испытание автомобильных покрышек. Четыре различных типа покрышек (фактор A на $p = 4$ уровнях) испытываются на автомобилях четырех различных марок (фактор B на $q = 4$ уровнях), причем измеряется износ в миллиметрах после пробега 40 000 км. Данные измерений, отнесенные к среднему износу в 10 мм, представлены в табл. 7.7. Требуется оценить влияние на износ различных типов покрышек и марок автомобилей. В табл. 7.8 сведены результаты расчетов при проведении дисперсионного анализа. При

Таблица 7.7

Данные измерений для примера с испытанием покрышек

i \ j		Марка автомобиля B				$\tilde{y}_{i\cdot}$
		1	2	3	4	
Тип покрышек A	1	4	1	−1	0	1,00
	2	1	1	−1	−2	−0,25
	3	0	0	−3	−2	−1,25
	4	0	−5	−4	−4	−3,25
$\tilde{y}_{\cdot j}$		1,25	−0,75	−2,25	−2,00	−0,94

Таблица 7.8

Дисперсионный анализ для примера с испытанием покрышек

Источник изменчивости	S	φ	s^2	F
Фактор A	30,6	3	10,2	7,8
Фактор B	38,6	3	12,9	9,9
Ошибка	11,7	9	1,3	
	80,9	15		

уровне значимости $\alpha = 0{,}01$ и числах степеней свободы $\varphi_1 = 3$ и $\varphi_2 = 9$ из табл. 15.5 находим значение $F_{кр} = 7{,}0$. Поскольку

$$F_A = 7{,}8 > 7{,}0 = F_{кр}$$

и

$$F_B = 9{,}9 > 7{,}0 = F_{кр},$$

то отвергаются обе гипотезы H_A: $a_i = 0$, $i = 1\,(1)\,4$, и H_B: $b_j = 0$, $j = 1\,(1)\,4$. Различные типы покрышек и марки автомобилей оказывают разное влияние на износ покрышек. Поэтому не безразлично, какие покрышки устанавливаются на автомобили той или иной марки.

7.2.4. Двухсторонняя классификация с многократными наблюдениями

Пусть факторы A и B имеют по-прежнему p и q уровней соответственно. В каждом из pq подклассов (i, j) теперь уже содержится r результатов наблюдений $\tilde{y}_{ij1}, \tilde{y}_{ij2}, \ldots, \tilde{y}_{ijr}$. Данные измерений можно упорядочить аналогично табл. 7.5, в которую они записываются столбцами по r значений, завершающихся значениями средних по столбцам. Включим теперь в рассматриваемую линейную математическую модель эффекты взаимодействий $(ab)_{ij}$

$$\tilde{y}_{ijk} = m + a_i + b_j + (ab)_{ij} + \tilde{e}_{ijk}. \tag{7.12}$$

Условия репараметризации в данном случае имеют вид

$$\begin{aligned} &\sum_{i=1}^{p} a_i = 0, \quad \sum_{j=1}^{q} b_j = 0, \\ &\sum_{i=1}^{p} (ab)_{ij} = 0, \qquad j = 1\,(1)\,q, \\ &\sum_{j=1}^{q} (ab)_{ij} = 0, \qquad i = 1\,(1)\,p. \end{aligned} \tag{7.13}$$

С помощью метода наименьших квадратов можно найти однозначные точечные оценки параметров m, a_i, b_j, $(ab)_{ij}$. Если ввести выборочные характеристики

$$\begin{aligned} \tilde{y}_{ij.} &= \frac{1}{r}\sum_{k=1}^{r} \tilde{y}_{ijk}, & \tilde{y}_{i..} &= \frac{1}{qr}\sum_{j=1}^{q}\sum_{k=1}^{r} \tilde{y}_{ijk}, \\ \tilde{y}_{.j.} &= \frac{1}{pr}\sum_{i=1}^{p}\sum_{k=1}^{r} \tilde{y}_{ijk}, & \tilde{y}_{...} &= \frac{1}{pqr}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} \tilde{y}_{ijk} \end{aligned} \tag{7.14}$$

и

$$
\begin{aligned}
S_A &= qr\sum_{i=1}^{p}(\tilde{y}_{i..}-\tilde{y}_{...})^2, & s_A^2 &= \frac{S_A}{p-1},\\
S_B &= pr\sum_{j=1}^{q}(\tilde{y}_{.j.}-\tilde{y}_{...})^2, & s_B^2 &= \frac{S_B}{q-1},\\
S_{A\times B} &= r\sum_{i=1}^{p}\sum_{j=1}^{q}(\tilde{y}_{ij.}-\tilde{y}_{...})^2, & s_{A\times B}^2 &= \frac{S_{A\times B}}{(p-1)(q-1)},\\
S_R &= r\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(\tilde{y}_{ijk}-\tilde{y}_{ij.})^2, & s_R^2 &= \frac{S_R}{pq(r-1)},
\end{aligned}
\qquad (7.15)
$$

то результаты дисперсионного анализа можно представить в виде табл. 7.9.

Таблица 7.9

Дисперсионный анализ при двухсторонней классификации с r-кратным повторением наблюдений в каждом классе

Источник изменчивости	S	φ	s^2	F
Фактор A	$S_A = qr\left(\sum_{i=1}^{p}\tilde{y}_{i..}^2 - p\tilde{y}_{...}^2\right)$	$p-1$	s_A^2	$F_A = \frac{s_A^2}{s_R^2}$
Фактор B	$S_B = pr\left(\sum_{j=1}^{q}\tilde{y}_{.j.}^2 - q\tilde{y}_{...}^2\right)$	$q-1$	s_B^2	$F_B = \frac{s_B^2}{s_R^2}$
Взаимодействие A×B	$S_{A\times B} = r\sum_{i=1}^{p}\sum_{j=1}^{q}\tilde{y}_{ij.}^2 - qr\sum_{i=1}^{p}\tilde{y}_{i..}^2 - pr\sum_{j=1}^{q}\tilde{y}_{.j.}^2 + pqr\tilde{y}_{...}^2$	$(p-1)(q-1)$	$s_{A\times B}^2$	$F_{A\times B} = \frac{s_{A\times B}^2}{s_R^2}$
Экспериментальная ошибка	$S_R = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\tilde{y}_{ijk}^2 - r\sum_{i=1}^{p}\sum_{j=1}^{q}\tilde{y}_{ij.}^2$	$pq(r-1)$	s_R^2	
	$S = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\tilde{y}_{ijk}^2 - pqr\tilde{y}_{...}^2$	$pqr-1$		

Величина $s_R^2 = s^2$ вновь представляет собой несмещенную оценку дисперсии экспериментальной ошибки σ^2.

Для проверки гипотез об эффектах

$$H_A\colon\ a_1 = a_2 = \ldots = a_p = 0, \quad H_B\colon\ b_1 = b_2 = \ldots = b_q = 0$$

и

$$H_{A\times B}\colon\ (ab)_{ij} = 0, \quad i = 1\,(1)\,p, \quad j = 1\,(1)\,q$$

применяется F-критерий. Гипотезы H_A, H_B или $H_{A\times B}$ отвергаются, если при заданном уровне значимости соответственно окажется, что $F_A > F_{кр}$, $F_B > F_{кр}$ или $F_{A\times B} > F_{кр}$, где $F_{кр}$ — кри-

Таблица 7.10

Данные измерений активности катализатора

Фактор B (i)	j	Фактор A					$\tilde{y}_{.j.}$
		500	540	570	600		
Фактор B	5	29	98	44	84	821	68,5
		30	128	81	100		
		20	67	77	63		
		79	293	202	247		
	10	22	35	53	103	704	58,6
		26	80	93	90		
		16	29	59	98		
		64	144	205	291		
	15	18	49	58	80	761	63,5
		17	68	103	91		
		11	61	128	77		
		46	278	289	248		
	20	38	68	87	86	852	71,0
		31	74	116	113		
		21	47	90	81		
		90	189	293	280		
		279	804	989	1066	3138	
$\tilde{y}_{i..}$		23,2	67,5	82,5	88,8		65,4

Таблица 7.11

Дисперсионный анализ активности катализатора

Источник изменчивости	S	φ	s^2	F
Фактор A	31 414	3	10 472	31,78
Фактор B	1 077	3	359	1,09
Взаимодействие	6 474	9	719	2,18
Ошибка	10 544	32	330	
	49 509	47		

тические значения F-распределения со степенями свободы $\varphi_1 = p-1$ и $\varphi_2 = pq(r-1)$, $\varphi_1 = q-1$ и $\varphi_2 = pq(r-1)$ или $\varphi_1 = (p-1)(q-1)$ и $\varphi_2 = pq(r-1)$.

Рассмотрим в качестве примера задачу исследования некоторого катализатора. Носитель катализатора — кизельгур — может иметь 4 различные величины зерен (фактор A). В процессе прокаливания при получении катализатора поддерживались 4 различных температурных режима (фактор B). Для каждого из 16 разных сортов катализатора было проведено по 3 измерения активности. Соответствующие числовые значения для $p = q = 4$ и $r = 3$ занесены в табл. 7.10. Целью исследования являлась оценка влияния температурного режима прокаливания и величины зерен на качество катализатора.

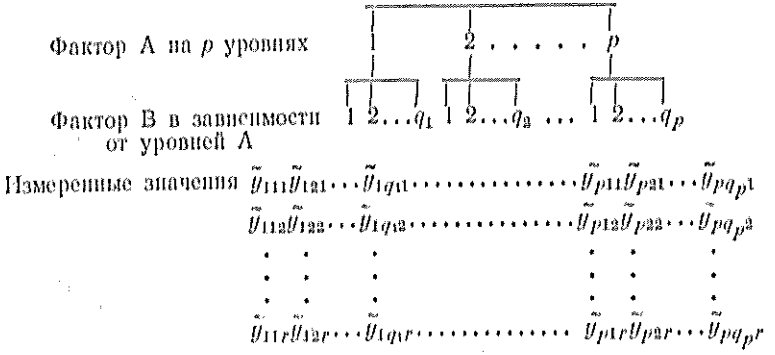
Табл. 7.11 содержит числовые данные, требуемые для дисперсионного анализа.

Для уровня значимости $\alpha = 0{,}01$ и при числах степеней свободы $\varphi_1 = p-1 = 3$ и $\varphi_2 = pq(r-1) = 32$ по табл. 15.5 находим критическое значение F-распределения $F_{кр} = 4{,}5$. Поскольку $F_{кр} = 4{,}5 < 31{,}78 = F_A$, гипотеза H_A отвергается. Таким образом, величина зерен кизельгура оказывает влияние на активность катализатора. Напротив, гипотеза H_B из-за $F_{кр} = 4{,}5 > 1{,}09 = F_B$ не может быть отвергнута. Следовательно, влияние различных температур прокаливания на активность катализатора доказать не удалось.

При числах степеней свободы $\varphi_1 = (p-1)(q-1) = 9$ и $\varphi_2 = pq(r-1) = 32$ из табл. 15.5 получаем $F_{кр} = 3{,}2$. Влиянием эффектов взаимодействий можно пренебречь, так как $F_{кр} = 3{,}2 > 2{,}18 = F_{A \times B}$.

7.2.5. Двухступенчатая иерархическая классификация

Иерархическая классификация применяется в случае, если появление некоторых уровней одного фактора связано с определенным уровнем другого фактора. Здесь мы рассмотрим лишь двухфакторную (двухступенчатую) иерархическую классификацию с равными числами наблюдений в наименьших подклассах. Обобщение на случай многоступенчатой иерархии не представляет особых трудностей. Экспериментальный план, построенный с учетом особенностей взаимодействия уровней факторов A и B, можно предетавить в следующем виде:



Предполагается, что $(q_1 + q_2 + \ldots + q_p)\,r$ выборочных значений могут быть представлены в виде

$$\tilde{y}_{ijk} = m + a_i + b_{ij} + \tilde{e}_{ijk}, \quad i = 1\,(1)\,p, \quad j = 1\,(1)\,q_i, \quad k = 1\,(1)\,r, \qquad (7.16)$$

где m, a_i, b_{ij} — константы, а e_{ijk} — независимые нормально распределенные случайные величины $N(0, \sigma^2)$.

Вводя условия репараметризации

$$\sum_{i=1}^{p} a_i = 0, \quad \sum_{j=1}^{q_i} b_{ij} = 0, \quad i = 1\,(1)\,p, \qquad (7.17)$$

с помощью метода наименьших квадратов находим оценки $\hat{m}$, $\hat{a}_i$ и $\hat{b}_{ij}$. При этом в основные эффекты фактора B входят также эффекты взаимодействия с i-м уровнем фактора A и разделить их оказывается невозможно.

При дисперсионном анализе проверяются две гипотезы. Гипотеза H_A: $a_i = 0$, $i = 1\,(1)\,p$, означает равенство основных эффектов фактора A. Для проверки гипотезы H_A используют случайную величину

$$F_A = \frac{s_A^2}{s_R^2} = \frac{(r-1)\sum_{i=1}^{p} q_i}{p-1} \frac{S_A}{S_R}, \qquad (7.18)$$

которая имеет F-распределение с числами степеней свободы $\varphi_1 = p-1$ и $\varphi_2 = (r-1)\sum_{i=1}^{p} q_i$. Гипотеза $H_{B(A)}$: $b_{ij} = 0$, $j = 1(1)q_i$, $i = 1(1)p$, означает равенство основных эффектов фактора B при i-м уровне фактора A. Для проверки гипотезы используют случайную величину

$$F_{B(A)} = \frac{s^2_{B(A)}}{s^2_R} = \frac{(r-1)\sum_{i=1}^{p} q_i}{\sum_{i=1}^{p} q_i - p} \frac{S_{B(A)}}{S_R}, \qquad (7.19)$$

которая также имеет F-распределение с числами степеней свободы

$$\varphi_1 = \sum_{i=1}^{p} q_i - p, \qquad \varphi_2 = (r-1)\sum_{i=1}^{p} q_i.$$

Числовые значения представляются в виде табл. 7.12. Как и раньше, в таблице используются величины

$$\tilde{y}_{ij.} = \frac{1}{r}\sum_{k=1}^{r} \tilde{y}_{ijk}, \quad \tilde{y}_{i..} = \frac{1}{q_i}\sum_{j=1}^{q} \tilde{y}_{ij.}, \quad \tilde{y}_{...} = \frac{1}{p}\sum_{i=1}^{p} \tilde{y}_{i..}. \qquad (7.20)$$

Таблица 7.12

Дисперсионный анализ при двухступенчатой иерархической классификации

Источник изменчивости	S	φ	s^2	F
Фактор A	$S_A = r\sum_{i=1}^{p} q_i(\tilde{y}_{i..} - \tilde{y}_{...})^2$	$p-1$	s^2_A	$F_A = \frac{s^2_A}{s^2_R}$
Фактор B (A)	$S_{B(A)} = r\sum_{i,j} (\tilde{y}_{ij.} - \tilde{y}_{i..})^2$	$\sum_{i=1}^{p} q_i - p$	$s^2_{B(A)}$	$F_{B(A)} = \frac{s^2_{B(A)}}{s^2_R}$
Ошибка наблюдений	$S_R = \sum_{i,j,k} (\tilde{y}_{ijk} - \tilde{y}_{ij.})^2$	$(r-1)\sum_{i=1}^{p} q_i$	s^2_R	
	$S = S_A + S_{B(A)} + S_R$	$r\sum_{i=1}^{p} q_i - 1$		

Гипотеза H_A или $H_{B(A)}$ отвергается, если для заданного уровня значимости α — вычисленное значение $F_A > F_{кр}$ или соответственно $F_{B(A)} > F_{кр}$, где $F_{кр}$ — критическое значение F-распределения, имеющее место при соответствующем числе степеней свободы. Если это не так, то результаты наблюдений не противоречат высказанной гипотезе.

7.2.6. Трехфакторная классификация в латинском квадрате

Многофакторная классификация может рассматриваться аналогично двухфакторной и не обнаруживает новых особенностей. Однако объем плана существенно возрастает, так что при n-факторной классификации с двухуровневыми факторами и однократными наблюдениями требуется произвести 2^n измерений отклика. Налагая ограничения на план, получают неполные классификации. Частным случаем неполной трехфакторной классификации является так называемый латинский квадрат, впервые примененный в сельскохозяйственных экспериментах.

Рассматриваются три фактора, каждый из которых имеет p уровней. Предполагается, что в модели отсутствуют эффекты взаимодействия факторов A и B, с одной стороны, и фактора C, — с другой. Пример экспериментального плана при $p = 4$ приведен в табл. 7.13. Уровень фактора C определяется третьим индексом k измеренных значений $\tilde{y}_{ijk}$. Принцип изменения k демонстрируется с помощью квадрата, помещенного справа от табл. 7.13. Для латинского квадрата такого типа характерно, что каждое значение $k = 1\,(1)\,p$ в каждой строке и каждом столбце квадрата встречается только один раз. Измеренные значения $\tilde{y}_{ijk}$ рассматриваются как реализации p^2 выборок объема 1, и для них постулируется представимость результатов наблюдений в виде

$$\tilde{y}_{ijk} = m + a_i + b_j + c_k + \tilde{e}_{ijk},$$

где m, a_i, b_j, c_k — константы, а e_{ijk} — случайная величина, имеющая нормальное распределение $N(0, \sigma^2)$. Используя условия репараметризации

$$\sum_{i=1}^{p} a_i = \sum_{j=1}^{p} b_j = \sum_{k=1}^{p} c_k = 0, \tag{7.21}$$

можно методом наименьших квадратов найти оценки $\hat{m}$, $\hat{a}_i$, $\hat{b}_j$, $\hat{c}_k$. Для проверки гипотез H_A: $a_i = 0$, $i = 1\,(1)\,p$; H_B: $b_j = 0$,

Таблица 7.13

Четырехуровневый латинский квадрат

i \ j		Уровни фактора B			
		1	2	3	4
Уровни фактора A	1	$\tilde{y}_{111}$	$\tilde{y}_{122}$	$\tilde{y}_{133}$	$\tilde{y}_{144}$
	2	$\tilde{y}_{212}$	$\tilde{y}_{221}$	$\tilde{y}_{234}$	$\tilde{y}_{243}$
	3	$\tilde{y}_{313}$	$\tilde{y}_{324}$	$\tilde{y}_{332}$	$\tilde{y}_{341}$
	4	$\tilde{y}_{414}$	$\tilde{y}_{423}$	$\tilde{y}_{431}$	$\tilde{y}_{442}$

1	2	3	4
2	1	4	3
3	4	2	1
4	3	1	2

$j=1\,(1)\,p$ и H_C: $c_k=0$, $k=1\,(1)\,p$ используются значения выборочных средних

$$\tilde{y}_{i..}=\frac{1}{p}\sum_{l,\,k}\tilde{y}_{ilk},\quad \tilde{y}_{.j.}=\frac{1}{p}\sum_{i,\,k}\tilde{y}_{ijk},$$
$$\tilde{y}_{..k}=\frac{1}{p}\sum_{i,\,l}\tilde{y}_{ilk},\quad \tilde{y}_{...}=\frac{1}{p^2}\sum_{i,\,l,\,k\in L}\tilde{y}_{ilk}, \tag{7.22}$$

где L — множество таких комбинаций индексов, которые включены в план эксперимента.

Обозначим далее

$$S=\sum_{i,\,l,\,k\in L}(\tilde{y}_{ilk}-\tilde{y}_{...})^2,\quad S_C=p\sum_{k=1}^{p}(\tilde{y}_{..k}-\tilde{y}_{...})^2,$$
$$S_A=p\sum_{i=1}^{p}(\tilde{y}_{i..}-\tilde{y}_{...})^2,\quad S_B=p\sum_{l=1}^{p}(\tilde{y}_{.l.}-\tilde{y}_{...})^2, \tag{7.23}$$
$$S_R=S-S_A-S_B-S_C,$$
$$s_A^2=\frac{S_A}{p-1},\quad s_B^2=\frac{S_B}{p-1},\quad s_C^2=\frac{S_C}{p-1},\quad s_R^2=\frac{S_R}{(p-1)(p-2)}.$$

Гипотезы H_A, H_B и H_C проверяются с помощью величин s_A^2/s_R^2, s_B^2/s_R^2 и s_C^2/s_R^2, подчиняющихся F-распределению с числами степеней свободы $\varphi_1=p-1$ и $\varphi_2=(p-1)\,(p-2)$. Эти гипотезы отвергаются, если вычисленные значения оказываются больше критического значения F-распределения, найденного для выбранного уровня значимости α и соответствующих чисел степеней свободы.

7.2.7. Четырехфакторная классификация в греко-латинском квадрате

Частным случаем планов эксперимента с четырехфакторной классификацией являются так называемые греко-латинские квадраты, которые так же, как и уже рассмотренные латинские

квадраты, возникли из практики проведения сельскохозяйственных экспериментов. В них исследуется действие четырех факторов A, B, C и D, варьируемых на p уровнях каждый. При этом необходимо устранить возможное систематическое взаимное влияние факторов A, B, с одной стороны, и C, D, с другой. Соответствующий план эксперимента для $p=4$ приведен в табл. 7.14. Для греко-латинского квадрата характерно то, что, во-первых, каждое сочетание пар индексов, одна из которых относится к факторам A, B, а другая — к факторам C, D, встречается точно один раз и, во-вторых, в любом столбце и в любой строке всякая пара появляется только однажды. Квадраты такого типа можно построить не для любого порядка; например, они не существуют для $p=6$. Математическая модель, выборочные характеристики, гипотезы и таблицы числовых значений для выполнения дисперсионного анализа представляют собой естественное обобщение трехфакторной классификации в латинском квадрате и поэтому здесь не рассматриваются.

Таблица 7.14

Четырехуровневый греко-латинский квадрат

l \ j		Уровни фактора B			
		1	2	3	4
Уровни фактора A	1	$\tilde{y}_{1111}$	$\tilde{y}_{1222}$	$\tilde{y}_{1333}$	$\tilde{y}_{1444}$
	2	$\tilde{y}_{2124}$	$\tilde{y}_{2213}$	$\tilde{y}_{2343}$	$\tilde{y}_{2431}$
	3	$\tilde{y}_{3132}$	$\tilde{y}_{3231}$	$\tilde{y}_{3414}$	$\tilde{y}_{3423}$
	4	$\tilde{y}_{4143}$	$\tilde{y}_{4231}$	$\tilde{y}_{4321}$	$\tilde{y}_{4412}$

(1,1)	(2,2)	(3,3)	(4,4)
(2,4)	(1,3)	(4,2)	(3,1)
(3,2)	(4,1)	(1,4)	(2,3)
(4,3)	(3,4)	(2,1)	(1,2)

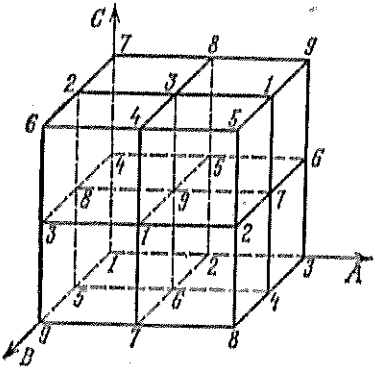
7.2.8. Пример латинского куба [5]

Рассмотрим пример исследования рецептуры получения некоторого полимерного материала. Пусть действуют четыре фактора: A (пластификатор), B (стабилизатор), C (количество трехосновного сульфата свинца в стабилизаторе — ТОСС) и D (наполнитель), причем первые три из них — на трех уровнях и последний фактор — на девяти уровнях. В основу дисперсионного анализа положена модель

$$\tilde{y}_{ijkl}=m+a_i+b_j+c_k+d_l+\tilde{e}_{ijkl},\ i,\ j,\ k=1\,(1)\,3,\ l=1\,(1)\,9, \quad (7.24)$$

где m — общее среднее, a_i, b_j, c_k, d_l — константы, а e_{ijkl} — независимые нормально распределенные случайные величины. В качестве плана эксперимента выбран латинский куб с однократ-

ными наблюдениями, показанный на фиг. 7.2. В табл. 7.15 содержатся используемые комбинации индексов $(i, j, k, l) \in L$ для измеренных значений. Построенный куб можно назвать латинским, так как каждый уровень фактора D в каждом сечении куба, параллельном его сторонам, встречается только один раз, как это отмечено цифрами от 1 до 9 на фиг. 7.2. Всего



Фиг. 7.2. Латинский куб для $p = 3$.

здесь необходимо провести $3^3 = 27$ измерений. При этом измерялись девять различных свойств материала:

y_1 — термостабильность;
y_2 — спектр отражения;
y_3 — морозостойкость;
y_4 — модуль эластичности;
y_5 — прочность;
y_6 — удлинение;
y_7 — число перегибов;
y_8 — липкость;
y_9 — блеск.

В табл. 7.15 приведены результаты измерений двух из девяти свойств: спектра отражения y_2 и прочности y_5. С помощью соответствующего преобразования данные измерений были приведены к интервалам $0 \leqslant \tilde{y}_r \leqslant 1$, $r = 1\,(1)\,9$. Далее были введены две обобщенные характеристики $\tilde{y}_I$ и $\tilde{y}_{II}$:

$$\tilde{y}_I = \sqrt[9]{\prod_{r=1}^{9} \tilde{y}_r}, \quad \tilde{y}_{II} = \sqrt[5]{\prod_{r=2,\,5,\,7,\,8,\,9} \tilde{y}_r}.$$

Характеристика $\tilde{y}_I$ является мультипликативным средним для всех технологических и эксплуатационных свойств, в то время как характеристика $\tilde{y}_{II}$ — мультипликативное среднее для свойств y_2, y_5, y_7, y_8 и y_9, служащее критерием эксплуатационных свойств полимерного материала. С помощью дисперсионного анализа исследовались характеристики качества $\tilde{y}_I$ и $\tilde{y}_{II}$. Результаты представлены в табл. 7.15. В дисперсионном анализе

Таблица 7.15

Влияющие факторы и данные измерений для пластмасс

Фактор	Обозначение	Уровни		
		1	2	3
Пластификатор	A	W-1	W-2	W-3
Стабилизатор	B	C-1	C-2	C-3
Количество ТОСС в стабилизаторе	C	2	3	4
Наполнитель	D	F1,	F2, ...,	F9

Факторы				Характеристики		Измерения	
A i	B j	C k	D l	$\tilde{v}_{\text{I}}$	$\tilde{v}_{\text{II}}$	$\tilde{v}_{\text{a}}$	$\tilde{v}_{\text{b}}$
1	1	1	1	0,650	0,539	—4,0	15
1	2	1	5	0,531	0,404	—10,0	—21
1	3	1	9	0,528	0,409	—18,5	16
2	1	1	2	0,000	0,000	—46,2	—2
2	2	1	6	0,519	0,395	—12,0	9
2	3	1	7	0,000	0,000	5,0	20
3	1	1	3	0,540	0,478	22,5	1
3	2	1	4	0,647	0,566	—13,5	30
3	3	1	8	0,615	0,492	3,0	15
1	1	2	4	0,610	0,605	0,0	40
1	2	2	8	0,000	0,000	—27,0	—6
1	3	2	3	0,000	0,000	22,0	—9
2	1	2	5	0,000	0,000	—11,5	1
2	2	2	9	0,000	0,000	—39,0	16
2	3	2	1	0,000	0,000	—3,0	14
3	1	2	6	0,571	0,569	—1,5	20
3	2	2	7	0,589	0,541	—2,0	25
3	3	2	2	0,000	0,000	—36,5	—9
1	1	3	7	0,448	0,399	1,0	19
1	2	3	2	0,000	0,000	—40,0	—8
1	3	3	6	0,511	0,674	20,0	14
2	1	3	8	0,000	0,000	—5,0	32
2	2	3	3	0,676	0,584	16,5	—11
2	3	3	4	0,720	0,576	9,0	13
3	1	3	9	0,000	0,000	—26,5	—1
3	2	3	1	0,630	0,531	—2,0	24
3	3	3	5	0,688	0,569	8,2	2

использовались два типа выборочных характеристик. Первые из них — это средние значения:

$$\tilde{y}_{i...} = \frac{1}{3^3}\sum_{j,k,l}\tilde{y}_{ijkl}, \quad \tilde{y}_{.j..} = \frac{1}{3^3}\sum_{i,k,l}\tilde{y}_{ijkl},$$
$$\tilde{y}_{..k.} = \frac{1}{3^3}\sum_{i,j,l}\tilde{y}_{ijkl}, \quad \tilde{y}_{....} = \frac{1}{3^3}\sum_{i,j,k,l\in L}\tilde{y}_{ijkl}. \tag{7.25}$$

Суммирование производится по всем комбинациям индексов, встречающимся в плане эксперимента (см. табл. 7.15). Например, $\tilde{y}_{..k.}$ получают путем усреднения по первым девяти значениям таблицы. Вторую группу характеристик составляют следующие суммы квадратов:

$$S_A = 3^3\sum_{i=1}^{3}(\tilde{y}_{i...}-\tilde{y}_{....})^2 = 9\sum_{i=1}^{9}\tilde{y}^2_{i...} - 27\tilde{y}^2_{....},$$
$$S_D = 3\sum_{l=1}^{9}(\tilde{y}_{...l}-\tilde{y}_{....})^2 = 3\sum_{l=1}^{9}\tilde{y}^2_{...l} - 27\tilde{y}^2_{....}, \tag{7.26}$$
$$S = \sum_{i,j,k,l\in L}(\tilde{y}_{ijkl}-\tilde{y}_{....})^2 = \sum_{i,j,k,l\in L}\tilde{y}^2_{ijkl} - 27\tilde{y}^2_{....},$$
$$S_R = S - S_A - S_B - S_C - S_D.$$

Суммы квадратов S_B и S_C получаются аналогично S_A, и для них число степеней свободы равно $\varphi = 3-1 = 2$. Числа степеней свободы таковы: для S_D $\varphi = 3^3-1 = 8$, для S $\varphi = 3^3-1 = 26$ и для остаточной суммы S_R $\varphi = (3^3-1)-(3^3-1)-3(3-1) = 12$. В дисперсионной таблице (табл. 7.16) представлены соответствующие числовые значения для анализа обеих характеристик $\tilde{y}_{\text{I}}$ и $\tilde{y}_{\text{II}}$. Нулевые гипотезы H_A, H_B, H_C и H_D проверялись с помощью F-критерия при доверительной вероятности $P = 0{,}8$. Из таблиц F-распределения [1] для уровня значимости $\alpha = 0{,}2$ находим $F_{\text{кр}} = 1{,}8$ при степенях свободы $\varphi_1 = 2$, $\varphi_2 = 12$ и $F_{\text{кр}} = 1{,}7$ при $\varphi_1 = 8$, $\varphi_2 = 12$.

Поскольку

$$F_A = \{2{,}51\ \ 3{,}02\} > 1{,}8,$$
$$F_B = \{0{,}28\ \ 0{,}12\} < 1{,}8,$$
$$F_C = \{2{,}63\ \ 2{,}05\} > 1{,}8,$$
$$F_D = \{1{,}89\ \ 2{,}30\} > 1{,}7,$$

то как для технико-эксплуатационной характеристики $\tilde{y}_{\text{I}}$, так и для чисто эксплуатационной характеристики $\tilde{y}_{\text{II}}$ гипотезы H_A,

[1] Соответствующие таблицы распределения F-критерия можно найти, например, в книге: Большев Л. Н., Смирнов Н. В., Таблицы математической статистики, изд-во „Наука", М., 1965. — *Прим. ред.*

H_C и H_D отвергаются, а гипотеза H_B принимается. Иначе говоря, на качество материала оказывают влияние следующие факторы: пластификатор, количество ТОСС и наполнитель. Влияние фактора „стабилизатор" оказалось несущественным.

Таблица 7.16

Дисперсионная таблица для примера с исследованием полимерного материала

	Источник изменчивости		S	φ	s^2	F
A	Пластификатор	I II	0,3132 0,2774	2	0,1566 0,1487	2,51 3,02
B	Стабилизатор	I II	0,0346 0,0108	2	0,0173 0,0054	0,28 0,12
C	Количество ТОСС в стабилизаторе	I II	0,3282 0,1881	2	0,1641 0,0941	2,63 2,05
D	Наполнитель	I II	0,9456 0,8452	8	0,1182 0,1057	1,89 2,30
R	Ошибка наблюдений	I II	0,7489 0,5503	12	0,0624 0,0459	
		I II	2,3705 0,8719	26		

7.3. Применение насыщенных регулярных дробных факторных планов для выделения существенных факторов

Насыщенные регулярные дробные факторные планы (см. разд. 5) зарекомендовали себя в качестве эффективного средства выделения существенных факторов при условии, что доминирующими являются линейные эффекты.

Зависимость между целевой величиной и факторами постулируется в виде линейной модели

$$y = a_0 + \sum_{i=1}^{n} a_i x_i. \qquad (7.27)$$

Насыщенные регулярные дробные факторные планы типа 2^{k-p} содержат число точек эксперимента N (на уровнях $+1$ и -1),

равное числу искомых параметров

$$N = n + 1. \tag{7.28}$$

С их помощью можно исследовать влияние $2^k - 1$ факторов, т. е. 3, 7, 15, 31, 63 и т. д. факторов. Оценки параметров a_l определяются независимо друг от друга и смешаны с взаимодействиями, если таковые имеются. Расчет параметров производится по формулам, приведенным в разд. 5. Значимость параметров (а следовательно, и факторов) проверяется с помощью t-критерия. При этом подвергается проверке справедливость неравенства

$$|\hat{a}_l| \geqslant t_{кр} s_l, \tag{7.29}$$

где $t_{кр}$ — критическое значение распределения Стьюдента для заданного уровня значимости α и соответствующего числа степеней свободы оценки дисперсии s_l $\varphi = N(\nu - 1)$ [1]. Если нулевое значение попадает в доверительный интервал для некоторого параметра, то соответствующий фактор признается несущественным, т. е. не оказывающим влияния на целевую величину.

Пример 7.1 [6]

Требуется найти факторы, оказывающие существенное влияние на процесс получения металлического титана, включающий стадию хлорирования титановых шлаков. Из теоретических соображений для исследования отобрано 15 факторов:

x_1 — расход хлора;
x_2 — температура;
x_3 — концентрация углерода в шлаке;
x_4 — концентрация TiO_2 в расплаве;
$x_5 - x_{11}$ — другие компоненты шлака;
$x_{12} - x_{15}$ — компоненты расплава.

В качестве отклика рассматривалась скорость хлорирования.

Решение

Для отбора существенных факторов применялся дробный факторный план типа 2^{15-11}, содержавший 16 экспериментальных точек ($N = n + 1$). Уровни факторов и результаты измерений в экспериментах (план был реализован дважды) представлены в табл. 7.17, в которую также включены рассчитанные значения параметров. Для применения t-критерия из таблиц получено значение $t_{кр} = 2{,}12$ при $\alpha = 0{,}05$ и $\varphi = 16$ степенях свободы. Факторы x_3, x_7, x_9, x_{10} и x_{14} оказались незначимыми, так как

[1] Здесь ν — число повторений наблюдений в каждой точке плана. — *Прим. ред.*

Таблица 7.17

План и результаты эксперимента (пример 7.1)

	Cl	T	C	t	Компоненты шлака							Компоненты расплава				Скорость хлорирования		
					Mg	Ca	Fe^{II}	Fe^{III}	$Fe^{III}Si$	Al	Mn	Ca	Fe	Na	1:2			
Основной уровень	16,6	750	3,5	2	5,65	1,0	8,3	1	6,65	5,7	4,7	7	7	8	1:1			
Интервал варьирования	5,2	50	1,5	1	1,35	0,5	1,7	1	1,35	1,3	1,3	3	3	2	1:0,5			
Верхний уровень	21,8	800	5	3	7,0	1,5	10,0	2	8,0	7,0	6,0	10	10	10	1:0,5			
Нижний уровень	11,4	700	2	1	4,3	0,5	6,6	0	5,3	4,4	3,4	4	4	6	1:2			
Нормированные переменные	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	$\tilde{y}^1$	$\tilde{y}^2$	$\tilde{y}$
Эксперимент																			
1	+	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+	39,2	48,5	43,85
2	+	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−	48,0	47,5	47,75
3	+	−	+	−	−	−	+	+	−	−	+	+	+	−	+	−	44,7	48,5	46,6
4	+	+	+	−	−	+	−	−	−	−	+	−	−	+	+	+	47,2	43,7	45,45
5	+	−	−	+	−	+	−	+	−	+	−	+	−	+	+	−	41,7	33,4	37,55
6	+	+	−	+	−	−	+	−	−	+	−	−	+	−	+	+	40,5	48,2	44,35
7	+	−	+	+	−	−	−	+	+	−	−	−	+	+	−	+	48,2	49,2	48,7
8	+	+	+	+	−	+	+	−	+	−	−	+	−	−	−	−	49,4	49,5	49,45
9	+	−	−	−	+	+	+	−	+	−	−	−	+	+	+	−	98,5	95,0	96,75
10	+	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	115,0	92,5	103,75
11	+	+	+	−	+	+	−	+	−	+	−	−	+	−	−	−	110,0	102,5	106,75
12	+	−	−	+	+	+	−	−	−	−	+	+	+	−	−	+	50,4	47,0	48,7
13	+	+	−	+	+	−	+	+	−	−	+	−	−	+	−	−	151,0	151,1	151,05
14	+	−	+	+	+	−	−	−	+	+	−	−	−	−	+	−	93,5	103,7	98,6
15	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	127,0	123,5	125,25
16	+	+	−	−	+	−	−	+	+	−	−	+	−	−	+	+	85,0	103,0	94,0
Регрессионные коэффициенты	75,25	8,69	3,75	1,20	28,79	8,38	−5,09	1,29	7,40	1,65	1,66	−5,12	−3,71	7,78	−0,68	−4,99			
Оценка наблюдений	$s_e^2 = 39{,}85$																		

для них

$$|\hat{a}_i| < t_{кр} s_i.$$

Поэтому число факторов было сокращено с 15 до 10.

7.4. Выбор существенных факторов с помощью планов Плакетта — Бермана

Применение насыщенных регулярных дробных факторных планов, т. е. полуреплик, четвертьреплик и т. д., оказывается возможным лишь для относительно небольшого числа различных наборов факторов ($n = 3, 7, 15, 31, 63$ и т. д.). Использование же ненасыщенных дробных факторных планов, как правило, бывает не очень эффективным, так как число проводимых экспериментов значительно превышает число определяемых параметров. Например, при исследовании действия 16 факторов нужно применять дробный факторный план типа 2^{16-11}, который состоит из 32 экспериментов.

Плакетт и Берман [7] расширили класс насыщенных ортогональных планов эксперимента за счет конструирования специальных матриц плана. Число экспериментов в этих матрицах кратно четырем ($N = 4k$), и с их помощью можно исследовать влияние ($4k-1$) факторов ($k = 2(1)25$, $k \neq 23$). При $4k = 2^n$ планы Плакетта — Бермана совпадают с насыщенными регулярными дробными факторными планами. Поскольку эти планы являются ортогональными, линейные эффекты факторов находятся независимо друг от друга.

В табл. 7.18 б приведены первые строки матриц планов, содержащих от 8 до 72 экспериментов. Факторы варьируются на уровнях $+1$ ($+$) и -1 ($-$). Полные матрицы планов конструируются следующим образом: исходя из заданной первой строки, вторую и последующие строки получают путем сдвига всех элементов предыдущей строки на одну позицию вправо (или влево) и перестановки последнего (первого) элемента на первую (последнюю) позицию. Этот процесс повторяется ($N-2$) раз. Последняя строка плана состоит только из элементов -1 ($-$). Матрица плана имеет размерность $N(N-1)$.

Примеры планов

Чтобы получить план Плакетта — Бермана для $N = 16$, выбирается первая строка плана из табл. 7.18 б и применяется сформулированное правило (см. табл. 7.18 а).

Для $N = 28$ используются приведенные в табл. 7.18 б блоки ABC, которые сдвигаются аналогично элементам, т. е. матрица

плана получается в виде

$$\begin{matrix} A & B & C \\ C & A & B \\ B & C & A \end{matrix}$$

Кроме того, добавляется еще строка с элементами (—).

Для $N = 52$ применяют заданную первую строку и столбец, который ей идентичен, и по тому же правилу сдвигают блоки $ABCDE$. Матрица плана завершается строкой, состоящей только из элементов (—).

Обработка результатов эксперимента

Вслед за реализацией плана эксперимента производится обработка его результатов, которая состоит из следующих операций:

1. Расчет эффектов отдельных факторов.

Оценка эффекта B_l равна разности между суммами значений целевой функции для фактора x_l на уровнях +1 и —1, поделенной на $N/2$;

$$\hat{B}_l = \frac{\sum_{i=1}^{N} \tilde{y}_i x_{il}}{N/2}. \qquad (7.30)$$

Из матрицы плана следует, что оценки эффектов могут быть рассчитаны независимо друг от друга (свойство ортогонально-

Таблица 7.18а

N	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
1	+	+	+	+	−	+	−	+	+	−	−	+	−	−	−
2	+	+	+	−	+	−	+	+	−	−	+	−	−	−	+
3	+	+	−	+	−	+	+	−	−	+	−	−	−	+	+
4	+	−	+	−	+	+	−	−	+	−	−	−	+	+	+
5	−	+	−	+	+	−	−	+	−	−	−	+	+	+	+
6	+	−	+	+	−	−	+	−	−	−	+	+	+	+	−
7	−	+	+	−	−	+	−	−	−	+	+	+	+	−	+
8	+	+	−	−	+	−	−	−	+	+	+	+	−	+	−
9	+	−	−	+	−	−	−	+	+	+	+	−	+	−	+
10	−	−	+	−	−	−	+	+	+	+	−	+	−	+	+
11	−	+	−	−	−	+	+	+	+	−	+	−	+	+	−
12	+	−	−	−	+	+	+	+	−	+	−	+	+	−	−
13	−	−	−	+	+	+	+	−	+	−	+	+	−	−	+
14	−	−	+	+	+	+	−	+	−	+	+	−	−	+	−
15	−	+	+	+	+	−	+	−	+	+	−	−	+	−	−
16	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−

$N = 8$

$N = 12$

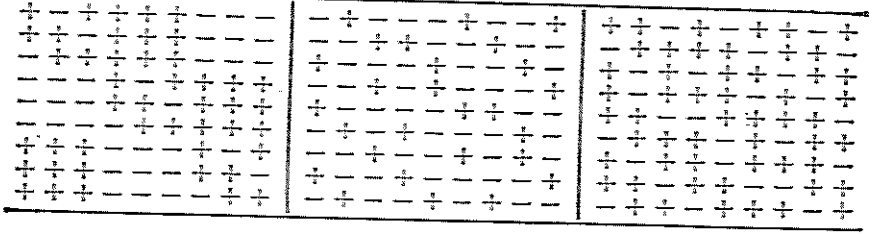
$N = 16$

$N = 20$. 

$N = 24$ 

$N = 28$ *Первые девять строк*

A *B*

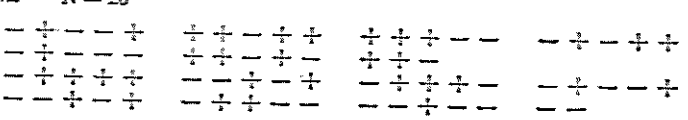


$N = 32$ 

$N = 36$ 

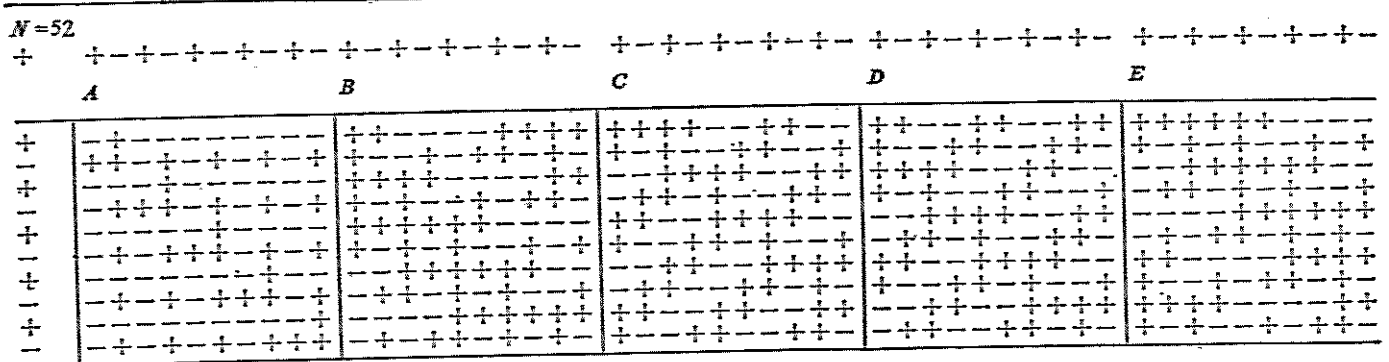
$N = 40$ *Удвоение плана* $N = 20$

$N = 44$ 

$N = 48$ 

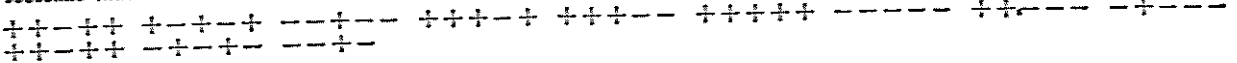
Продолжение табл. 7.18 б

$N=52$



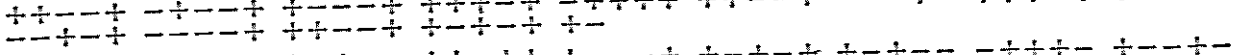
$N=56$ *Удвоение плана* $N=28$

$N=60$

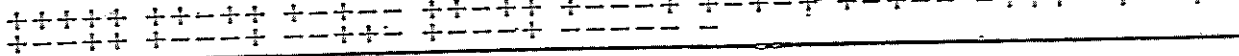


$N=64$ *Удвоение плана* $N=32$

$N=68$



$N=72$



сти). Значения $\hat{a}_l$ равны половинам соответствующих оценок эффектов.

2. Проверка значимости параметров.

Для выявления существенных факторов используется t-критерий и проверяется условие

$$|\hat{a}_l| \geqslant t_{кр} s_l, \tag{7.31}$$

где $t_{кр}$ — критическое значение t-распределения для уровня значимости α и φ степеней свободы; s_l^2 — оценка дисперсии коэффициента $\hat{a}_l$.

Дисперсия ошибок наблюдений оценивается с помощью специальных экспериментов, например дублированием наблюдений или введением в план фиктивных факторов от x_{l+1} до x_{N-1}. Если, например, план строится для анализа 12 факторов, то можно добавить к нему 3 фиктивных фактора и применить план типа $N = 16$. Эффекты этих фиктивных переменных будут равны нулю лишь в том случае, если не имеется взаимодействий и измерения являются абсолютно точными. Поскольку на практике это обычно не выполняется, их можно использовать для расчета оценки дисперсии наблюдений.

Обозначим

$$s_l^2 = 4k\,(\hat{a}_{l+1}^2 + \hat{a}_{l+2}^2 + \ldots + \hat{a}_{N-1}^2)/(4k - l - 1) \tag{7.32}$$

и

$$s_l = \sqrt{s_l^2}. \tag{7.33}$$

Уровень значимости обычно выбирают равным $\alpha = 0{,}05$ и из таблиц t-распределения находят значение $t_{0{,}05}[4k - l - 1]$. Дисперсия оценок параметров $\hat{a}_l$ равна [1]

$$s_{\hat{a}}^2 = \frac{s_l^2}{4k}. \tag{7.34}$$

Значимость параметров проверяется обычным способом путем проверки неравенства

$$|\hat{a}_l| \geqslant t_{кр} s_l.$$

Пример 7.2. Применение планов Плакетта — Бермана [8]. Рассматривается процесс получения катализатора по схеме

$$M_1Cl + M_2Cl + HAN + OH^- \longrightarrow M_1M_2AN\downarrow + H_2O + Cl^-,$$

где M_1, M_2 — ионы металлов, а AN — анион. Процесс протекает в аппарате-смесителе, в котором в результате нейтрализации

[1] В (7.32) s_l^2 есть по существу остаточная дисперсия, которая используется в качестве оценки дисперсии ошибок наблюдений. — *Прим. ред.*

каталитически активных компонентов водным раствором аммиака (NH_4OH) образуется осадок M_1M_2AN, который затем отфильтровывают, промывают, высушивают и смешивают с другими компонентами. На активность катализатора могут оказывать влияние 12 факторов. Чтобы применить план Плакетта—Бермана для выделения существенных переменных, добавим к ним три фиктивных фактора, в результате чего получаем следующий список факторов:

x_1 — значение pH раствора, при котором происходит образование осадка;
x_2 — соотношение ионов металлов в растворе на выходе;
x_3 — температура процесса образования осадка;
x_4 — избыток анионов;
x_5 — объемная скорость добавления кислоты;
x_6 — источник ионов металла M_2;
x_7 — фиктивный фактор;
x_8 — исходное состояние, определяемое условием 1 или 2;
x_9 — время образования осадка;
x_{10} — источник ионов OH^- (либо водный раствор аммиака NH_4OH, либо водный раствор смеси аммиака и едкого натра $NH_4OH/NaOH$);
x_{11} — фактор, определяющий условия (1 или 2) грануляции выпавшей в осадок массы катализатора;
x_{12} — количество компонента x, смешиваемого с массой катализатора перед грануляцией;
x_{13} — количество компонента y, смешиваемого с массой катализатора перед грануляцией;
x_{14} — фиктивный фактор;
x_{15} — фиктивный фактор.

Для решения задачи применяется план Плакетта—Бермана с $N = 16$ экспериментами (см. табл. 7.18а).

В табл. 7.19 приведены уровни факторов и эффекты $\hat{B}$ (или коэффициенты $\hat{a}$), вычисленные в соответствии с уравнением (7.30). Для расчета дисперсии из (7.32) получаем выражение

$$s_l^2 = 4k(\hat{a}_7^2 + \hat{a}_{14}^2 + \hat{a}_{15}^2)/(4k - l - 1),$$

где $k = 4$, $l = 12$;

$$s_l^2 = 16(0{,}116^2 + 3{,}974^2 + 0{,}354^2)/3 \approx 5{,}34 \times 16;$$

$$s_{\hat{l}}^2 = \frac{s_l^2}{4k}; \quad s_{\hat{l}} = 2{,}31.$$

Значимость параметров проверяем по (7.31), где для $\alpha = 0{,}05$ и $\varphi = 3$ из табл. 15.3 находим $t_{кр} = 3{,}18$ и

$$|\hat{a}_l| \geqslant 3{,}18 \times 2{,}31 \approx 7{,}35.$$

Таблица 7.19

	Уровень фактора		$\hat{a}_l$	Значимость
	−	+		
x_1	Нижняя граница	Верхняя граница	−19,4	Да
x_2	» »	» »	5,3	Нет
x_3	20°С	30°С	6,6	»
x_4	5%	20%	2,6	»
x_5	2 л/ч	3 л/ч	0,5	»
x_6	1	2	1,2	»
x_7	—	—	0,116	—
x_8	1	2	2,3	Нет
x_9	0	20 ч	0,3	»
x_{10}	NH_4^+	$NH_4^+/Na+$	−7,8	Да
x_{11}	1	2	1,5	Нет
x_{12}	0	x	2,2	»
x_{13}	0	y	−1,5	»
x_{14}	—	—	3,974	—
x_{15}	—	—	0,354	—

В последнем столбце табл. 7.19 содержатся результаты проверки значимости. Существенными оказались лишь факторы x_1 (значение pH) и x_{10} (источник ионов OH^-).

7.5. Метод случайного баланса

Если число факторов очень велико, имеет смысл применять сверхнасыщенные планы, чтобы число экспериментов оставалось в разумных пределах.

Метод случайного баланса разработан Сатерзуайтом [9] и предназначен для выделения существенных факторов из множества переменных и их парных взаимодействий. При этом предполагается, что из всего множества факторов только небольшое их число оказывает существенное влияние на целевую функцию. Факторы с незначительным воздействием можно назвать „шумом". Сила воздействия факторов должна убывать по закону, близкому к экспоненциальному, как показано на фиг. 7.1.

Планы эксперимента для метода случайного баланса являются сверхнасыщенными ($\varphi < 0$), т. е. число всех изучаемых факторов превышает число экспериментов. Поскольку по предположению лишь небольшая часть всех факторов является существенной, то после выделения этих факторов можно по результатам реализованных экспериментов произвести также их количественную оценку. В самом деле, относительно существенных факторов план эксперимента может оказаться насыщенным или даже ненасыщенным.

Предполагается, что математическая модель объекта включает линейные эффекты и парные взаимодействия факторов

$$\tilde{y} = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n + a_{12}x_1x_2 + a_{13}x_1x_3 + \ldots$$
$$\ldots + a_{n-1,\,n}x_{n-1}x_n + \tilde{e}, \quad (7.35)$$

где a_i — коэффициенты при линейных членах $i = 1, 2, \ldots, n$; a_0 — свободный член; a_{ij} $(i, j = 1\,(1)\,n,\ i \neq j)$ — коэффициенты при парных взаимодействиях; e — ошибка наблюдений.

Задача состоит в таком изменении модели, чтобы она содержала лишь существенные факторы:

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_kx_k + a_{12}x_1x_2 + \ldots$$
$$\ldots + a_{k-1,\,k}x_{k-1}x_k + \tilde{e}, \quad (7.36)$$

где k — число существенных факторов $(k < n)$; e — ошибка наблюдений, включающая также влияние отброшенных членов.

Целью эксперимента в методе случайного баланса является упорядочение факторов по степени их влияния на целевую величину (фиг. 7.1) и оценка коэффициентов модели (7.36).

Построение плана эксперимента в этом методе можно производить либо с применением случайных чисел для установления последовательности уровней факторов в столбцах матрицы плана („чистый" случайный баланс), либо путем случайного смешивания дробных факторных планов. В первом случае число опытов может быть произвольным, но по крайней мере равным числу ожидаемых значимых эффектов. Во втором случае N равно некоторой степени числа 2. Мы здесь подробнее опишем второй метод (случайное смешивание дробных факторных планов).

Все факторы разбиваются на группы (обычно от 4 до 6 факторов в группе), и для каждой из них выбирается дробный факторный план по возможности меньшего объема (см. разд. 5). Смешивание этих планов производится случайным образом с применением случайных чисел.

В некоторых задачах имеется априорная информация (например, полученная путем опроса экспертов) о том, какие факторы являются доминирующими. Тогда целесообразно так составить план эксперимента, чтобы первую его часть составил дробный факторный план (без случайного смешивания), построенный только для этих факторов.

Если априорное упорядочение совпадает с апостериорным, то уже предварительный эксперимент позволяет получить несмешанные оценки коэффициентов при значимых факторах. При использовании случайных выборок из полного факторного плана оценки параметров будут случайным образом смешаны друг с другом. То же самое будет иметь место и при использо-

вании случайных чисел для выбора уровней переменных в матрице плана.

Обработка результатов эксперимента достаточно проста и поясняется нижеследующим примером, охватывающим все операции метода случайного баланса.

Пример 7.3 [10][1]

Теоретический анализ показывает, что на скорость реакции в некотором процессе могут оказывать воздействие 15 факторов $x_1, \ldots, x_{15}$. Это значит, что в модель может войти до 15 линейных эффектов и до 105 парных взаимодействий. Эти факторы должны варьироваться на двух уровнях: нижнему уровню соответствует меньшее значение фактора ($x_{мин}$), верхнему — большее значение ($x_{макс}$). Эти значения выбирают так, чтобы их разность была по возможности наибольшей; следовательно, они должны находиться на границах области планирования. Целесообразно применять нормирование факторов по формуле

$$x_l = \frac{x_l^* - x_{l0}^*}{L_l}, \qquad (7.37)$$

где x_l^* — фактор в натуральном масштабе; x_{l0}^* — основной (нулевой) уровень для фактора в натуральном масштабе; L_l — интервал варьирования, $L_l = (x_{l\,макс}^* - x_{l\,мин}^*)/2$. При этом получаем, что $x_{l\,мин} = -1$ и $x_{l\,макс} = +1$. В табл. 7.20 представлены значения факторов в натуральном масштабе.

Таблица 7.20

Границы области планирования для 15 факторов (пример 7.3)

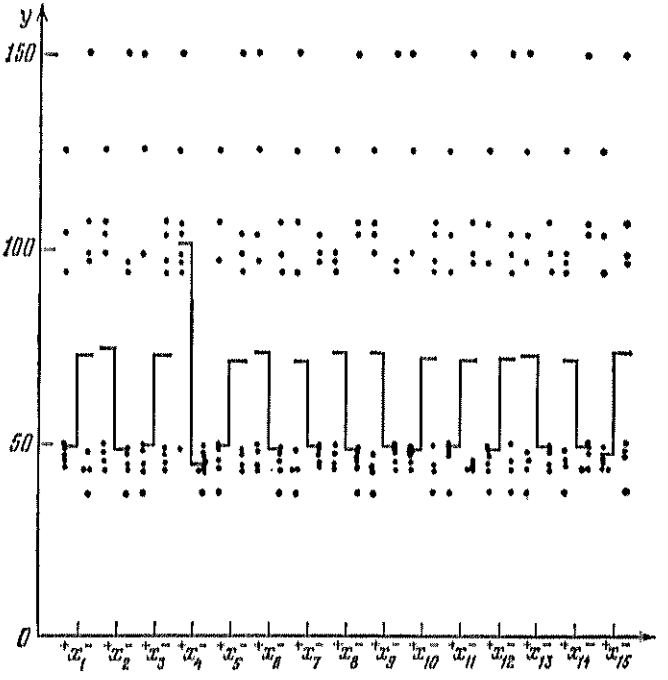
Факторы		x_1^*	x_2^*	x_3^*	x_4^*	x_5^*	x_6^*	x_7^*	x_8^*
Уровни	+1	21,8	800	5	3	7	1,5	10	2
	−1	11,4	700	2	1	4,3	0,5	6,6	0
Факторы		x_9^*	x_{10}^*	x_{11}^*	x_{12}^*	x_{13}^*	x_{14}^*	x_{15}^*	
Уровни	+1	8	7	6	10	10	10	1:0,5	
	−1	5,3	4,4	3,4	4	4	6	1:2	

[1] Приведенный здесь пример лишь частично совпадает с примером из работы [10], отличаясь от последнего как исходными данными, так и численными результатами анализа. — *Прим. ред.*

Для построения плана эксперимента мы применим второй метод, т. е. смешивание случайных выборок из дробных факторных планов. Разобьем факторы на следующие группы:

1. x_1, x_2, x_3, x_4,
2. x_5, x_6, x_7, x_8,
3. x_9, x_{10}, x_{11}, x_{12},
4. x_{13}, x_{14}, x_{15}

и используем планы эксперимента типа 2^4 (см. разд. 5). С помощью случайных чисел из матрицы плана 2^4 для каждой группы факторов случайным образом выбираются строки (16 экспери-



Фиг. 7.3. Диаграмма рассеяния для всех факторов (пример 7.3).

ментов — 16 строк). В нашем примере были получены следующие последовательности строк:

группа 1: 10,6, 3,16, 4,15, 14,1, 5,9, 7,13, 2,8, 12,11;
группа 2: 13,3, 7,16, 2,9, 2,16, 6,12, 13,7, 9,12, 3,6;
группа 3: 5,10, 15,16, 3,4, 15,4, 6,9, 9,3, 16,5, 6,10;
группа 4: 7,15, 11,8, 16,3, 5,13, 4,12, 6,2, 2,1, 14,9.

Матрица плана и результаты эксперимента представлены в табл. 7.21.

Таблица 7.21

План и результаты эксперимента для метода случайного баланса (пример 7.3)

Номер эксперимента	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	$\bar{y}$	$\bar{y}^{I}$	$\bar{y}^{II}$	$\bar{y}^{III}$	$\bar{y}^{IV}$	$\bar{y}^{V}$	$\bar{y}^{VI}$
1	+	−	−	+	−	−	+	+	−	−	+	−	−	+	+	94,00	36,0	36,0	22,0	42,0	33,0	33,0
2	+	−	+	−	−	+	−	−	+	−	−	+	−	+	+	44,35	44,4	44,4	27,4	38,4	37,4	33,4
3	−	+	−	−	−	+	+	−	−	+	+	+	−	+	−	48,05	48,1	41,1	10,1	19,1	27,1	23,1
4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	125,25	67,3	45,3	14,3	44,3	43,3	35,3
5	+	+	−	−	+	−	−	−	−	+	−	−	+	+	+	45,45	45,5	23,5	23,5	44,5	35,5	31,5
6	−	+	+	+	−	−	−	+	+	+	−	−	−	+	−	98,60	40,6	33,6	33,6	33,6	33,6	25,6
7	+	−	+	+	+	−	−	−	−	+	+	+	−	−	+	48,70	−9,3	−9,3	−9,3	20,7	19,7	15,7
8	−	−	−	+	+	+	+	+	+	+	−	−	−	−	+	43,85	43,9	43,9	12,9	33,9	33,9	25,9
9	−	−	+	−	+	−	+	−	+	−	+	−	+	+	−	37,55	37,6	22,6	8,6	27,6	27,6	23,6
10	−	−	−	+	+	+	−	+	−	−	−	+	+	+	−	96,75	38,8	23,8	6,8	16,8	24,8	24,8
11	−	+	+	−	−	−	+	+	−	−	−	+	+	−	+	43,60	43,6	21,6	7,6	18,6	26,6	26,6
12	−	−	+	+	−	+	+	−	−	+	−	−	+	−	−	150,05	92,1	77,1	46,1	46,1	46,1	42,1
13	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−	47,75	47,8	32,8	32,8	41,8	40,8	32,8
14	+	+	+	−	+	+	−	+	−	−	+	+	−	−	−	49,45	49,5	42,5	25,5	44,5	35,5	35,5
15	+	+	−	+	−	+	−	−	+	−	+	−	+	−	+	103,75	45,8	23,8	6,8	26,8	17,8	13,8
16	−	+	−	+	+	−	+	−	+	−	−	+	−	−	−	106,75	48,8	41,8	27,8	37,8	45,8	41,8

На следующем этапе строится диаграмма рассеяния. Для этого на графике над отдельными факторами, варьируемыми на двух уровнях $+1$ и -1, наносятся значения целевой величины и определяются медианы для обоих уровней каждого фактора (фиг. 7.3). Чем больше разность между медианами, тем больше воздействие соответствующего фактора. Из диаграммы видно, что наибольшее воздействие оказывает фактор x_4. Визуальное выявление доминирующих факторов по величине медиан можно

Таблица 7.21а

	$+x_{13}$		$-x_{13}$	
	$+x_2$	$-x_2$	$+x_2$	$-x_2$
$+x_{14}$	67,3 45,5 112,8 $\bar{y}_1^m = 56,4$	37,6 38,8 76,4 $\bar{y}_2^m = 38,2$	48,1 40,6 88,7 $\bar{y}_3^m = 44,35$	36,0 44,4 80,4 $\bar{y}_4^m = 40,2$
$-x_{14}$	43,6 45,8 89,4 $\bar{y}_5^m = 44,7$	47,8 92,1 139,9 $\bar{y}_6^m = 69,95$	49,5 48,8 98,3 $\bar{y}_7^m = 49,15$	43,9 −9,275 34,625 $\bar{y}_8^m = 17,312$

$$\hat{B}_2 = \frac{194,6 - 165,662}{4} = 7,25$$

$$\hat{B}_{13} = \frac{209,25 - 151,012}{4} = 14,56$$

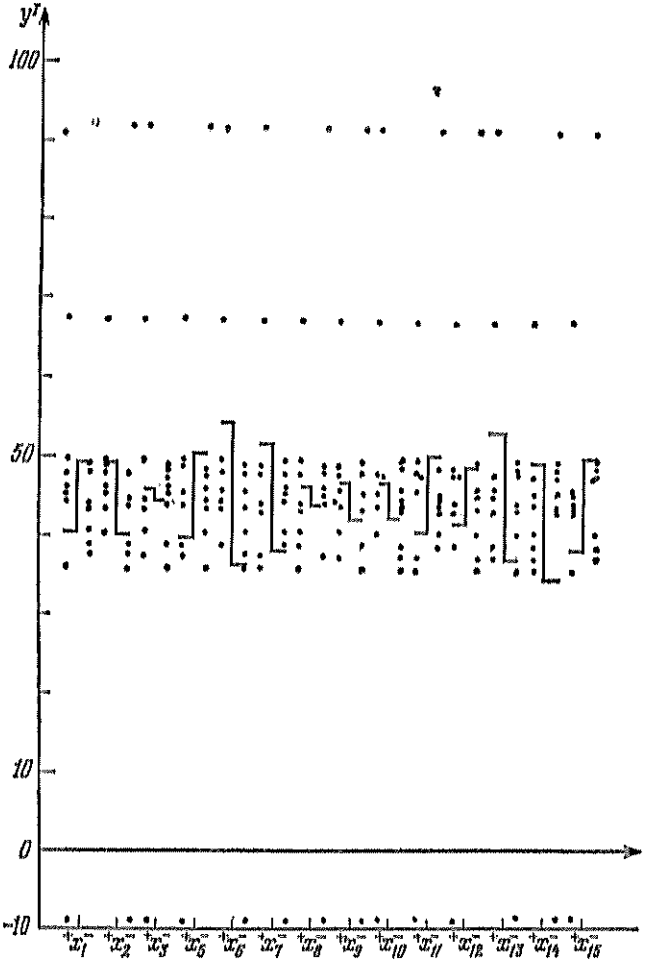
$$\hat{B}_{14} = \frac{179,15 - 181,112}{4} = -0,94$$

заменить методом, основанным на определении числа „выделяющихся" точек.

„Выделяющимися" для некоторого уровня фактора (например, для $x_l = +1$) называются такие значения $\bar{y}$, которые больше наибольшего (или меньше наименьшего) значения $\tilde{y}$ при другом уровне фактора (т. е. для $x_l = -1$). Общее число таких точек можно использовать в качестве характеристики значимости факторов.

Для расчета оценок вкладов и коэффициентов определяют среднее значение всех $\bar{y}$, для которых фактор x_4 находится на уровне $+1$, и вычитают из него среднее значение $\tilde{y}$ на уровне -1. Оценка B_4 вклада фактора x_4 равна $B_4 = 57,575$. Величина $\hat{B}_4$ позволяет получить оценку параметра $\hat{a}_4$: $\hat{a}_4 \approx B_4/2 \approx 29$.

Для нахождения других существенных факторов необходимо устранить влияние фактора x_4 на $\tilde{y}$. С этой целью вычитают величину $2\hat{a}_4 = 58$ из всех $\tilde{y}$, для которых x_4 находился на



Фиг. 7.4. Диаграмма рассеяния для примера 7.3 после устранения влияния фактора x_4.

уровне $+1$, и тем самым эти значения освобождаются от влияния фактора x_4. После этой операции получают новый вектор результатов эксперимента $\tilde{y}^1$ и строят диаграмму рассеяния

(фиг. 7.4), в которой фактор x_4 уже не оказывает влияния [1]. Изучая эту диаграмму, мы видим, что наибольшее влияние имеют факторы x_2, x_{13} и x_{14} [2].

При расчете оценок $\hat{a}_2$, $\hat{a}_{13}$ и $\hat{a}_{14}$ удобно пользоваться табл. 7.21а, построенной для факторов x_2, x_{13} и x_{14}. Способ построения таблицы достаточно очевиден. Для каждой группы чисел рассчитывается среднее значение $\tilde{y}_l$. Оценки вкладов факторов определяются разностью между суммами средних значений для уровней „плюс" и „минус". Например, для B_2

$$\hat{B}_2 = \frac{\tilde{y}_1^m + \tilde{y}_3^m + \tilde{y}_5^m + \tilde{y}_7^m}{4} - \frac{\tilde{y}_2^m + \tilde{y}_4^m + \tilde{y}_6^m + \tilde{y}_8^m}{4}, \tag{7.38}$$

$$\hat{B}_{13} = \frac{\tilde{y}_1^m + \tilde{y}_2^m + \tilde{y}_5^m + \tilde{y}_6^m}{4} - \frac{\tilde{y}_3^m + \tilde{y}_4^m + \tilde{y}_7^m + \tilde{y}_8^m}{4} \tag{7.39}$$

и т. д. Оценки коэффициентов находятся по формуле

$$\hat{a}_l = \frac{\hat{B}_l}{2}.$$

Рассчитанные таким образом величины лучше характеризуют существенность факторов, чем разность меднан. Для оценок $\hat{a}_2$, $\hat{a}_{13}$ и $\hat{a}_{14}$ получены следующие значения:

$$\hat{a}_2 = 3{,}61,$$
$$\hat{a}_{13} = 7{,}28,$$
$$\hat{a}_{14} = -0{,}245.$$

Значение $\hat{a}_{14} \approx -0{,}25$ мало по сравнению с $\hat{a}_2$ и $\hat{a}_{13}$, и поэтому $\tilde{y}^{\mathrm{I}}$ будет освобождаться только от влияния факторов x_2 и x_{13}. Это достигается за счет вычитания из тех $\tilde{y}^{\mathrm{I}}$, для которых $x_2 = +1$ или $x_{13} = +1$, величин $\hat{B}_2$ или $\hat{B}_{13}$ соответственно. После этого получают вектор результатов $\tilde{y}^{\mathrm{II}}$ и строят новую диаграмму рассеяния (фиг. 7.5). Из этой диаграммы видно, что наибольшее влияние имеют факторы x_6 и x_7. Для расчета оценок параметров применяют табл. 7.22 с двумя входами (x_6 и x_7). Эта таблица строится и используется аналогично табл. 7.21а. В результате расчета получаем

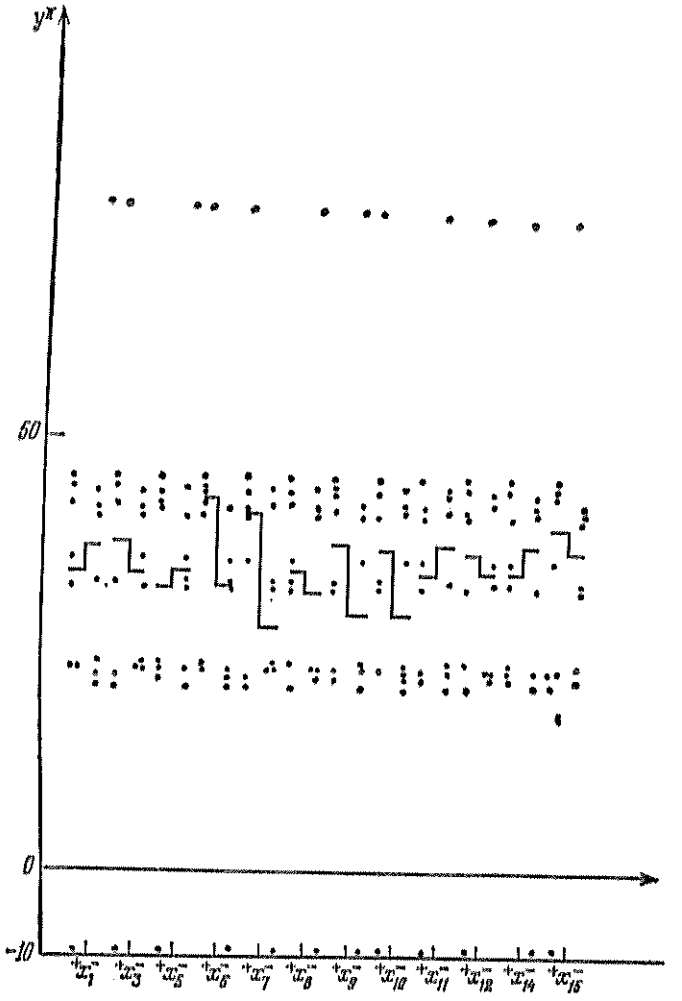
$$\hat{a}_6 = 8{,}71,$$
$$\hat{a}_7 = 7{,}14.$$

---

[1] Значения $\tilde{y}^{\mathrm{I}}$ в табл. 7.21 даются округленно. — *Прим. ред.*

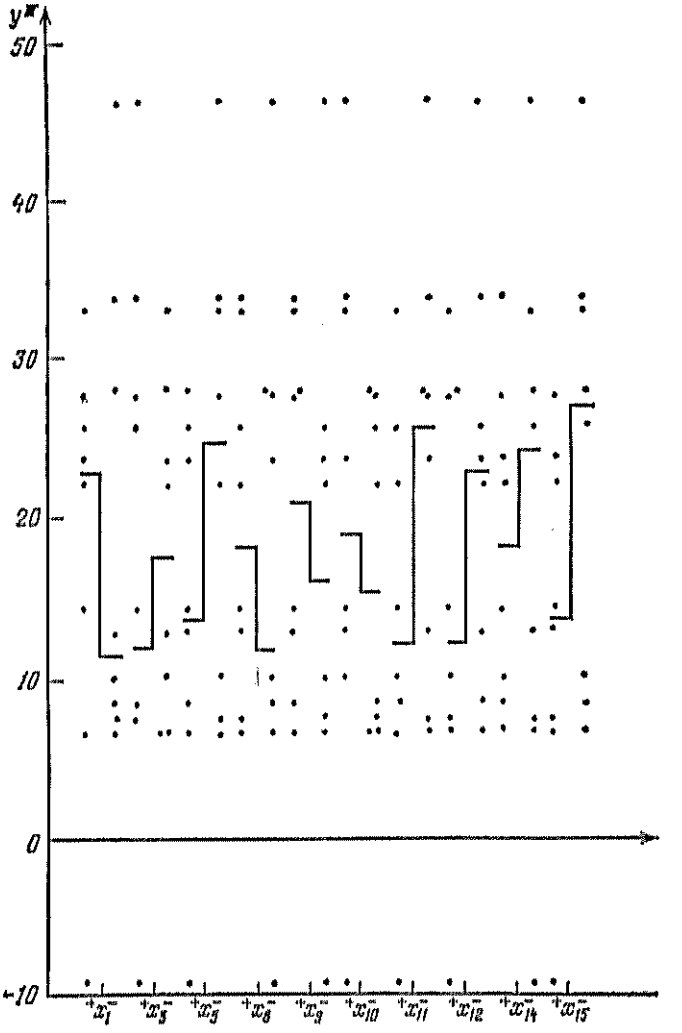
[2] Судя по диаграмме рассеяния, фактор 13 вряд ли можно на этом этапе выделить как существенный. Возможно, правильнее было бы вместо фактора 13 выбрать фактор 6. — *Прим. ред.*

Для выделения существенных факторов из числа оставшихся необходимо освободить вектор результатов $\tilde{y}^{\text{II}}$ от влияния факторов x_0 и x_7. Этого можно достичь, вычитая из $\tilde{y}^{\text{II}}$ значения



Фиг. 7.5. Диаграмма рассеяния для примера 7.3 после устранения влияния факторов x_2, x_4 и x_{13}.

$\hat{B}_0$ и $\hat{B}_7$ при $x_0 = +1$ или $x_7 = +1$ соответственно. Строится новая диаграмма рассеяния для полученного вектора $\tilde{y}^{\text{III}}$ (фиг. 7.6). Визуальная обработка и расчет медиан указывают, что среди оставшихся факторов доминирующими являются x_6, x_{11} и x_{15}.

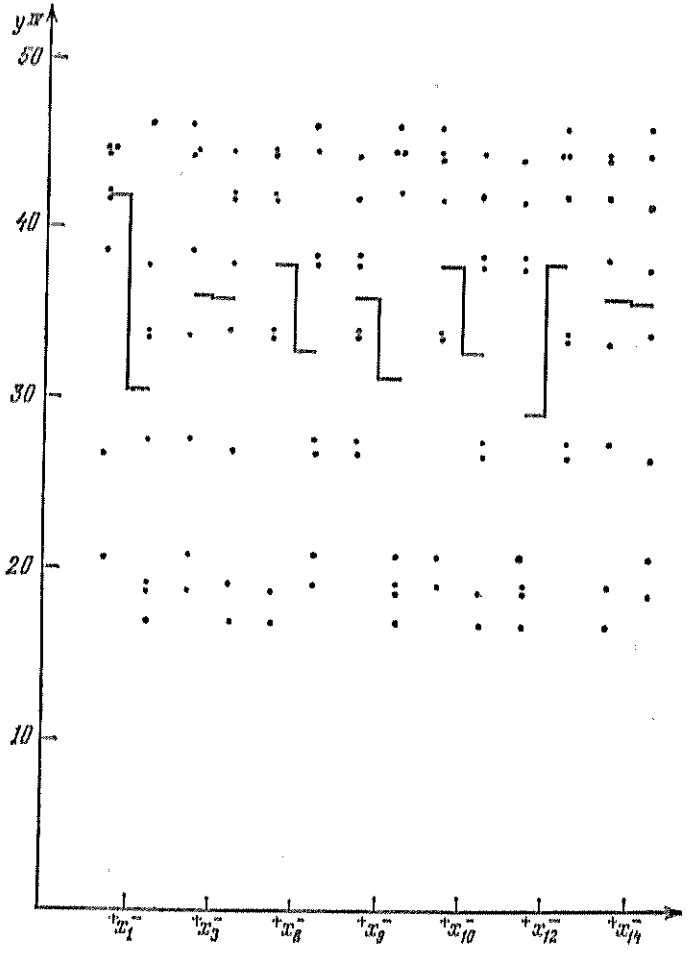


Фиг. 7.6. Диаграмма рассеяния для примера 7.3 после устранения влияния факторов x_2, x_4, x_8, x_7 и x_{13}.

Далее заполняется новая таблица для расчета оценок $\hat{a}_5$, $\hat{a}_{11}$ и $\hat{a}_{15}$ (см. табл. 7.23). Они оказались равными

$$\hat{a}_5 = -4{,}75,$$
$$\hat{a}_{11} = -4{,}66,$$
$$\hat{a}_{15} = -5{,}41.$$

По уже известному правилу исключается влияние отобранных факторов на отклик, и в итоге получается вектор результатов $\tilde{y}^{IV}$, для которого вновь строится диаграмма рассеяния (фиг. 7.7).



Фиг. 7.7. Диаграмма рассеяния для примера 7.3 после устранения влияния факторов x_2, x_4, x_5, x_6, x_7, x_{11}, x_{13} и x_{15}.

Визуально выделяются доминирующие факторы x_1 и x_{12}. С помощью соответствующей таблицы, построенной для этих факторов (табл. 7.24), рассчитываются следующие оценки коэффи-

Таблица 7.22

	$+x_6$	$-x_6$
$+x_7$	41,1 45,3 43,9 77,1 207,4 $\bar{y}_1''' = 51,85$	36,0 22,6 21,6 41,8 122,0 $\bar{y}_2''' = 30,5$
$-x_7$	44,4 23,8 42,5 23,8 134,5 $\bar{y}_3''' = 33,625$	23,5 33,6 −9,3 32,8 80,6 $\bar{y}_4''' = 20,15$

$$\hat{B}_6 = \frac{85{,}475 - 50{,}65}{2} \cong 17{,}41$$

$$\hat{B}_7 = \frac{82{,}35 - 53{,}775}{2} \cong 14{,}29$$

Таблица 7.23

	$+x_5$		$-x_5$	
	$+x_{15}$	$-x_{15}$	$+x_{15}$	$-x_{15}$
$+x_{11}$	14,3 −9,3 5,0 $\bar{y}_1''' = 2,5$	8,6 25,5 34,1 $\bar{y}_2''' = 17,05$	22,0 6,8 28,8 $\bar{y}_3''' = 14,4$	10,1 32,8 42,9 $\bar{y}_4''' = 21,45$
$-x_{11}$	23,5 12,9 36,4 $\bar{y}_5''' = 18,2$	6,8 27,8 34,6 $\bar{y}_6''' = 17,3$	7,6 27,4 35,0 $\bar{y}_7''' = 17,5$	33,6 46,1 79,7 $\bar{y}_8''' = 39,85$

$$\hat{B}_5 = \frac{55{,}05 - 93{,}20}{4} = -9{,}50$$

$$\hat{B}_{11} = \frac{55{,}4 - 92{,}85}{4} = -9{,}31$$

$$\hat{B}_{15} = \frac{52{,}6 - 95{,}65}{4} = -10{,}81$$

Таблица 7.24

	$+x_1$	$-x_1$
$+x_{12}$	38,4 44,3 20,7 41,8 145,2 $\bar{y}_1^m = 36{,}3$	19,1 16,8 18,6 37,8 92,3 $\bar{y}_2^m = 23{,}075$
$-x_{12}$	26,8 44,5 44,5 42,0 157,8 $\bar{y}_3^m = 39{,}45$	46,1 27,6 33,9 33,6 141,2 $\bar{y}_4^m = 35{,}3$

$$\hat{B}_1 = \frac{75{,}75 - 58{,}375}{2} = 8{,}68$$

$$\hat{B}_{12} = \frac{59{,}375 - 74{,}75}{2} = -7{,}69$$

циентов:

$$\hat{a}_1 = 4{,}34,$$
$$\hat{a}_{12} = 3{,}84.$$

После устранения влияния факторов x_1 и x_{12} на функцию отклика получается вектор $\tilde{y}^V$. Диаграмма рассеяния для $\tilde{y}^V$ (фиг. 7.8) показывает, что влияние оставшихся факторов невелико и примерно одного порядка [1].

Таким образом, найдены следующие оценки:

$$\hat{a}_1 = 4{,}34;\ \hat{a}_2 = 3{,}62;\ \hat{a}_4 = 29;\ \hat{a}_5 = 4{,}75;\ \hat{a}_6 = 8{,}71;$$
$$\hat{a}_7 = 7{,}14;\ \hat{a}_{11} = -4{,}66;\ \hat{a}_{12} = -3{,}84;\ \hat{a}_{13} = 7{,}28;\ \hat{a}_{15} = -5{,}41.$$

В заключение этого этапа с помощью t-критерия проверяется значимость коэффициентов. Условие значимости a_l записывается

[1] Если продолжить процесс выделения существенных переменных, то, согласно фиг. 7.8, следует принять за существенные факторы x_9 и x_{10}. Исключение влияния этих факторов на выходную переменную приводит к диаграмме рассеяния, показанной на фиг. 7.9.

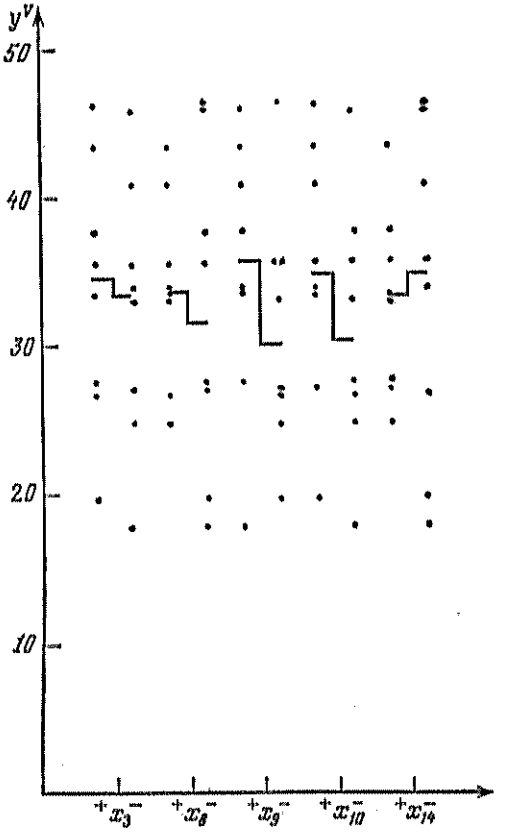
в виде

$$|\hat{a}_l| \geqslant t_{кр} s, \tag{7.40}$$

где

$$s^2 = \frac{1}{n^2} s_F^2 \sum_{l=1}^{n} \frac{1}{t_l}, \tag{7.41}$$

$t_{кр}$ — табулированное значение t-распределения для определенного уровня значимости (как правило, $\alpha = 0{,}05$) и соответст-



Фиг. 7.8. Диаграмма рассеяния для примера 7.3 после устранения влияния факторов x_1, x_2, x_4, x_5, x_6, x_7, x_{11}, x_{12}, x_{13} и x_{15}.

вующего числа степеней свободы

$$\varphi = \sum_{l=1}^{n} t_l - n,$$

l_i — число значений $\tilde{y}$ в i-й клетке таблицы; s^2_F — усредненное значение дисперсий всех клеток соответствующей таблицы:

$$s^2_F = \frac{\sum_{i=1}^{n} s_i^2 \varphi_i}{\sum_{i=1}^{n} \varphi_i}, \qquad (7.42)$$

где s_i^2 — дисперсия i-й клетки таблицы ($i = 1, 2, \ldots, n$):

$$s_i^2 = \frac{\sum_{l=1}^{l_i} (\tilde{y}_{il} - \tilde{y}_i^{(n)})^2}{\varphi_i}, \quad \varphi_i = l_i - 1. \qquad (7.43)$$

Такую проверку значимости можно произвести лишь в том случае, если каждая клетка таблицы содержит по меньшей мере два значения $\tilde{y}$.

В качестве примера приведем проверку значимости коэффициента a_1 (см. табл. 7.24):

$$s_1^2 = \frac{(38{,}4-36{,}3)^2 + (44{,}3-36{,}3)^2 + (20{,}7-36{,}3)^2 + (41{,}8-36{,}3)^2}{3} = 114,$$

$$s_2^2 = 98, \quad s_3^2 = 74, \quad s_4^2 = 60,$$

$$s_F^2 = \frac{114\times 3 + 98\times 3 + 74\times 3 + 60\times 3}{12} = 86{,}4,$$

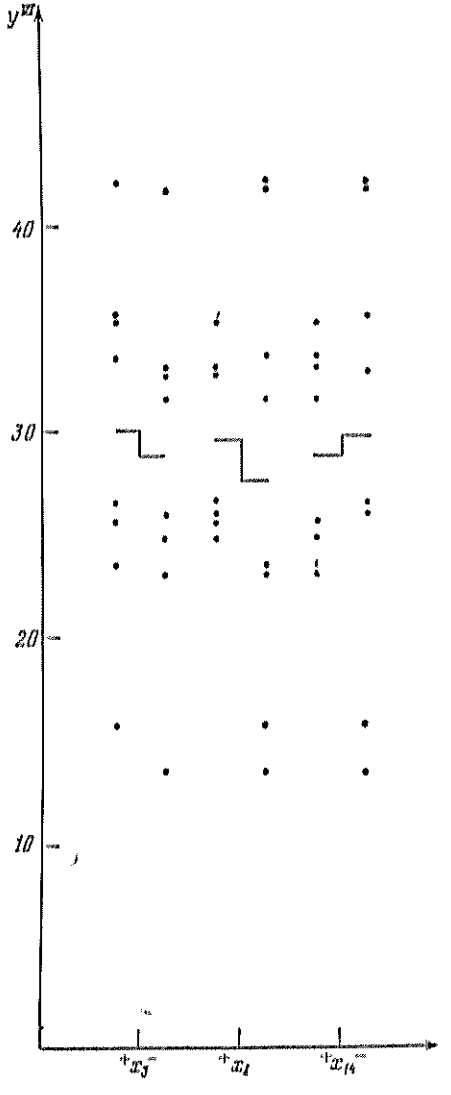
$$s^2 = \frac{1}{16}\times 86{,}4\,(4/4) = 5{,}4,$$

$$s = \sqrt{5{,}4} = 2{,}32, \quad \varphi = 12.$$

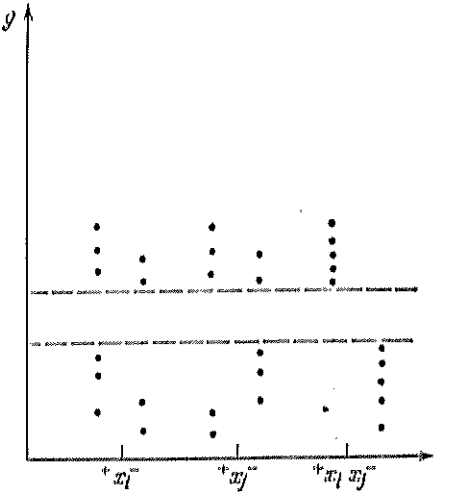
При $\alpha = 0{,}05$ критическое значение $t_{кр}$ равно 2,18 (см табл. 15,3), поэтому оценка $\hat{a}_1$ незначимо отличается от нуля.

После нахождения линейных эффектов необходимо выделить важнейшие из парных взаимодействий, общее число которых равно числу сочетаний из 15 по 2, т. е. $C^2_{15} = 105$. Построить диаграмму рассеяния для такого большого числа взаимодействий очень трудно, да и не нужно. Важнейшие парные взаимодействия можно найти с помощью эвристического визуального метода отбора, основу которого составляет следующее правило. Для того чтобы парное взаимодействие факторов x_i и x_j относилось к числу существенных, необходимо появление выделяющихся точек как на уровне $x_i x_j = +1$, так и на уровне $x_i x_j = -1$. В первом случае x_i и x_j должны иметь одинаковые знаки, а во втором — разные. Следовательно, существенным будет парное взаимодействие таких факторов, у которых число выделяющихся точек будет велико как на одинаковых, так и на разных уров-

ниях. Поэтому у таких факторов нижняя часть диаграммы должна быть похожа на зеркальное отображение верхней (фиг. 7.10).

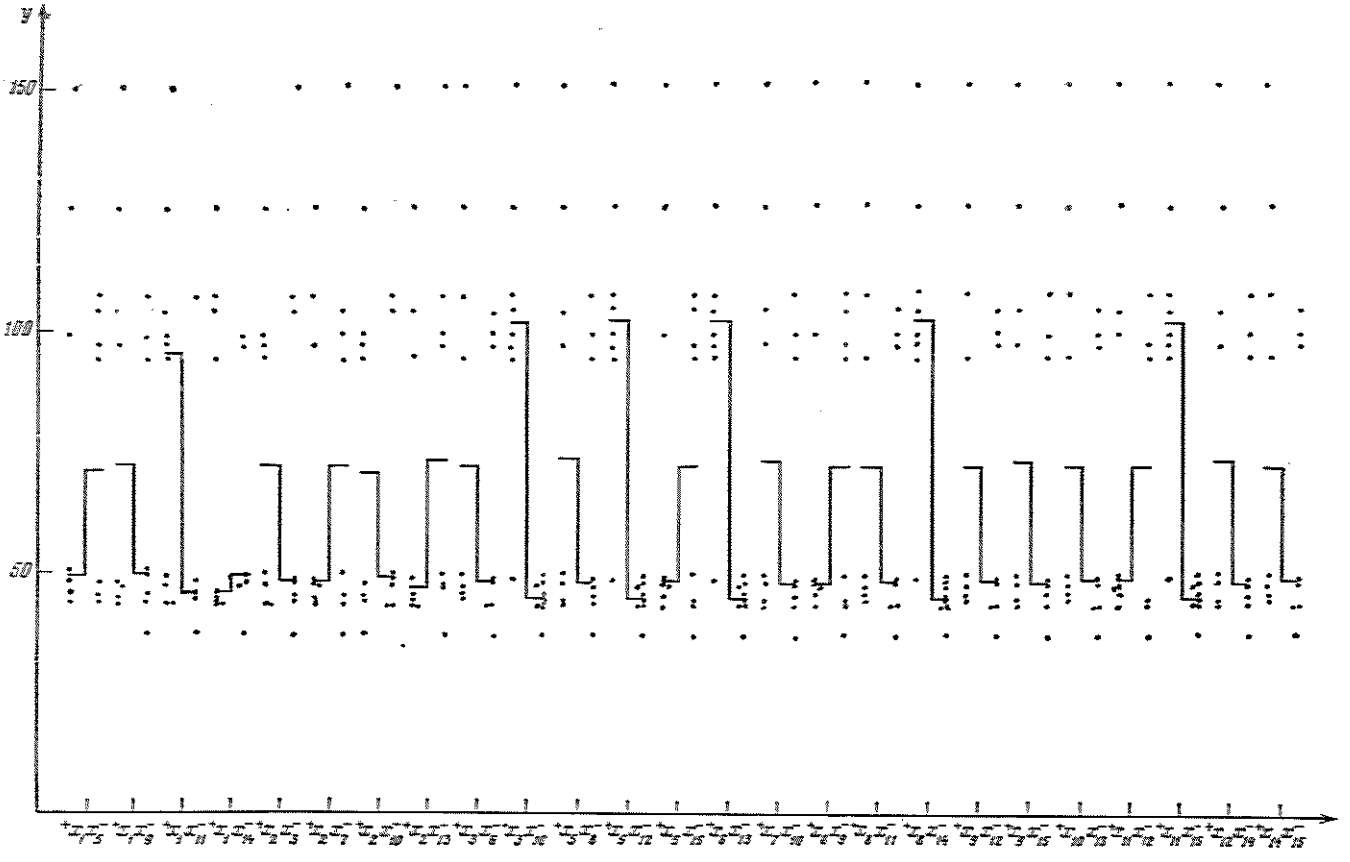


Фиг. 7.9. Диаграмма рассеяния для примера 7.3 после устранения влияния факторов x_1, x_2, x_4, x_6, x_8, x_7, x_9, x_{10}, x_{11}, x_{12}, x_{13} и x_{15}.



Фиг. 7.10. Расположение результатов эксперимента на диаграмме рассеяния для факторов x_i, x_j и их взаимодействия $x_i x_j$.

По этому правилу из диаграммы рассеяния (фиг. 7.3) визуально были выбраны парные взаимодействия, для которых была построена новая диаграмма (фиг. 7.11). Для расчета параметров надо составить соответствующие таблицы. Из-за корреляции



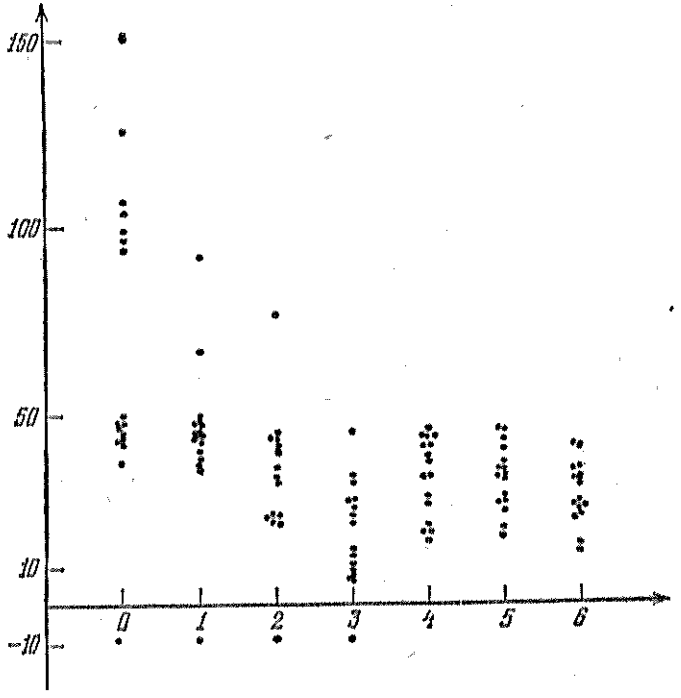
Фиг. 7.11. Диаграмма рассеяния для выбранных парных взаимодействий (пример 7.3).

между некоторыми строками не все клетки таблиц оказываются заполненными. Поэтому количественную оценку парных взаимодействий можно произвести только по медианам.

Существенными являются следующие парные взаимодействия:

$$x_1x_{11},\ x_9x_{10},\ x_5x_{12},\ x_6x_{13},\ x_8x_{14},\ x_{11}x_{15}.$$

Процесс выделения можно продолжать произвольно долго, поэтому необходимо сформировать правило остановки. Здесь можно



Фиг. 7.12. Точечная диаграмма, характеризующая изменение дисперсий результатов эксперимента (пример. 7.3).

применить F-критерий:

$$F = \frac{s_m^2}{s^2},$$

где s_m^2—оценка дисперсии результатов эксперимента относительно их арифметического среднего значения на r-м шаге процедуры, s^2—оценка дисперсии ошибок наблюдений, вычисляемая на основании результатов нескольких параллельных наблюдений. Изменение дисперсий $s_m^2(r)$ в зависимости от номера шага r процедуры выделения факторов можно проследить на фиг. 7.12.

Из этой диаграммы следует, что уже после пяти шагов эта дисперсия оказывается достаточно малой. Путем дублирования экспериментов в одной точке плана была найдена оценка $s^2 = 32\,(\varphi = 7)$. При этом для $s_m^2(5) = 74{,}4\,(\varphi_m = 12)$ получаем

$$F = \frac{74{,}4}{32} = 2{,}32 < 3{,}5 = F_{\text{кр}},\ \alpha = 0{,}05.$$

Это означает, что расчеты по данной процедуре можно прекратить, поскольку выделены все существенные факторы.

7.6. Обсуждение и сравнение методов

Методы, описанные в данном разделе, являются важным вспомогательным средством для построения моделей, предназначенных для оптимизации и управления. В основу этих методов положены следующие требования:

— выделение всех существенных факторов должно производиться при возможно меньшем числе экспериментов;

— точность оценок параметров может быть относительно невысокой.

В отличие от дисперсионного анализа, дробных факторных планов и планов Плакетта—Бермана метод случайного баланса долгое время считался эвристическим и лишь в последнее время была сделана попытка дать ему строгое теоретическое обоснование [12]. Метод случайного баланса является не слишком чувствительным, и с его помощью не всегда удается выделить все существенные факторы. Эффективность этого метода удается повысить, например, путем модификации процесса обработки с применением ЭВМ. Были предприняты попытки распространения этого метода на случай большего числа уровней и построения для него почти ортогональных планов эксперимента [11].

Помимо ранее названных свойств, можно отметить следующие достоинства, недостатки и области применения описанных методов.

1. Дисперсионный анализ

Могут быть выделены линейные эффекты и парные взаимодействия дискретно изменяющихся факторов, а также может быть проверена значимость их влияния на целевую величину. Выделяются все существенные факторы. Метод обработки экспериментальных данных достаточно прост, но требует больших затрат времени.

Рекомендуется применять вычислительные машины. Число экспериментов здесь может быть относительно велико. Применение специальных планов, например латинских и греко-латинских квадратов, позволяет уменьшить затраты на эксперимент.

2. Дробные факторные планы и планы Плакетта — Бермана

Число факторов должно быть не очень большим либо эксперименты должны быть дешевыми и не слишком длительными. Благодаря ортогональности этих планов их результаты легко поддаются обработке. Линейные эффекты рассчитываются независимо друг от друга. Факторы варьируются только на двух уровнях (если применяются планы типа 2^k).

3. Метод случайного баланса

Может быть рассмотрено большое число факторов. Не все факторы должны быть существенными. Эксперименты могут быть дорогими и длительными. Невысокая чувствительность. Простая обработка. Возможно варьирование факторов на нескольких уровнях.

ЛИТЕРАТУРА

1. Hicks C. R., Fundamental concepts in the Design of Experiments, Holt, Rinfart a. Winston, New York, 1964; русский перевод: Хикс Ч., Основные принципы планирования эксперимента, изд-во „Мир", М., 1967.
2. Ahrens H., Varianzanalyse, Akademie-Verlag, Berlin, 1967.
3. Müller P. H., Lexikon-Wahrscheinlichkeitsrechnung und mathematische Statistik, Akademie-Verlag, Berlin, 1970.
4. Scheffe H., The Analysis of variance, New York, 1959; русский перевод: Шеффе Г., Дисперсионный анализ, Физматгиз, М., 1963.
5. Маркова Е. В. и др., *Заводск. лаб.*, 35, № 7 (1969).
6. Хомяков Н. П. и др., *Заводск. лаб.*, 29, № 1 (1963).
7. Plackett R. L., Burman I. P., *Biometrika*, 33, № 4, 305 (1946).
8. Stowe R. A., Mayer R. P., *Ind. Eng. Chem.*, 58, № 2, 36 (1966).
9. Satterthwaite F. E., *Technometrics*, 1, № 2, 111 (1959).
10. Экспериментально-статистические методы получения математического описания и оптимизации сложных технологических процессов, составители Круг Г. К., Иванов А. З. и Чирков И. М., Руководящие технические материалы НИИТЭХим, вып. 2, М., 1964.
11. Booth K. H. V., *Technometrics*, 4, № 4, 489 (1962).
12. Мешалкин Л. Д., *Заводск. лаб.*, 36, № 3, 316 (1970).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Маркова Е. В., Неполноблочные планы, Препринт № 15, МГУ, 1970.
2. Маркова Е. В., Лисенков А. Н., Планирование эксперимента в условиях неоднородностей, изд-во „Наука", М., 1973.
3. Длин А. М., Математическая статистика в технике, изд-во „Советская наука", М., 1958.
4. Налимов В. В., Чернова Н. А., Статистические методы планирования экстремальных экспериментов, изд-во „Наука", М., 1965.
5. Налимов В. В., Применение математической статистики при анализе вещества, Физматгиз, М., 1960.
6. Brownlee K. A., Statistical Theory and Metodology in Science and Engineering, Wiley, New York, London, 1961.
7. Лисенков А. Н., Маркова Е. В., Адлер Ю. П., О классификации экспериментальных планов, АН СССР, Научный совет по комплексной проблеме „Кибернетика", *Информационные материалы*, 8 (45), М. (1970).

8.

АПРИОРНОЕ МОДЕЛИРОВАНИЕ

8.1. Постановка задачи априорного моделирования

Эффективность управления большими системами в значительной степени зависит от качества решений, принимаемых в сложных ситуациях на основе оценок и мнений специалистов, т. е. на основе экспертных оценок. Экспертные оценки могут явиться важным источником информации при решении задач управления, формировании целевой функции управляемых объектов, при исследовании объектов, выборе влияющих переменных, определяющих характер протекающих в них процессов, и т. д.

Экспертные оценки отражают опыт, интуицию и знания специалистов относительно исследуемого объекта и, несмотря на их субъективность, содержат полезную объективную информацию. Под априорным моделированием мы будем здесь понимать методы выявления, формализации и обработки неявной, качественной, субъективной информации, которая может содержаться в мнениях, поведении и высказываниях людей [1]. Таким образом, задачей априорного моделирования является получение количественного описания объекта управления путем обработки данных, полученных в результате направленного опроса специалистов.

Методы априорного моделирования раньше всего стали применять в психологии [1—3], позднее они нашли применение также и в социологии [4, 5]. В последнее время этим методам начали уделять большое внимание специалисты в области управления [6—12]. Объектами априорного моделирования могут быть технологические процессы, промышленные установки, отрасли производства и т. д. На фиг. 8.1 представлена структура применения методов априорного моделирования.

Исследование, проводимое группой специалистов (исследовательская группа), состоит из нескольких этапов:

1. Формулирование конкретной цели исследования.
2. Выбор экспертов, которые должны быть опрошены.
3. Выбор метода опроса P.

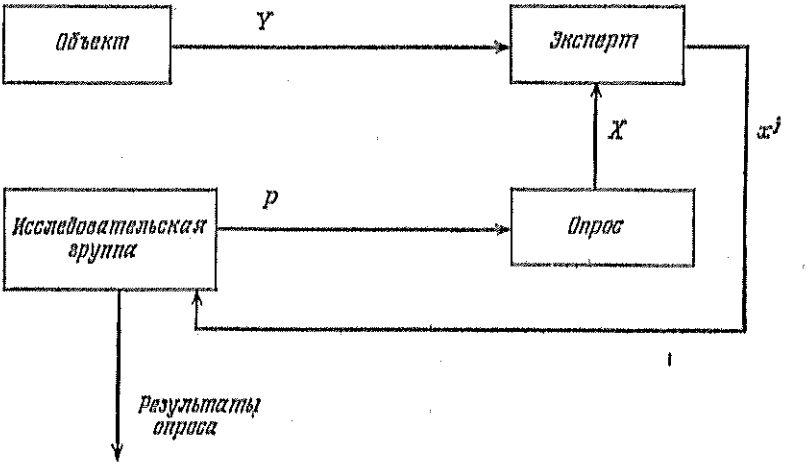
---

[1] В настоящее время методы, рассмотренные в этом разделе, чаще объединяют под названием „методы экспертных оценок".— *Прим. ред.*

4. Разработка опросного листа (анкеты). Анкета состоит из вопросов, на которые эксперты должны дать ответ в определенной форме. Ответ j-го эксперта на i-й вопрос анкеты мы будем обозначать через x_i^j.

5. Обработка результатов опроса.

Опыт показывает, что к опросу следует привлекать экспертов, принадлежащих к возможно большему числу различных направлений или научных школ в соответствующей области. Это



Фиг. 8.1. Структурная схема применения методов априорного моделирования.

позволяет рассмотреть объект с различных точек зрения и ограждает от ошибок, связанных с неправильной постановкой задачи. При составлении экспертной группы необходимо предусмотреть возможность взвешивания ответов экспертов согласно их компетентности. Учет компетентности экспертов может существенно изменить результаты обработки данных опроса [13].

Метод опроса в значительной степени определяется особенностями исследуемого процесса, числом экспертов, располагающих необходимой информацией, а также временем и средствами, имеющимися в распоряжении исследовательской группы. Под методом опроса P мы будем подразумевать метод составления анкеты (p_1), число вопросов в анкете (p_2), число повторных опросов (p_3), позволяющих скорректировать анкету на основе предыдущих опросов.

Опрос может быть очным и заочным. При заочном опросе личный контакт исследователя с экспертом отсутствует. Преимущество этого метода заключается в его простоте и дешевизне, однако этот метод дает большое число незаполненных или не-

правильно заполненных анкет. Очный опрос дает лучшие результаты, однако требует большей затраты времени и средств. Кроме того, во время личной беседы с экспертом исследователь, помимо своего желания, может определенным образом воздействовать на ответы эксперта. Поэтому предварительно должен быть составлен и испытан план личной беседы, а в процессе опроса исследователю необходимо строго придерживаться определенных заранее формулировок вопросов.

8.2. Порядковые методы измерения качественной информации

Априорная информация, т. е. сведения об объекте, которыми располагают эксперты, чаще всего носит качественный характер. Поэтому для количественного представления такой информации нужно применять специальные методы. Один из способов измерения качественной информации связан с использованием порядковых шкал. Данные, измеренные в порядковой шкале, позволяют установить между объектами отношения „равно", „больше" или „меньше". Одним из методов измерения данных в порядковой шкале является метод ранжирования (упорядочения). Этот метод состоит в расположении объектов в порядке убывания или возрастания какого-либо свойства, присущего этим объектам. Обычно степень, с которой то или иное свойство присуще объектам, не поддается количественному измерению и оценивается только качественно.

Пусть n элементов, обладающих свойством X, расположены экспертами в порядке возрастания или убывания степени обладания этим свойством. Обозначим через x_i место (ранг) i-го элемента (объекта) среди остальных $(n-1)$ элементов (объектов) в этой последовательности. Сумма рангов в таком ряду составляет

$$\sum_{i=1}^{n} x_i = \frac{n(n+1)}{2}, \qquad (8.1)$$

где n — число ранжируемых элементов (объектов). Этот метод формализации априорной информации обычно применяется, когда число ранжируемых объектов невелико ($n \leqslant 15$). Если эксперты затрудняются присвоить всем объектам различные ранги, они могут приписать двум или нескольким объектам одинаковые ранги, так что общее число различных рангов N будет меньше числа исследуемых объектов n. В этом случае полученную ранжировку необходимо привести к нормальному виду, т. е. к такому виду, при котором условие (8.1) выполняется. Для этого объектам, имеющим одинаковые ранги, приписывается ранг, равный среднему значению мест, которые объекты поделили между собой

в ранжировке с совпавшими рангами. Пусть, например, имеется ранжировка (x_i) шести объектов с тремя различными рангами, т. е. $N=3$, $n=6$:

Объекты	i	1	2	3	4	5	6
Ранги	x_i	1	2	3	3	2	3

Объекты 2-й и 5-й поделили между собой места 2 и 3. Поэтому в новой ранжировке приписываем этим объектам одинаковый ранг, равный $(2+3)/2=2{,}5$. Объекты 3-й, 4-й и 6-й поделили между собой места 4, 5 и 6, так что мы приписываем им ранг, равный $(4+5+6)/3=5$. Тогда нормальная ранжировка (x_i^*) имеет вид

Объекты	i	1	2	3	4	5	6
Ранги	x_i^*	1	2,5	5	5	2,5	5

8.3. Анализ ранжированных данных. Коэффициент ранговой корреляции

В результате формализации данных порядковым методом получается упорядоченный ряд, состоящий из рангов. Будем считать ранги случайными числами и введем для них статистику связи. Показателем связи ранжированных рядов может служить коэффициент ранговой корреляции [13,7].

Пусть n объектов ранжированы сначала по степени обладания свойством X, а затем по степени обладания свойством Y. Коэффициент ранговой корреляции оценивает степень связи между этими рядами. Ранжировки представим в виде

$$x_1 x_2 \ldots x_n,$$
$$y_1 y_2 \ldots y_n.$$

При этом предполагается, что выполнено условие (8.1). Ранг i-го объекта в ранжировке по свойству X здесь обозначен через x_i, а в ранжировке по свойству Y — через y_i.

Пусть требуется определить связь между свойствами X и Y для n объектов. Обозначим связь между рангами x_i и x_j через a_{ij}, а связь между рангами y_i и y_j через b_{ij}. Примем, кроме

того, что

$$a_{ij} = -a_{ji}, \qquad a_{ii} = 0,$$
$$b_{ij} = -b_{ji}, \qquad b_{ii} = 0.$$

Тогда коэффициент корреляции определяется как

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}^2}.$$

Если в этой формуле положить $a_{ij} = x_j - x_i$, $b_{ij} = y_j - y_i$ и учесть, что ранги x_i и y_i суть числа натурального ряда, то путем несложных преобразований получим коэффициент ранговой корреляции, предложенный Спирмэном [13]:

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2-1)} = 1 - \frac{6S}{n(n^2-1)}, \tag{8.2}$$

где $S = \sum_{i=1}^{n}(x_i - y_i)^2$.

В случае если ранжировки содержат совпадающие ранги, выражение для ρ имеет вид

$$\rho = \frac{\frac{1}{6}(n^3-n) - \sum_{i=1}^{n}(x_i - y_i)^2 - T - U}{\sqrt{\left\{\frac{1}{6}(n^3-n) - 2T\right\}\left\{\frac{1}{6}(n^3-n) - 2U\right\}}},$$

где

$$T = \frac{1}{12}\sum_{i=1}^{n} t_i(t_i^2-1),$$
$$U = \frac{1}{12}\sum_{i=1}^{n} u_i(u_i^2-1);$$

здесь t_i и u_i — числа повторений i-го ранга в ранжировках по X и Y соответственно.

Проверка значимости коэффициента ранговой корреляции

Исследование распределения вероятностей коэффициента ранговой корреляции показывает, что при отсутствии связи в ранжировках распределение величины ρ в формуле (8.2) стремится

к нормальному распределению с дисперсией $\sigma_\rho^2 = 1/(n-1)$. Поэтому для оценки значимости коэффициента ρ можно использовать нормальный закон распределения. Рассмотрим пример вычисления коэффициента ранговой корреляции и оценки его значимости.

Пример 8.1

На предприятии по производству синтетического каучука требовалось установить, существует ли связь между степенью износа сита и производительностью лентоотливочной машины. Для этого были проранжированы степень износа сита (X) и производительность (Y) для различных ($n=12$) моментов времени (табл. 8.1). Связь между X и Y определялась путем расчета коэффициента ранговой корреляции и оценки его значимости. В табл. 8.1 приведены расчеты для определения значения коэффициента ρ:

$$\hat{\rho} = 1 - \frac{6S}{n(n^2-1)} = 1 - \frac{6 \cdot 30}{12\,(143)} = 0{,}895.$$

Таким образом, между производительностью лентоотливочной машины и степенью износа сита существует положительная связь, оцениваемая коэффициентом ранговой корреляции, равным 0,895.

Таблица 8.1

Результаты опроса (пример 8.1)

Износ сита x_t	1	2	3	4	5	6	7	8	9	10	11	12
Производительность y_t	2	3	1	4	6	5	7	10	11	8	12	9
$\lvert x_t - y_t \rvert$	1	1	2	0	1	1	0	2	2	2	1	3
$(x_t - y_t)^2$	1	1	4	0	1	1	0	4	4	4	1	9
$S = \sum_{t=1}^{12} (x_t - y_t)^2$	30											

Для оценки значимости полученного коэффициента воспользуемся таблицей нормального распределения. Для этого вычислим среднеквадратичное отклонение распределения коэффициента ρ при $n=12$:

$$\sigma_\rho = \sqrt{\frac{1}{n-1}} = \sqrt{\frac{1}{12-1}} \approx 0{,}3.$$

Примем уровень значимости $\alpha = 0{,}05$. Гипотеза о том, что истинное значение коэффициента ранговой корреляции равно нулю, принимается, если

$$Ta_{\text{н}} \geqslant p = \frac{\hat{\rho}}{\sigma_\rho} = \frac{0{,}895}{0{,}3} = 2{,}98.$$

Здесь $p_{кр}$ — значение аргумента функции Лапласа $\Phi(p)$, удовлетворяющее уравнению $\Phi(p_{кр}) = (1/2)(1-\alpha)$. Пользуясь табл. 15.1, находим $p_{кр} = 1{,}96$. Следовательно, гипотеза о равенстве коэффициента ρ нулю отвергается и коэффициент ρ считается значимым.

Конкордация

Степень связи между несколькими ранжировками оценивается коэффициентом конкордации (коэффициентом согласия). Коэффициент конкордации определяет согласованность экспертов при ранжировании n объектов по степени обладания некоторым свойством X.

Пусть имеется n объектов $1, 2, \ldots, i, \ldots, n$, в разной степени обладающих свойством X, и пусть m экспертов ранжируют

Таблица 8.2

Матрица рангов

Эксперты	Объекты					
	1	2	...	i	...	n
1	x_{11}	x_{12}	...	x_{1i}	...	x_{1n}
2	x_{21}	x_{22}	...	x_{2i}	...	x_{2n}
⋮	⋮					
j	x_{j1}	x_{j2}	...	x_{ji}	...	x_{jn}
⋮	⋮					
m	x_{m1}	x_{m2}	...	x_{mi}	...	x_{mn}
$\sum_{j=1}^{m} x_{ji}$						

эти объекты по свойству X (табл. 8.2). В нижней строке табл. 8.2 приведена суммарная ранжировка

$$\sum_{j=1}^{m} x_{j1}, \quad \sum_{j=1}^{m} x_{j2}, \ldots, \sum_{j=1}^{m} x_{jn}.$$

Средний ранг в этом ряду составляет

$$a = \frac{1}{2} m (n+1).$$

Сумма квадратов разностей между членами суммарной ранжировки и членами ряда, составленного из средних значений a,

равна

$$S = \sum_{j=1}^{n} \left\{ \sum_{i=1}^{m} (x_{ij}) - \frac{1}{2} m (n+1) \right\}^2.$$

Величина S достигает максимума, когда все эксперты дают одинаковые ранжировки. Если определить согласованность экспертов как отношение реальной суммы квадратов разностей S к максимально возможной сумме $S_{\text{макс}}$, получается выражение для коэффициента конкордации, предложенное Кендаллом [13]:

$$W = \frac{S}{S_{\text{макс}}} = \frac{12S}{m^2 (n^3 - n)}. \tag{8.3}$$

Величина W изменяется от 0 до 1. $W = 1$ означает, что все эксперты дали одинаковые ранжировки; $W = 0$ означает, что связь между ранжировками, данными m экспертами, отсутствует. Если в ранжировках присутствуют совпадающие ранги, формула для W принимает вид

$$W = \frac{S}{\frac{1}{12} m^2 (n^3 - n) - m \sum_{j=1}^{m} T_j}, \tag{8.4}$$

где

$$T_j = \frac{1}{12} \sum_{l=1}^{n} (t_l^3 - t_l),$$

t_l — число повторений l-го ранга в j-м ряду.

Для оценки значимости коэффициента конкордации используется χ^2-распределение с числом степеней свободы $\varphi = n - 1$, которому подчинена величина $Wm(n-1)$ [1]. При $n < 10$ распределение величины $m(n-1)W$ отличается от χ^2-распределения, поэтому для оценки значимости приходится пользоваться специальными таблицами [13].

---

[1] Проверка значимости сводится к проверке статистической гипотезы о равенстве коэффициента конкордации нулю. При заданном уровне значимости α эта гипотеза отвергается, когда

$$m(n-1)\,\hat{W} \geqslant \chi^2_{\text{кр}},$$

где $\chi^2_{\text{кр}}$ — критическое значение χ^2-распределения при числе степеней свободы $\varphi = n - 1$. В этом случае конкордация считается значимой. При $\varphi = n - 1 > 30$ для проверки значимости используют нормальное распределение (см. примечание 2 на стр. 520). — *Прим. ред.*

8.4. Метод парных сравнений

На практике встречаются случаи, когда исследователи затрудняются расположить объекты в порядке возрастания или убывания признака X (например, когда число объектов превышает 20—30). В таких случаях, даже если и удается получить ранжировки, нельзя быть уверенным в их достоверности, поскольку сами эксперты не уверены в своих оценках. Тогда суммарную ранжировку можно получить другими путями, например методом парных сравнений. При использовании этого метода каждый эксперт сравнивает друг с другом все n объектов по данному свойству X. В результате проведения C_n^2 таких сравнений эксперт отдает предпочтение каким-то C_n^2 объектам по отношению к остальным C_n^2 объектам.

Пусть эксперт произвел парные сравнения n объектов по свойству X. В результате сравнений заполняется матрица парных сравнений, в которой единица в клетке с номерами i и j означает, что эксперт предпочитает объект i объекту j в отношении свойства X; соответственно в j-й строке и i-м столбце при этом записывается нуль. Такую таблицу составляет каждый из m экспертов. Данные этих m таблиц можно свести в одну общую таблицу, которая и будет обобщенной матрицей парных сравнений m экспертов. В этой матрице число γ_{ij}, записанное в i-й строке и j-м столбце, показывает, как часто эксперты предпочитают i-й объект j-му в отношении свойства X.

Расположение объектов в порядке возрастания или убывания сумм чисел γ_{ij} по столбцам или строкам суммарной матрицы соответствует ранжировке объектов по степени обладания свойством X. При полном согласии экспертов C_n^2 ячеек матрицы будут содержать число $\gamma_{ij} = m$, а в остальных ячейках будут нули. Коэффициент согласия экспертов при парных сравнениях подсчитывается по формуле

$$V = \frac{4\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} C_{\gamma_{ij}}^2\right\}}{m(m-1)n(n-1)}, \qquad (8.5)$$

где γ_{ij} — число, стоящее в матрице парных сравнений в ячейке с номерами i и j, а $C_{\gamma_{ij}}^2$ — число сочетаний из γ_{ij} по 2[1].

Практическое вычисление величины $\sum_{i}^{n}\sum_{j}^{n} C_{\gamma_{ij}}^2$ довольно трудоемко; путем несложных алгебраических преобразований его

[1] Знаменатель в выражении (8.5) представляет собой нормирующий множитель, равный максимально возможной величине числителя. Числитель имеет максимальную величину в случае, когда все исследователи дают одинаковые матрицы сравнений. При этом $V = 1$. — *Прим. ред.*

можно привести к следующему выражению, более удобному для вычислений:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} C^2_{\gamma_{ij}} = Q = \sum_{(i,j)}^{n} \gamma_{ij}^2 - m\sum_{(i,j)}^{n} \gamma_{ij} + C_m^2 C_n^2, \qquad (8.6)$$

где суммирование производится только по ячейкам матрицы, лежащим выше (или ниже) главной диагонали [1].

Для оценки значимости коэффициента V используется χ^2-критерий. Величина

$$\chi^2 = \frac{4}{m-2}\left\{Q - \frac{1}{2} C_n^2 C_m^2 \frac{m-3}{n-2}\right\} \qquad (8.7)$$

имеет при больших n и m χ^2-распределение с

$$\varphi = C_n^2 \frac{m(m-1)}{(m-2)^2}$$

степенями свободы [2]. При небольших значениях m и n ($m < 6$, $n < 8$) для оценки значимости коэффициента согласия используются специально рассчитанные таблицы распределения вероятностей величины Q при различных сочетаниях m и n [13].

Пример 8.2

В табл. 8.3 представлена суммарная матрица парных сравнений $n = 5$ объектов, полученная путем сложения матриц парных сравнений $m = 10$ экспертов. Здесь же приведены суммы по строкам и столбцам суммарной матрицы и ранжировка (последняя строка таблицы), полученная путем расположения объектов в порядке убывания сумм по столбцам. Рассчитаем коэффициент согласия экспертов и оценим его значимость.

Для ускорения вычислений целесообразно столбцы или строки матрицы поменять местами так, чтобы они были расположены в порядке убывания или возрастания сумм по столбцам или строкам. Расположив элементы в порядке убывания сумм по строкам, мы получаем табл. 8.4.

Величину Q вычисляем по формуле (8.6). Так как числа, стоящие в ячейках, расположенных ниже главной диагонали, меньше чисел, расположенных выше диагонали, то проще про-

---

[1] При выводе выражения (8.6) использовано условие $\gamma_{ij} = m - \gamma_{ji}$. При этом $C^2_{\gamma_{ij}} + C^2_{\gamma_{ji}} = C^2_{\gamma_{ij}} + C^2_{m-\gamma_{ij}} = \gamma_{ij}^2 - m\gamma_{ij} + C_m^2$. — *Прим. ред.*

[2] Коэффициент согласия считается значимым, если рассчитанная по (8.7) величина χ^2 превосходит критическое значение $\chi^2_{кр}$, определяемое при выбранном уровне значимости α и числе степеней свободы φ по табл. 15.2. — *Прим. ред.*

Таблица 8.3

Результаты парных сравнений (пример 8.2)

Объекты	Объекты					Суммы
	1	2	3	4	5	
1	—	10	7	6	1	24
2	0	—	3	5	2	10
3	3	7	—	8	3	21
4	4	5	2	—	2	13
5	9	9	7	8	—	32
Суммы	16	30	19	27	8	100
Ранги	4	1	3	2	5	

Таблица 8.4

Таблица с упорядоченными суммами (пример 8.2)

l / i	5	1	3	4	2	Сумма
5	—	9	7	8	8	32
1	1	—	7	6	10	24
3	1	3	—	8	7	21
4	2	4	2	—	5	13
2	2	0	3	5	—	10

изводить вычисления для нижней половины матрицы[1)]

$$\hat{Q} = 81 - 10 \times 25 + 45 \times 10 = 281.$$

Коэффициент согласия $\hat{V}$ вычисляем по формуле (8.5):

$$\hat{V} = \frac{4 \times 281}{90 \times 20} = 0{,}624.$$

Для оценки значимости коэффициента согласия при $m = 10$, $n = 5$,

[1)] В этом случае формулу (8.6) можно записать следующим образом:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{i-1} \gamma_{ij}^{a} - m \sum_{i=1}^{n} \sum_{j=1}^{i-1} \gamma_{ij} + C_m^a C_n^a.$$

— *Прим. ред.*

$\hat{Q} = 281$ найдем величину $\hat{\chi}^2$ по формуле (8.7):

$$\hat{\chi}^2 = \frac{4}{10-2}\left\{281 - \frac{1}{2}\times 10\times 45\times\frac{10-3}{10-2}\right\} = 42,$$

$$\varphi = C_n^a \frac{m(m-1)}{(m-2)^2} = 10\times\frac{10\times 9}{8^2} = 14{,}05.$$

Для уровня значимости $\alpha = 0{,}01$ из таблиц χ^2-распределения (табл. 15.2) находим при $\varphi = 14$ $\chi^2_{кр} = 29{,}14$. Так как $\hat{\chi}^2 = 42 > 29{,}14 = \chi^2_{кр}$, то можно считать, что коэффициент согласия $\hat{V} = 0{,}624$ значимо отличается от нуля и гипотеза о наличии неслучайного согласия между экспертами относительно полученной суммарной ранжировки принимается.

8.5. Применение экспертных оценок при выявлении факторов, существенно влияющих на ход технологического процесса

Для процесса производства синтетического каучука путем статистической обработки экспериментальных данных были выделены наиболее существенные факторы, влияющие на ход процесса. После некоторого усовершенствования технологии встал вопрос о справедливости выбора тех же факторов в условиях новой технологии. Был проведен опрос специалистов, работающих с данным оборудованием или хотя бы просто знакомых с процессом. Анкета, предъявленная специалистам (экспертам), приведена в табл. 8.5. Эксперты должны были проранжировать $n = 11$ факторов, перечисленных в анкете, по степени их влияния на скорость высыхания крошки каучука.

Таблица 8.5

Анкета

Номер параметра	Название фактора	Ранг
1	Относительная влажность крошки на входе в сушилку	
2	pH в 6-м аппарате	
3	pH в 7-м аппарате	
4	Расход NaCl	
5	Расход серума в 6-м аппарате	
6	Расход серума в 7-м аппарате	
7	Температура серума	
8	Дефо латекса	
9	Процентное содержание масла в каучуке	
10	Расход латекс-масла на машину	
11	Поверхностное натяжение латекса	

В табл. 8.6 приведена матрица рангов, полученная из анкет в результате опроса $m = 18$ экспертов. Матрица была преобразована таким образом, чтобы выполнялось условие (8.1). Преобразованная матрица приведена в табл. 8.7.

Таблица 8.6

Матрица рангов

Номер эксперта	*Номер параметра*										
	1	2	3	4	5	6	7	8	9	10	11
1	1	2	2	2	3	3	4	1	1	2	4
2	1	8	3	6	11	10	7	2	9	5	4
3	1	4	7	8	6	10	11	3	5	1	9
4	3	8	1	4	8	8	5	8	6	2	7
5	7	9	1	5	6	6	2	8	3	1	4
6	1	1	3	2	2	3	4	1	2	1	4
7	1	1	2	2	2	3	4	1	2	2	4
8	1	2	3	3	3	3	4	2	3	4	4
9	1	3	4	3	4	4	4	2	4	1	2
10	1	4	4	2	4	4	4	3	3	5	6
11	2	5	5	6	7	7	7	3	4	1	8
12	2	1	4	3	2	6	1	1	3	1	3
13	3	2	5	4	5	6	2	1	3	3	4
14	3	2	4	2	6	7	5	1	4	1	8
15	2	1	9	5	7	8	10	3	4	6	2
16	1	5	3	5	6	6	6	2	4	1	5
17	1	4	10	9	7	8	6	2	5	3	11
18	4	4	2	3	5	5	5	1	3	5	6

В последнем столбце табл. 8.7 записаны числа совпавших рангов в каждой строке. В нижней строке матрицы приведена сумма рангов факторов $\sum_{j=1}^{m} x_{jl}$. По этим данным рассчитан коэффициент конкордации

$$\hat{W} = \frac{14066{,}5}{\frac{1}{12}\times 18^2\times 11\times 120 - \frac{1}{12}\times 18\times 1194} = 0{,}415.$$

Оценка значимости этого коэффициента проводилась по критерию χ^2:

$$\hat{\chi}^2 = m(n-1)\hat{W} = 18\times 10\times 0{,}415 = 74{,}5.$$

Таблица 8.7

Преобразованная матрица рангов

Номер эксперта	*Номер параметра*											
	1	2	3	4	5	6	7	8	9	10	11	t_t
1	2	5,5	5,5	5,5	8,5	8,5	10,5	2	2	5,5	10,5	3 + 4 + 2 + 2
2	1	8	3	6	11	10	7	2	9	5	4	0
3	1	4	7	8	6	10	11	3	5	2	9	0
4	3	9	1	4	9	9	5	9	6	2	9	5
5	9	11	1,5	6	7,5	7,5	3	10	4	1,5	5	2 + 2
6	2,5	2,5	8,5	6	6	8,5	10,5	2,5	6	2,5	10,5	4 + 2 + 3 + 2
7	2	2	6	6	6	9	10,5	2	6	6	10,5	3 + 5 + 2
8	1	2,5	6	6	6	6	10	2,5	6	10	10	2 + 5 + 3
9	1,5	5,5	9	5,5	9	9	9	3,5	9	1,5	3,5	2 + 2 + 2 + 5
10	1	7	7	2	7	7	7	3,5	3,5	10	11	2 + 5
11	2	5,5	5,5	7	9	9	9	3	4	1	11	2 + 3
12	5,5	2,5	10	8	5,5	11	2,5	2,5	8	2,5	8	4 + 2 + 3
13	5	2,5	9,5	7,5	9,5	11	2,5	1	5	5	7,5	3 + 2 + 2 + 2
14	5	3,5	6,5	3,5	9	10	8	1,5	6,5	1,5	11	2 + 2 + 2
15	2	1	9	5	7	8	10	3	4	6	11	0
16	1,5	7	4	7	10	10	10	3	5	1,5	7	2 + 3 + 3
17	1	4	10	9	7	8	6	2	5	3	11	0
18	5,5	5,5	2,0	3,5	8,5	8,5	8,5	1	3,5	8,5	11	2 + 2 + 4
$\sum_{j=1}^{m} x_{ji}$	51,5	88,5	111	105,5	141,5	160	140,0	57	97,5	75	160,5	

Для уровня значимости $\alpha = 0{,}01$ и для $\varphi = n - 1 = 10$ степеней свободы из табл. 15.2 находим

$$\chi^2_{кр} = 23{,}2.$$

Так как $\chi^2_{кр} = 23{,}2 < \hat{\chi}^2 = 74{,}5$, то гипотезу о том, что между ранжировками специалистов существует согласованность, принимаем. Степень согласованности специалистов оценивается коэффициентом $\hat{W} = 0{,}415$. Отобранные ранее статистическими методами шесть факторов оказались в новой ранжировке на втором — седьмом местах соответственно. На первом месте оказался новый фактор — влажность крошки каучука, который в прежних расчетах не фигурировал. Таким образом, с помощью опроса удалось выделить наиболее существенные факторы, влияющие на процесс при новой технологии.

8.6. Сочетание опроса с экспериментом

Получение информации путем опроса может оказаться полезным не только тогда, когда невозможно прямое физическое измерение определенных данных, но и в тех случаях, когда измерение каких-либо количественных свойств объекта сопряжено с большими затратами времени или средств. В таких ситуациях целесообразно с целью сокращения числа опытов сочетать использование априорных данных с данными физического эксперимента. В работе [12] рассмотрен пример сочетания опроса с экспериментом при построении математической модели объекта.

Пример 8.3

Экспериментальным путем требуется оценить зависимость целевой величины y некоторого технологического процесса от входных величин x_1, x_2 и x_3. Предполагаем, что модель имеет вид

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_{23}x_2x_3 + \ldots + a_{33}x_3^2. \qquad (8.8)$$

Для определения оценок коэффициентов a предполагается реализовать план, представленный в табл. 8.8. Чтобы уменьшить число физически реализуемых экспериментов, можно часть из них заменить опросом экспертов.

Первый шаг состоит в предъявлении экспертам всех точек центрального композиционного плана, приведенного в табл. 8.8. Эксперты дают оценки y_1 выхода для каждой точки плана. По данным шага I в табл. 8.8 методом наименьших квадратов вычисляются коэффициенты уравнения (8.8):

$$\hat{y}_1 = 10{,}889 + 1{,}625x_1 - 0{,}125x_2 - 0{,}1375x_3 + 0{,}139x_1^2 - \\ - 0{,}361x_2^2 + 0{,}139x_3^2 - 0{,}250x_1x_2 - 0{,}250x_1x_3 - 0{,}250x_2x_3.$$

Таблица 8.8

План эксперимента (пример 8.3) [1]

Номер опыта	x_1	x_2	x_3	*Шаг I* $\bar{y}_1$	*Шаг II* $\bar{y}_2$	*Шаг III* $\bar{y}_3$
1	−1	−1	−1	10	*8	*8
2	−1	−1	+1	8	6	*4
3	−1	+1	−1	10	8	*12
4	−1	+1	+1	6	*8	*8
5	+1	−1	−1	14	12	*8
6	+1	−1	+1	10	*8	*8
7	+1	+1	−1	14	*12	*12
8	+1	+1	+1	10	8	*12
9	−2	0	0	9	7	8
10	+2	0	0	15	13	12
11	0	−2	0	10	8	7
12	0	+2	0	10	8	9
13	0	0	−2	14	12	12
14	0	0	+2	10	8	6
15	0	0	0	12	*10	*10

[1] Звездочкой отмечены точки, в которых проводится активный эксперимент.

Второй шаг состоит в проведении эксперимента в точках 1, 4, 6 и 7 плана (шаг II в табл. 8.8). Эти точки составляют полуреплику трехфакторного эксперимента на двух уровнях с центральной точкой. Результаты физического эксперимента, помеченные в табл. 8.8 звездочкой, предъявляются экспертам, которые несколько исправляют свои первые оценки. Данные шага II используются для вычисления новых оценок коэффициентов. Значения оценок коэффициентов, рассчитанные для каждого шага, приведены в табл. 8.9.

На третьем шаге проводится физический эксперимент в точках 2, 3, 5, 8—вторая полуреплика трехфакторного эксперимента на двух уровнях. Результаты этого эксперимента и скорректированные в соответствии с ними оценки выхода, сделанные экспертами, приведены в табл. 8.8, а полученные оценки коэффициентов—в табл. 8.9, из которой видно, как изменялись оценки $\hat{a}$ при проведении последовательных экспериментов в сочетании с опросом.

Из табл. 8.9 следует, что коэффициенты, полученные уже на основе первого опроса, не сильно отличаются от коэффициентов, полученных при реализации физического эксперимента во всех 15 точках плана, приведенных в столбце IV табл. 8.9. Сочетание

Таблица 8.9

Оценки коэффициентов модели (пример 8.3)

Коэффициент	Шаг			
	I	II	III	IV
$\hat{a}_0$	10,889	9,333	9,875	10,1
$\hat{a}_1$	1,625	1,375	1,000	1,0
$\hat{a}_2$	−0,125	0,125	1,250	2,0
$\hat{a}_3$	−0,137	−1,125	−1,250	−1,0
$\hat{a}_{11}$	0,139	0,083	0,000	0,0
$\hat{a}_{22}$	−0,361	−0,417	−0,500	−1,0
$\hat{a}_{33}$	0,139	0,083	−0,250	0,0
$\hat{a}_{12}$	−0,250	0,250	0,000	0,0
$\hat{a}_{13}$	−0,250	0,750	1,000	1,0
$\hat{a}_{23}$	−0,250	0,250	0,000	0,0

опроса с физическим экспериментом (шаги II и III в табл. 8.9) дают еще лучшее приближение к результатам физического эксперимента. Приведенный пример показывает, что комбинация опроса с физическим экспериментом позволяет уменьшить число опытов при построении модели.

ЛИТЕРАТУРА

1. Thurstone L. L., Contributions to mathematical psychology, New York, 1964.
2. Thurstone L. L., Law of comparative judgement, *Psychol. Rev.*, 34, 273—286 (1927).
3. Guilford J. P., Psychometric methods, New York, 1954.
4. Bert F., Green, Attitude Measurement, см. Handbook of Social Psychology, ed. by G. Lindsey, vol. I, Addison-Wesley Publ. Co., 1959.
5. Шубкин В. Д., Количественные методы в социологии, изд-во «Наука», М., 1966.
6. Мот Ж., Статистические методы. Предвидения и решения на предприятии, М., 1966.
7. Дьякова Н. С., Круг Г. К., Руководящие технические материалы. Экспериментально-статистические методы получения математического описания и оптимизации сложных технологических процессов (ранговая корреляция), вып. 3, ОКБА, М., 1966.
8. Круг Г. К., Дьякова Н. С., Сумароков А. В., Ранкор для оценки качества, *Вестн. высшей школы*, № 5 (1969).
9. Кириченко Г. С., Макалец В. Н., Сравнение методов априорного ранжирования, *Заводск. лаб.*, № 1 (1971).
10. Розанов Г. В., Френкель А. А., Об одной многоэтапной процедуре формализации априорной информации, *Заводск. лаб.*, № 3 (1970)
11. Bischoff Th., Die Befragung als Hilfmittel zur Datengewinnung, *Agrarwirtschaft*, № 6 (1964).
12. Hinchen J. D., From Opinion to Fact throught Sequential Experimentation, *Industr. Quality Control*, 23, № 3 (1966).
13. Kendall M. G., Rank correlation Methods, New York, 1955; русский перевод: Кендалл М. Г., Ранговая корреляция, М., 1975.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Бешелев С. Д., Гурвич Ф. Г., Математико-статистические методы экспертных оценок, изд-во „Статистика", М., 1974.
2. Глушков В. М., О прогнозировании на основе экспертных оценок, МДНТП, М., 1969.
3. Янч Э., Прогнозирование научно-технического прогресса, из-во „Прогресс", М., 1970.
4. Саркисян С. А., Голованов Л. Е., Прогнозирование развития больших систем, изд-во „Статистика", М., 1975.

9.

МЕТОДЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ОБЪЕКТОВ С ИЗМЕНЯЮЩИМИСЯ ВО ВРЕМЕНИ ХАРАКТЕРИСТИКАМИ (ПРИ НАЛИЧИИ НЕКОНТРОЛИРУЕМОГО ДРЕЙФА)

9.1. Введение

Химико-технологические процессы часто подвержены действию неконтролируемых возмущений. К числу таких возмущений относятся, например, изменения активности катализатора во времени, процессы старения и износа оборудования, случайные изменения трудноизмеримых характеристик входного продукта и т. п. Подобного рода возмущения мы будем называть неконтролируемыми, если они не могут быть количественно оценены (измерены). В отличие от неконтролируемых измеряемые возмущения мы будем называть контролируемыми (измеряемый расход сырья, измеряемые характеристики выходного продукта и пр.). При построении по экспериментальным данным математического описания объекта мы можем вводить в модель лишь контролируемые (измеряемые) возмущения (контролируемые переменные объекта). Часто, однако, не удается построить работоспособное математическое описание процесса, если не принимать во внимание факта воздействия неконтролируемых возмущений (неконтролируемых переменных) на характер протекания процессов в объекте.

Влияние неконтролируемых возмущений на математическую модель процесса формально можно свести к следующей схеме. Предположим, что вид уравнения, описывающего исследуемый процесс, задан с точностью до значения вектора коэффициента a, истинные величины которых зависят от значений неконтролируемых переменных объекта. Тогда, вследствие того что неконтролируемые воздействия изменяются во времени, истинные значения коэффициентов a будут также со временем изменяться. Таким образом, при наличии неконтролируемых возмущений коэффициенты математической модели процесса некоторым образом изменяются во времени.

Часто процесс изменения неконтролируемых переменных значительно медленнее процесса изменения контролируемых факторов. Так, если изменение контролируемых свойств входных компонентов и управляющих воздействий производится при управлении некоторым объектом, например, в среднем один раз в час, то неконтролируемая активность катализатора может практически не меняться в течение суток, старение агрегата

будет сказываться на выходном продукте лишь после нескольких месяцев эксплуатации и т. д. Неконтролируемые воздействия такого типа вызывают медленное изменение истинных значений коэффициентов модели. Этот процесс часто называют дрейфом характеристик, а объекты, обладающие таким свойством,— дрейфующими.

В этом разделе мы будем рассматривать методы построения математических моделей в условиях дрейфа характеристик объекта, причем наше рассмотрение ограничено здесь лишь случаями непрерывного временно́го дрейфа. Кроме того, мы будем рассматривать лишь такие задачи, где построение математической модели сводится к нахождению оценок некоторого числа параметров зависимости между входными и выходной переменными объекта, причем выходная переменная является линейной функцией искомых коэффициентов.

Целесообразно различать следующие два вида дрейфа:

1) аддитивный дрейф,

2) неаддитивный дрейф.

В первом случае неконтролируемые возмущения влияют только на величину свободного члена в уравнении, описывающем исследуемый процесс. Этому виду дрейфа соответствует такая модель объекта:

$$y = a_0(t) + \sum_{l=1}^{k} a_l f_l(x), \tag{9.1}$$

где x — вектор входных (независимых) переменных объекта, $f_l(x)$ — заданная система функций; a_l, $l = 0, 1, \ldots, k$, — неизвестные коэффициенты; y — выходная (зависимая) переменная; t — время. Таким образом, аддитивный дрейф не изменяет степени влияния независимых переменных x на переменную y. Его результатом является лишь изменение не зависящей от x составляющей выходного процесса.

В случае неаддитивного дрейфа неконтролируемые возмущения приводят к изменению во времени степени влияния на выходную переменную каких-либо из независимых переменных или функций от независимых переменных. Это означает, что при неаддитивном дрейфе некоторые или все коэффициенты a_l, $l = 1, 2, \ldots, k$, модели объекта изменяются во времени. Неаддитивный дрейф имеет место, например, у объекта, модель которого имеет вид

$$y = a_0(t) + \sum_{l=1}^{k} a_l(t) f_l(x). \tag{9.2}$$

Методы построения математических моделей дрейфующих объектов часто оказываются различными для объектов с аддитивным и неаддитивным дрейфом. Пусть ищется зависимость некоторого

выходного параметра y от параметров x с целью получения в дальнейшем с помощью этой зависимости значения x^*, обеспечивающего, например, максимум y. Тогда величина свободного члена в уравнении, описывающем эту зависимость, может не интересовать исследователя, так как она никак не повлияет на выбор x^*. Изменение свободного члена во времени (т. е. аддитивный дрейф) выступает при этом в качестве „мешающего" фактора, и естественно, что исследователь в этом случае должен стремиться как можно лучше „очистить" эксперимент от влияния дрейфа, или, другими словами, элиминировать дрейф. Если ему это удается, дальнейший путь решения не отличается от методов, используемых при отсутствии дрейфа.

В отличие от случая аддитивного дрейфа при неаддитивном дрейфе экспериментатора во всякий момент времени интересуют истинные значения дрейфующих коэффициентов. Методы решения подобных задач можно получить с помощью теории оптимальной фильтрации.

Для выбора метода построения модели дрейфующего объекта необходимо сделать предположение о типе процесса, описывающего дрейф. Мы будем различать следующие два типа процессов изменения коэффициентов модели объекта:

1) детерминированный (или регулярный) процесс,

2) случайный процесс.

В первом случае изменение во времени значения произвольного i-го коэффициента модели может быть представлено с помощью зависимости

$$a_i(t) = d_{0i}\varphi_{0i}(t) + d_{1i}\varphi_{1i}(t) + d_{2i}\varphi_{2i}(t) + \ldots + d_{mi}\varphi_{mi}(t), \quad (9.3)$$

где $\varphi_{ji}(t)$, $j = 0, \ldots, m$, —заданная система непрерывных функций. Процесс изменения коэффициента a_i, описываемый функцией (9.3), мы будем называть детерминированным дрейфом. Метод решения задачи построения математической модели объекта в условиях детерминированного аддитивного дрейфа описан в разд. 9.3. В разд. 9.4, помимо этого, излагается метод планирования эксперимента в случаях аддитивного дрейфа, описываемого полиномом известной степени.

При случайном характере изменения во времени коэффициентов модели объекта мы будем говорить о случайном дрейфе характеристик объекта. Построение в этом случае аналитическим путем оптимального в каком-либо смысле алгоритма решения задачи требует достаточно полного знания статистических свойств процесса изменения коэффициентов модели объекта [1—3]. В практических задачах эта информация, как правило, отсутствует, и поэтому аналитический путь синтеза наилучшего алгоритма часто оказывается невозможным. В таких ситуациях может быть использован адаптивный подход [4]. Метод реше-

ния задачи, основанный на этом подходе, рассмотрен в разд. 9.5 применительно к случаю неаддитивного временно́го дрейфа.

Когда исследователь приступает к решению задачи построения математической модели некоторого процесса, ему часто приходится решать вопрос о том, следует ли к исследуемому объекту подходить как к дрейфующему. Для этой цели может оказаться полезным метод проверки условий отсутствия дрейфа, рассмотренный в разд. 9.2.

9.2. Метод проверки условий отсутствия дрейфа характеристик объекта

Идея рассматриваемого здесь метода чрезвычайно проста. Она вытекает из следующих очевидных рассуждений. Пусть на исследуемом объекте проведено на разнесенных интервалах времени m серий экспериментов и по данным каждой серии построена математическая модель заданного вида. Предположим также, что математическое описание найдено по результатам всех экспериментов. Тогда в случае отсутствия дрейфа полученные модели должны быть достаточно близкими. Это означает, например, что не должно наблюдаться значимого различия в точности оценки выходной переменной для любой серии опытов по уравнению, полученному по данным этой серии, и по уравнению, полученному по результатам всех опытов.

Для реализации этой идеи необходимо найти критерий, который позволил бы количественно оценить близость различных моделей и в то же время дал бы возможность свести задачу проверки условий отсутствия дрейфа к задаче статистической проверки некоторой гипотезы.

Пусть модель исследуемого объекта имеет вид

$$y = a'f(x), \tag{9.4}$$

где a — вектор-столбец коэффициентов модели объекта размерности $k+1$; x — вектор независимых переменных размерности n; $f(x)$ — вектор функций от независимых переменных размерности $k+1$.

Предположим, что коэффициенты модели остаются неизменными в течение каждой серии опытов. Это значит, что, если дрейф имеет место, он настолько медленен, что на интервалах времени, которые выбраны для проведения каждой из m серий опытов, коэффициенты модели практически не изменяются. Наблюдаемые при проведении экспериментов значения выходной переменной

$$\tilde{y} = \bar{y} + \tilde{e}, \tag{9.5}$$

где e — случайная ошибка наблюдений; $\bar{y}$ — действительное значение выходной переменной. Мы будем считать, что ошибка e

независима и имеет нормальный закон распределения с нулевым математическим ожиданием и дисперсией, неизменной для всех серий опытов. Приведенные здесь условия являются обычными предпосылками метода регрессионного анализа. Требование нормальности закона распределения ошибки будет использовано ниже при проверке гипотезы отсутствия дрейфа.

Обозначим через X_l матрицу размера $N_l \times n$ (N_l — число опытов в l-й серии), определяющую план экспериментов в l-й серии:

$$X_l = \begin{bmatrix} x_{1l}^1 & x_{2l}^1 & \dots & x_{nl}^1 \\ x_{1l}^2 & x_{2l}^2 & \dots & x_{nl}^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{1l}^{N_l} & x_{2l}^{N_l} & & x_{nl}^{N_l} \end{bmatrix}. \tag{9.6}$$

Через $\tilde{Y}_l$ обозначим матрицу размера $N_l \times 1$, представляющую результаты измерения зависимой переменной в l-й серии:

$$\tilde{Y}_l = \begin{bmatrix} \tilde{y}_l^1 \\ \tilde{y}_l^2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{y}_l^{N_l} \end{bmatrix}. \tag{9.7}$$

Наконец, через F_l обозначим матрицу значений функций независимых переменных в l-й серии опытов:

$$F_l = \begin{bmatrix} f_{0l}^1 & f_{1l}^1 & \dots & f_{kl}^1 \\ f_{0l}^2 & f_{1l}^2 & \dots & f_{kl}^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ f_{0l}^{N_l} & f_{1l}^{N_l} & \dots & f_{kl}^{N_l} \end{bmatrix} = \begin{bmatrix} (f_l^1)' \\ (f_l^2)' \\ \cdot \\ \cdot \\ \cdot \\ (f_l^{N_l})' \end{bmatrix}. \tag{9.8}$$

Применяя аппарат метода множественной регрессии, найдем оценки коэффициентов локальных моделей для всякой l-й серии опытов. Для этого необходимо решить относительно $\hat{a}^{(l)}$ систему нормальных уравнений вида

$$F_l' F_l \hat{a}^{(l)} = F_l' \tilde{Y}_l. \tag{9.9}$$

Из (9.9) следует

$$\hat{a}^{(l)} = (F_l' F_l)^{-1} F_l' \tilde{Y}_l.$$

Составляя систему уравнений (9.9) для $i = 1, 2, \ldots, m$, находим соответственно оценки $\hat{a}^{(1)}, \hat{a}^{(2)}, \ldots, \hat{a}^{(m)}$.

Найдем теперь оценки коэффициентов по результатам всех опытов. Общее число экспериментов в m сериях обозначим через N:

$$N = N_1 + N_2 + \ldots + N_m. \tag{9.10}$$

Объединяя матрицы X_i, $\tilde{Y}_i$ F_i, $i = 1, 2, \ldots, m$, получим

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_m \end{bmatrix}, \tag{9.11}$$

$$\tilde{Y} = \begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{Y}_m \end{bmatrix}, \tag{9.12}$$

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \cdot \\ \cdot \\ \cdot \\ F_m \end{bmatrix}. \tag{9.13}$$

Оценки наименьших квадратов для коэффициентов модели (9.4), рассчитанные по результатам N опытов, находятся путем решения относительно $\hat{a}$ системы нормальных уравнений

$$F'F\hat{a} = F'\tilde{Y}. \tag{9.14}$$

Из (9.14) следует

$$\hat{a} = (F'F)^{-1} F'\tilde{Y}.$$

Запишем выражение для остаточной суммы квадратов S_R для общей модели, построенной по всем опытам:

$$S_R = \sum_{j=1}^{N} (\tilde{y}^j - \hat{a}' f^j)^2 = \tilde{Y}'\tilde{Y} - \hat{a}' F'\tilde{Y}. \tag{9.15}$$

Выражение для остаточной суммы квадратов можно предста-

вить в следующем виде:

$$S_R = \tilde{Y}'\tilde{Y} - \hat{a}'F'\tilde{Y} = \left(\tilde{Y}'\tilde{Y} - \sum_{i=1}^{m} \hat{a}^{(i)\prime} F_i' \tilde{Y}_i\right) + \left(\sum_{i=1}^{m} \hat{a}^{(i)\prime} F_i' \tilde{Y}_i - \hat{a}'F'\tilde{Y}\right) = Q_2 + Q_1. \quad (9.16)$$

Рассмотрим поочередно величины Q_2 и Q_1

$$Q_2 = \tilde{Y}'\tilde{Y} - \sum_{i=1}^{m} \hat{a}^{(i)\prime} F_i' \tilde{Y}_i = \sum_{i=1}^{m} (\tilde{Y}_i' \tilde{Y}_i - \hat{a}^{(i)\prime} F_i' \tilde{Y}_i) = \sum_{i=1}^{m} \sum_{j=1}^{N_i} (\tilde{y}_j^i - \hat{a}^{(i)\prime} f_j^i)^2 = \sum_{i=1}^{m} S_i. \quad (9.17)$$

Здесь через S_i обозначена остаточная сумма квадратов для модели, построенной по i-й серии экспериментов. Величина Q_2, таким образом, представляет собой сумму остаточных сумм квадратов по всем сериям опытов и характеризует рассеяние экспериментальных данных относительно локальных математических моделей, построенных для каждой серии опытов. Причина, вызвавшая это рассеяние, — наличие ошибки ε при наблюдении значений зависимой переменной.

С помощью величины Q_2 можно найти оценку s_ε^2 дисперсии ошибки единичного эксперимента. Для этого необходимо Q_2 разделить на число степеней свободы. Число степеней свободы легко получить из следующих соображений. Оценка дисперсии ошибки единичного опыта, полученная по некоторой i-й серии опытов, имеет число степеней свободы $N_i - (k+1)$. Оценка, вычисленная путем усреднения результатов по всем сериям опытов, будет иметь число степеней свободы

$$\varphi_2 = \sum_{i=1}^{m} (N_i - (k+1)) = N - m(k+1). \quad (9.18)$$

Таким образом,

$$s_{\varepsilon\text{д}}^2 = \frac{Q_2}{N - m(k+1)}. \quad (9.19)$$

Рассмотрим теперь величину

$$Q_1 = \sum_{i=1}^{m} \hat{a}^{(i)\prime} F_i' \tilde{Y}_i - \hat{a}'F'\tilde{Y}. \quad (9.20)$$

Нетрудно показать, что

$$Q_1 = \sum_{i=1}^{m} \sum_{j=1}^{N_i} (\hat{a}^{(i)\prime} f_j^i - \hat{a}' f_0^i)^2. \quad (9.21)$$

Это означает, что величина Q_1 представляет собой сумму квадратов разностей между оценками выходной переменной, полученными по локальным моделям и по общей модели. Величина Q_1 будет тем больше, чем сильнее различие в коэффициентах локальных моделей, т. е. чем большее влияние оказывает дрейф на величины коэффициентов. Если дрейф отсутствует, единственной причиной появления рассеяния Q_1 является наличие ошибок e наблюдения зависимой переменной. Это значит, что с помощью величины Q_1 можно в случае отсутствия дрейфа получить оценку s^2_{e1} для дисперсии ошибки единичного опыта. Для этого необходимо Q_1 разделить на число степеней свободы φ_1, где

$$\varphi_1 = (m-1)(k+1). \tag{9.22}$$

Таким образом,

$$s^2_{e1} = \frac{Q_1}{(m-1)(k+1)}. \tag{9.23}$$

Величины s^2_{e1} и s^2_{e2} в случае отсутствия дрейфа являются оценками одной и той же дисперсии ошибки единичного эксперимента. Если имеется дрейф, то дисперсия, связанная с величиной Q_1, будет превосходить дисперсию ошибки эксперимента. Это значит, что при наличии дрейфа оценка s^2_{e1} должна быть значительно больше оценки s^2_{e2}, а в случае отсутствия дрейфа эти оценки должны быть близкими.

Таким образом, задачу проверки условий отсутствия дрейфа можно сформулировать как задачу проверки статистической гипотезы о равенстве генеральных значений двух дисперсий по их оценкам s^2_{e2} с числом степеней свободы $\varphi_2 = N - m(k+1)$ и s^2_{e1} с числом степеней свободы $\varphi_1 = (m-1)(k+1)$.

В качестве критерия для проверки этой гипотезы можно воспользоваться отношением

$$F = \frac{s^2_{e1}}{s^2_{e2}}. \tag{9.24}$$

В условиях нормальности ошибки эксперимента это отношение имеет F-распределение с числами степеней свободы φ_1 и φ_2 [5]. Заметим, что привлекать F-распределение для проверки предположения об отсутствии дрейфа необходимо только тогда, когда $s^2_{e1} > s^2_{e2}$. Случай, когда $s^2_{e1} \leqslant s^2_{e2}$, свидетельствует о том, что результаты экспериментов не противоречат гипотезе об отсутствии дрейфа.

Процедура проверки гипотезы о равенстве дисперсий состоит в следующем (см. разд. 2.2.4): для заданного уровня значимости α и чисел степеней свободы φ_1 и φ_2 находят критическое значение F-распределения $F_{кр}$, пользуясь статистическими таблицами (см. табл. 15.4 и 15.5); если расчетное значение $F \leqslant F_{кр}$,

принимается гипотеза о равенстве дисперсий и делается вывод об отсутствии дрейфа; если $F > F_{кр}$, гипотеза об отсутствии дрейфа отклоняется. В последнем случае необходимо использовать методы построения моделей, учитывающие факт наличия дрейфа характеристик объекта.

Подводя итог содержанию этого подраздела, перечислим основные этапы решения задачи с помощью рассмотренного здесь метода.

1. Произвести m серий экспериментов на разнесенных интервалах времени.

2. По результатам каждой серии найти коэффициенты локальных моделей исследуемого процесса, решая систему линейных уравнений (9.9).

3. По формуле (9.17) рассчитать сумму квадратов $Q_э$, связанную с рассеянием наблюдений относительно локальных моделей.

4. Рассчитать оценку $s^2_{ээ}$ по формуле (9.19).

5. Путем решения системы линейных уравнений (9.14) найти оценки коэффициентов общей модели, построенной на основе наблюдений всех m серий.

6. По формуле (9.20) рассчитать сумму квадратов $Q_д$, связанную с различиями локальных моделей.

7. Вычислить оценку $s^2_{эд}$ по формуле (9.23).

8. Рассчитать $F = s^2_{эд}/s^2_{ээ}$ и, пользуясь F-распределением, проверить гипотезу о равенстве дисперсий, которым соответствуют оценки $s^2_{эд}$ и $s^2_{ээ}$. Если гипотеза принимается, делается вывод об отсутствии дрейфа.

Пример 9.1

Предположим, что зависимость между входной и выходной переменной объекта ищется в виде

$$y = a_0 + a_1 x_1. \tag{9.25}$$

Для проверки предположения об отсутствии дрейфа проведены $m = 3$ серии наблюдений на разнесенных во времени интервалах, каждая из которых состояла из трех опытов. Значения независимых переменных для этих опытов и результаты наблюдений зависимой переменной приведены в табл. 9.1.

Рассчитаем оценки коэффициентов модели для каждой серии экспериментов. Для этого необходимо для всякой серии, т. е. для $l = 1, 2, 3$, решить систему линейных уравнений типа (9.9):

$$\begin{aligned} \hat{a}_0^{(l)} N_l + \hat{a}_1^{(l)} \sum_{i=1}^{N_l} x_{1i}^l &= \sum_{i=1}^{N_l} \tilde{y}_i^l, \\ \hat{a}_0^{(l)} \sum_{i=1}^{N_l} x_{1i}^l + \hat{a}_1^{(l)} \sum_{i=1}^{N_l} (x_{1i}^l)^2 &= \sum_{i=1}^{N_l} \tilde{y}_i^l x_{1i}^l. \end{aligned} \tag{9.26}$$

Получаем

$$\hat{a}_0^{(1)} = 0{,}466, \quad \hat{a}_1^{(1)} = 0{,}85,$$
$$\hat{a}_0^{(2)} = 0{,}9, \quad \hat{a}_1^{(2)} = 0{,}6,$$
$$\hat{a}_0^{(3)} = 0{,}133, \quad \hat{a}_1^{(3)} = 1{,}1.$$

Остаточную сумму квадратов S_l для каждой серии опытов рассчитываем по формуле

$$S_l = \sum_{j=1}^{N_l} (\tilde{y}_l^j - \hat{a}_0^{(l)} - \hat{a}_1^{(l)} x_{1l}^j)^2. \tag{9.27}$$

Таблица 9.1

l	Серия 1		Серия 2		Серия 3	
	x_{11}^j	$\tilde{y}_1^j$	x_{12}^j	$\tilde{y}_2^j$	x_{13}^j	$\tilde{y}_3^j$
1	0	0,5	0	0,7	0	0,2
2	1	1,3	1	1,6	1	1,2
3	−1	−0,4	−1	−0,4	−1	−1

Получаем

$$S_1 = 0{,}0018, \quad S_2 = 0{,}06, \quad S_3 = 0{,}00668.$$

Отсюда находим величину Q_2:

$$Q_2 = \sum_{l=1}^{m} S_l = 0{,}06848.$$

Оценка $s_{\varepsilon 2}^2$ с числом степеней свободы

$$\varphi_2 = \sum_{l=1}^{3} (N_l - (k+1)) = 9 - 6 = 3$$

оказывается равной

$$s_{\varepsilon 2}^2 = \frac{0{,}06848}{3} = 0{,}0228.$$

Найдем теперь оценки коэффициентов a_0 и a_1 по результатам всех наблюдений. Для этого составим и решим систему уравнений (9.14):

$$\begin{aligned} \hat{a}_0 N + \hat{a}_1 \sum_{l=1}^{m} \sum_{j=1}^{N_l} x_{1l}^j &= \sum_{l=1}^{m} \sum_{j=1}^{N_l} \tilde{y}_l^j, \\ \hat{a}_0 \sum_{l=1}^{m} \sum_{j=1}^{N_l} x_{1l}^j + \hat{a}_1 \sum_{l=1}^{m} \sum_{j=1}^{N_l} (x_{1l}^j)^2 &= \sum_{l=1}^{m} \sum_{j=1}^{N_l} \tilde{y}_l^j x_{1l}^j. \end{aligned} \tag{9.28}$$

Получаем

$$\hat{a}_0 = 0{,}5, \quad \hat{a}_1 = 0{,}85.$$

Сумму квадратов Q_1 находим по формуле

$$Q_1 = \sum_{l=1}^{m} \sum_{f=1}^{N_l} (\hat{a}_0^{(l)} + \hat{a}_1^{(l)} x_{lf} - \hat{a}_0 - \hat{a}_1 x_{lf})^2. \tag{9.29}$$

Получаем

$$Q_1 = 1{,}13765.$$

Оценка $s_{\varepsilon 1}^2$ с числом степеней свободы $\varphi_1 = (m-1)(k+1) = 4$ оказывается равной

$$s_{\varepsilon 1}^2 = \frac{1{,}13765}{4} = 0{,}284.$$

Расчет F-отношения дает

$$F = s_{\varepsilon 1}^2 / s_{\varepsilon 2}^2 = 0{,}284/0{,}0228 = 12{,}4.$$

Пользуясь таблицами F-распределения (разд. 15), проверяем гипотезу об отсутствии дрейфа. Для уровня значимости $\alpha = 0{,}05$ и чисел степеней свободы $\varphi_1 = 4$ и $\varphi_2 = 3$ с помощью табл. 15.4 находим $F_{кр} = 9{,}1$. Так как $F_{кр} = 9{,}1 < F = 12{,}4$, то гипотеза об отсутствии дрейфа отклоняется.

9.3. Построение математических моделей в условиях аддитивного дрейфа, описываемого детерминированной функцией времени

Пусть модель исследуемого объекта имеет вид

$$y = a_0(t) + \sum_{i=1}^{k} a_i f_i(x). \tag{9.30}$$

Свободный член в уравнении (9.30) является функцией времени. Такая модель соответствует случаю аддитивного временно́го дрейфа. Мы будем предполагать, что изменение свободного члена уравнения (9.30) описывается следующей функцией времени:

$$a_0(t) = d_0\varphi_0(t) + d_1\varphi_1(t) + \ldots + d_m\varphi_m(t), \tag{9.31}$$

где $\varphi_j(t)$—заданная система функций.

Пусть в некоторые моменты времени $t_1, t_2, \ldots, t_N$ проведены эксперименты, т. е. измерены значения независимых переменных x и зависимой переменной y. Используя эти данные, можно методом наименьших квадратов найти оценки как коэффициентов a_i, $i = 1, 2, \ldots, k$, так и коэффициентов d_j, $j = 0, 1, 2, \ldots, m$. Мы будем предполагать, что общее число оцениваемых коэффициентов $k + m + 1 \leqslant N$.

Введем обозначения

$$\begin{aligned} f_0^j &= \varphi_0(t_j) = 1, \\ f_1^j &= f_1(t_j), \\ &\cdots \\ f_k^j &= f_k(t_j), \\ f_{k+1}^j &= \varphi_1(t_j) = \varphi_1^j, \\ &\cdots \\ f_{k+m}^j &= \varphi_m(t_j) = \varphi_m^j. \end{aligned} \tag{9.32}$$

Верхний индекс в (9.32) указывает номер эксперимента. Обозначим, кроме того,

$$a_0 = d_0; \quad a_{k+1} = d_1; \quad \ldots a_{k+m} = d_m.$$

Тогда имеем вектор a:

$$a = \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_{k+m} \end{bmatrix} = \begin{bmatrix} d_0 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_k \\ d_1 \\ d_m \end{bmatrix}$$

размерности $k+m+1$.

Матрица F, элементами которой являются значения функций f_l в N опытах, имеет вид

$$F = \begin{bmatrix} f_0^1 & f_1^1 & \ldots & f_{k+m}^1 \\ f_0^2 & f_1^2 & \ldots & f_{k+m}^2 \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ f_0^N & f_1^N & \ldots & f_{k+m}^N \end{bmatrix} = \begin{bmatrix} 1 & f_1^1 & \ldots & f_k^1 & \varphi_1^1 & \ldots & \varphi_m^1 \\ 1 & f_1^2 & \ldots & f_k^2 & \varphi_1^2 & \ldots & \varphi_m^2 \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ 1 & f_1^N & \ldots & f_k^N & \varphi_1^N & \ldots & \varphi_m^N \end{bmatrix}. \tag{9.33}$$

Размер этой матрицы $N \times (k+m+1)$. Матрица $F'F$ при этих условиях есть квадратная матрица размера $k+m+1$.

Оценки коэффициентов, соответствующие критерию наименьших квадратов, определяются из решения системы $k+m+1$ линейных уравнений

$$F'F\hat{a} = F'\tilde{Y}, \tag{9.34}$$

где $\tilde{Y}$ — вектор-столбец измеренных значений выходной переменной в N опытах:

$$\tilde{Y} = \begin{bmatrix} \tilde{y}^1 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{y}^N \end{bmatrix}.$$

При применении изложенного здесь метода может оказаться, что один или несколько столбцов φ_i матрицы F находятся в сильной корреляционной связи с каким-либо столбцом f_j. Это приведет к тому, что оценки коэффициентов при функциях φ_i и $f_j(x)$ окажутся сильно коррелированными. В случае если какой-то столбец φ_i окажется линейно связанным с каким-либо из столбцов f_j, оценки коэффициентов при функциях φ_i и f_j будут смешанными. При этом нельзя будет сделать правильного вывода о влиянии функции f_j на зависимую переменную. Естественно, что в случае линейной связи φ_i и f_j одна из этих функций должна быть исключена из уравнения регрессии, т. е. из матрицы F должен быть удален один из столбцов.

Выше мы уже указывали, что при построении математических моделей исследователя часто интересуют только коэффициенты при функциях от независимых переменных, ибо лишь величины этих коэффициентов определяют выбор значений независимых переменных, обеспечивающих экстремальную величину y. Поэтому влияние параметров дрейфа на величины коэффициентов при функциях независимых переменных является нежелательным. Чтобы избежать такого влияния, следует стремиться к тому, чтобы столбцы матрицы F, соответствующие функциям φ_i, были бы ортогональны столбцам, соответствующим функциям f_j. Условие ортогональности столбцов матрицы записывается следующим образом:

$$\sum_{\mu=1}^{N} \varphi_i^{\mu} f_j^{\mu} = 0, \quad i = 0, 1, 2, \ldots, m; \; j = 1, 2, \ldots, k. \quad (9.35)$$

При этом оценки коэффициентов при функциях от независимых переменных будут некоррелированы с оценками коэффициентов функции, описывающей дрейф, т. е. дрейф не будет оказывать влияния на оценки коэффициентов при функциях от независимых переменных.

Выполнение условия (9.35) может быть достигнуто двумя путями: 1) путем специального выбора функций $\varphi_i(t)$, когда выборка наблюдений и система функций $f_j(x)$ заданы; 2) путем выбора точек выборки x^j, $j = 1, 2, \ldots, N$, при заданной си-

стеме функций $\varphi_l(t)$. Последняя задача носит название задачи планирования эксперимента ортогонально к непрерывному временно́му дрейфу. К ее рассмотрению мы теперь и перейдем.

9.4. Планирование эксперимента в условиях аддитивного дрейфа

Пусть модель исследуемого объекта имеет вид

$$y = a_0(t) + \sum_{l=1}^{k} a_l x_l. \tag{9.36}$$

Эта модель соответствует случаю аддитивного дрейфа. Предположим, что дрейф описывается степенной функцией времени:

$$a_0(t) = c_0 + c_1 t + \ldots + c_m t^m. \tag{9.37}$$

Пусть исследователь имеет возможность поставить N экспериментов в моменты времени $t_1, t_2, \ldots, t_N$ (где $N \geqslant k + m + 1$), по результатам которых необходимо с помощью метода наименьших квадратов найти оценки коэффициентов a_l, $l = 1, 2, \ldots, k$, характеризующих степень влияния независимых переменных x на выходную переменную (отклик) y. Предположим, что в проводимых опытах исследователь имеет возможность выбирать значения переменных x^j, $j = 1, 2, \ldots, N$, по своему усмотрению, т. е. при решении задачи могут быть применены методы планирования эксперимента.

Задача планирования эксперимента здесь может быть поставлена следующим образом: найти такой план эксперимента X

$$X = \begin{bmatrix} x_1^1 & \ldots & x_k^1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_1^N & \ldots & x_k^N \end{bmatrix},$$

чтобы

1) исключить влияние дрейфа на величины оценок коэффициентов $\hat{a}_l$, $l = 1, 2, \ldots, k$;

2) обеспечить минимальные величины σ_l^2, $l = 1, 2, \ldots, k$, дисперсий оценок коэффициентов при независимых переменных.

Как уже указывалось выше, для выполнения первого требования необходимо обеспечить условие (9.35) ортогональности столбцов матрицы F, соответствующих коэффициентам при функциях, описывающих дрейф, и столбцов независимых переменных. Это требование мы будем называть требованием ортогональности плана к дрейфу.

Второе условие требует максимальной точности оценок коэффициентов модели. В соответствии с теоремой Бокса [6, 7]

при оценивании $k+1$ коэффициентов зависимости вида

$$y = \sum_{i=0}^{k} a_i x_i$$

с помощью плана X, где X — матрица ранга $k+1$, для которой имеют место условия

$$\sum_{l=1}^{N} x_l^i = 0, \tag{9.38а}$$

$$\sum_{l=1}^{N} (x_l^i)^2 = N, \tag{9.38б}$$

необходимым и достаточным условием того, что дисперсии оценок коэффициентов принимают минимальные значения, является условие ортогональности столбцов матрицы X, т. е.

$$\sum_{l=1}^{N} x_l^i x_l^r = 0 \text{ при } i, r = 1, 2, \ldots, k;\ i \neq r. \tag{9.38в}$$

Заметим, что план X, удовлетворяющий условию (9.38в), называется ортогональным. Таким образом, среди всех планов, для которых справедливы условия (9.38а) и (9.38б), наилучшим с точки зрения дисперсий оценок коэффициентов является ортогональный план.

Рассматриваемая здесь задача построения зависимости (9.36), линейной относительно переменных x, часто возникает при осуществлении экспериментального поиска наилучших условий ведения процесса в условиях быстрого временного дрейфа [11].

Идея излагаемого ниже метода планирования эксперимента заключается в следующем. Строятся ортогональные степенные полиномы $\varphi_l(t)$ (полиномы Чебышева) для значений аргумента $t_1, t_2, \ldots, t_N$. Для данного числа N экспериментов можно найти N ортогональных полиномов $\varphi_l(t)$ с максимальной степенью $N-1$. Значения полиномов $\varphi_l(t)$ для N значений аргумента t, соответствующих моментам проведения опытов, образуют матрицу Φ размера $N \times N$.

$$\Phi = \begin{bmatrix} \varphi_0^1 & \varphi_1^1 & \ldots & \varphi_{N-1}^1 \\ \varphi_0^2 & \varphi_1^2 & \ldots & \varphi_{N-1}^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \varphi_0^N & \varphi_1^N & \ldots & \varphi_{N-1}^N \end{bmatrix}.$$

Первые $m+1$ столбцов матрицы Φ соответствуют полиномам степени $0, 1, \ldots, m$. Эти полиномы используются для описания

дрейфа:

$$a_0(t) = d_0\varphi_0(t) + d_1\varphi_1(t) + \ldots + d_m\varphi_m(t). \qquad (9.39)$$

Оставшиеся $N-(m+1)$ столбцов матрицы Φ используются для построения плана эксперимента X, например, следующим образом. Положим

$$x_j^{\mu} = \varphi_{m+j}^{\mu} \text{ при } \mu = 1, 2, \ldots, N \text{ и } j = 1, 2, \ldots, k. \qquad (9.40)$$

Тогда оценка коэффициента при переменной x_j в силу ортогональности функций φ_l будет не коррелирована с оценками коэффициентов при функциях, использованных для описания дрейфа, что и составляло первое требование к некомому плану. Кроме того, полученный план является ортогональным и, следовательно, наилучшим с точки зрения дисперсий оценок коэффициентов среди всех планов, для которых имеют место условия (9.38а) и (9.38б).

В практических задачах, однако, выбирать план эксперимента, исходя из условия (9.40), было бы неправильным, если допускается возможность ошибки в априорном задании степени полинома, описывающего дрейф. Действительно, пусть исследователь предполагает, что дрейф описывается полиномом степени m, тогда как на самом деле характер дрейфа соответствует полиному степени $m+1$. Тогда выбор плана из условий (9.40) приведет к тому, что оценка коэффициента при переменной x_1 будет отражать воздействие на выходную переменную как переменной x_1, так и описывающей дрейф функции $\varphi_{m+1}(t)$. Поэтому при выборе плана применяют специальный прием рандомизации. Ниже этот прием будет подробно рассмотрен.

Приведем поэтапное описание процедуры выбора плана эксперимента и построения математической модели объекта в условиях аддитивного дрейфа. Первый этап — построение системы ортогональных полиномов до степени $N-1$ включительно для значений аргумента $t_1, t_2, \ldots, t_N$. Напомним, что N — число планируемых экспериментов, выбранное из условия $N \geqslant k+m+1$, а t_l — момент времени проведения l-го эксперимента. Условия ортогональности для некоторой системы полиномов $\varphi_0(t), \varphi_1(t), \ldots, \varphi_{N-1}(t)$ при значениях аргумента $t_1, t_2, \ldots, t_N$ имеют следующий вид:

$$\sum_{j=1}^{N} \varphi_l(t_j)\varphi_r(t_j) \begin{cases} = 0 \text{ при } l \neq r, \\ \neq 0 \text{ при } l = r, \end{cases} \quad l, r = 0, 1, \ldots, N-1. \qquad (9.41)$$

Вычисление ортогональных полиномов $\varphi_l(t)$ может быть осуществлено с помощью следующей рекуррентной процедуры. Полином $\varphi_0(t)$ принимается равным единице. Полином $\varphi_1(t)$

ищется в виде

$$\varphi_1(t) = t - b_{10}\varphi_0(t), \tag{9.42}$$

где коэффициент b_{10} определяется из условия

$$\sum_{j=1}^{N} \varphi_1(t_j)\,\varphi_0(t_j) = 0. \tag{9.43}$$

Нетрудно убедиться, что

$$b_{10} = \frac{\sum_{j=1}^{N} t_j \varphi_0(t_j)}{\sum_{j=1}^{N} \varphi_0^2(t_j)} = \frac{\sum_{j=1}^{N} t_j}{N}.$$

Полином $\varphi_2(t)$ ищется в виде

$$\varphi_2(t) = t^2 - b_{21}\varphi_1(t) - b_{20}\varphi_0(t). \tag{9.44}$$

Коэффициенты b_{21} и b_{20} определяются из условий

$$\begin{aligned} \sum_{j=1}^{N} \varphi_2(t_j)\,\varphi_0(t_j) &= 0, \\ \sum_{j=1}^{N} \varphi_2(t_j)\,\varphi_1(t_j) &= 0 \end{aligned} \tag{9.45}$$

по формулам

$$b_{20} = \frac{\sum_{j=1}^{N} t_j^2}{N}, \qquad b_{21} = \frac{\sum_{j=1}^{N} t_j^2 \varphi_1(t_j)}{\sum_{j=1}^{N} \varphi_1^2(t_j)}.$$

Соответственно l-й ортогональный полином имеет вид

$$\varphi_l(t) = t^l - b_{l,\,l-1}\varphi_{l-1}(t) - b_{l,\,l-2}\varphi_{l-2}(t) - \ldots - b_{l0}\varphi_0(t), \tag{9.46}$$

где

$$b_{lr} = \frac{\sum_{j=1}^{N} t_j^l \varphi_r(t_j)}{\sum_{j=1}^{N} \varphi_r^2(t_j)}. \tag{9.47}$$

Для случая когда эксперименты проводятся через равные интервалы времени, значения ортогональных полиномов Чебышева для различных чисел опытов сведены в таблицы [8]. Аргумент t_j в этих таблицах принимает значения 1, 2, 3, ..., N,

Таблица 9.2

j	$\varphi^j_{0\tau}$	$\varphi^j_{1\tau}$	$\varphi^j_{2\tau}$	$\varphi^j_{3\tau}$	$\varphi^j_{4\tau}$	$\varphi^j_{5\tau}$	$\varphi^j_{6\tau}$	$\varphi^j_{7\tau}$	$\varphi^j_{8\tau}$	$\varphi^j_{9\tau}$	$\varphi^j_{10\tau}$	$\varphi^j_{11\tau}$
1	1	−11	55	−33	33	−33	11	−55	+11	−11	1	−1
2	1	−9	25	3	−27	57	−31	225	−61	79	−9	11
3	1	−7	1	21	−33	21	11	−251	119	−227	35	−55
4	1	−5	−17	25	−13	−29	25	−83	−65	303	−75	165
5	1	−3	−29	19	12	−44	4	204	−74	−102	90	−330
6	1	−1	−35	7	28	−20	−20	140	70	−210	−42	462
7	1	+1	−35	−7	28	20	−20	−140	70	210	−42	−462
8	1	3	−29	−19	12	44	4	−204	−74	102	90	330
9	1	5	−17	−25	−13	29	25	83	−65	−303	−75	−165
10	1	7	1	−21	−33	−21	11	251	+119	227	+35	55
11	1	9	25	−3	−27	−57	−31	−225	−61	−79	−9	−11
12	1	11	55	33	33	33	11	55	11	11	1	1
$\sum_{j=1}^{N} (\varphi^j_{i\tau})^2$	12	572	12012	5148	8008	15912	4488	369512	65208	408408	33592	705432

совпадающие с номером эксперимента. Отметим, что в таблицах приводят обычно значения полиномов, отличающиеся от рассчитанных по описанной выше процедуре на некоторый постоянный для данного полинома множитель, т. е. $\varphi_{lk}^{t} = \lambda_l \varphi_{lk}^{l}$, где φ_{lk}^{t}—табличное значение полинома, λ_l—постоянный множитель. Значения λ_l выбираются так, чтобы в таблице содержались только целые числа. Очевидно, что введение множителей λ_l не отражается на условии ортогональности полиномов. Пример ортогональных полиномов для $N = 12$ приведен в табл. 9.2. В последней графе табл. 9.2 указаны суммы квадратов каждого из ортогональных полиномов, подсчитанные по N опытам. Эта величина используется при нормировании полиномов. Нормированные полиномы имеют значение суммы квадратов, равное N, и вычисляются по формуле

$$\psi_l^j = \psi_l(l_j) = \frac{\varphi_l^j \sqrt{N}}{\sqrt{\sum_{j=1}^{N} (\varphi_l^j)^2}}. \qquad (9.48)$$

Матрицу размера $N \times N$ значений нормированных ортогональных полиномов в N опытах обозначим через Ψ:

$$\Psi = \begin{bmatrix} \psi_0^1 & \psi_1^1 & \dots & \psi_{N-1}^1 \\ \psi_0^2 & \psi_1^2 & \dots & \psi_{N-1}^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \psi_0^N & \psi_1^N & \dots & \psi_{N-1}^N \end{bmatrix}.$$

Рассмотрим теперь второй этап решения задачи—составление плана эксперимента. Выше уже говорилось о том, что для этой цели используются столбцы значений ортогональных полиномов степени $m+1$, $m+2$, ..., $N-1$. Опишем здесь прием рандомизации плана [9, 10]. Цель рандомизации заключается в том, чтобы придать случайный характер степени влияния неучтенных параметров дрейфа (т. е. коэффициентов при $\varphi_{m+j}(l)$, $j > 0$, функции, описывающей дрейф) на оценки коэффициентов модели. Для этой цели план эксперимента X выбирают следующим образом:

$$X = \Psi_m W, \qquad (9.49)$$

где X—матрица размера $N \times k$, определяющая значения независимых переменных в N планируемых опытах: Ψ_m—матрица нормированных значений ортогональных полиномов степени

$m+1$, $m+2$, ..., $N-1$ в N опытах:

$$\Psi_m = \begin{bmatrix} \psi_{m+1}^1 & \psi_{m+2}^1 & \dots & \psi_{N-1}^1 \\ \psi_{m+1}^2 & \psi_{m+2}^2 & \dots & \psi_{N-1}^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \psi_{m+1}^N & \psi_{m+2}^N & \dots & \psi_{N-1}^N \end{bmatrix}; \tag{9.50}$$

W — случайная матрица размера $(N-m-1)\times k$:

$$W = \begin{bmatrix} w_1^1 & w_2^1 & \dots & w_k^1 \\ w_1^2 & w_2^2 & \dots & w_k^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ w_1^{N-m-1} & w_2^{N-m-1} & \dots & w_k^{N-m-1} \end{bmatrix}. \tag{9.51}$$

Предположим, что матрица W нормирована таким образом, что

$$\sum_{j=1}^{N-m-1} (w_l^j)^2 = 1. \tag{9.52}$$

Очевидно, что значение l-й переменной (l-го фактора) в j-м опыте равно

$$x_l^j = \sum_{r=1}^{N-m-1} \psi_{r+m}^j w_l^r, \tag{9.53}$$

т. е. представляет собой линейную комбинацию значений $N-m-1$ ортогональных полиномов в j-м опыте, взятых со случайными весами. Нетрудно убедиться в том, что столбцы матрицы X ортогональны m первым столбцам матрицы Ψ, т. е. план X будет ортогонален к дрейфу, описываемому полиномом степени m.

Так как из выражений (9.43) и (9.45) следует, что

$$\sum_{j=1}^{N} \varphi_l(t_j) = \sum_{j=1}^{N} \varphi_l^j = 0,$$

то $\sum_{j=1}^{N} \psi_l^j = 0$, и условие (9.38а) для плана X выполняется. Условие (9.38б) также будет выполняться для плана X, если имеет место (9.52). Легко убедиться теперь, что для того, чтобы план X был ортогональным, достаточно, чтобы матрица W была бы ортогональной. С учетом условия (9.52) требование

ортогональности матрицы W можно записать так:

$$\sum_{j=1}^{N-m-1} w_i^j w_r^j = \begin{cases} 0 \text{ при } i \neq r, \\ 1 \text{ при } i = r. \end{cases} \tag{9.54}$$

Таким образом, если выбрать план по правилу (9.49), где Ψ_m и W — нормированные ортогональные матрицы, то полученный план будет ортогонален к дрейфу и обеспечит минимальные дисперсии оценок коэффициентов модели среди всех планов, для которых имеют место условия (9.38а) и (9.38б). Кроме того, так как столбцы такого плана формируются в виде линейной комбинации столбцов матрицы Ψ_m, взятых со случайными коэффициентами, то степень влияния некоторого неучтенного параметра дрейфа на оценки коэффициентов модели становится случайной величиной.

Приведем здесь процедуру получения W [10, 11].

1. Из таблиц случайных чисел [8] выбирают k столбцов случайных чисел с $N-m-1$ элементом в каждом:

$$\xi_1 = \begin{bmatrix} \xi_1^1 \\ \xi_1^2 \\ \cdot \\ \cdot \\ \cdot \\ \xi_1^{N-m-1} \end{bmatrix}; \ \ldots; \ \xi_k = \begin{bmatrix} \xi_k^1 \\ \xi_k^2 \\ \cdot \\ \cdot \\ \cdot \\ \xi_k^{N-m-1} \end{bmatrix}. \tag{9.55}$$

2. Полученные векторы-столбцы ортогонализируют. Для этого полагают

$$\begin{aligned} \xi_1^* &= \xi_1, \\ \xi_2^* &= \xi_2 - c_{21}\xi_1^*, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \xi_l^* &= \xi_l - c_{l,\,l-1}\xi_{l-1}^* - \ldots - c_{l1}\xi_1^*. \end{aligned} \tag{9.56}$$

Коэффициенты c_{lj} определяются последовательно из условия ортогональности ξ_l^* всем ранее построенным векторам ξ_j^* по формуле

$$c_{lj} = \frac{\sum_{n=1}^{N-m-1} \xi_l^n \xi_j^{*n}}{\sum_{n=1}^{N-m-1} [\xi_j^{*n}]^2}, \quad j = 1, 2, \ldots, j-1. \tag{9.57}$$

3. Полученные векторы ξ_l^* нормируют по формуле

$$w_l^j = \frac{\xi_l^{*j}}{\sum_{j=1}^{N-m-1} [\xi_l^{*j}]^2}. \tag{9.58}$$

Пример 9.2

Рассмотрим в качестве примера формирование ортогональной случайной матрицы W размера 8×2 [10]. Пользуясь таблицами случайных чисел, находим векторы ξ_1 и ξ_2:

$$\xi_1 = \begin{bmatrix} 0{,}101 \\ 0{,}432 \\ 0{,}555 \\ -0{,}114 \\ 1{,}918 \\ -0{,}739 \\ -0{,}044 \\ -0{,}044 \\ 1{,}515 \end{bmatrix}, \qquad \xi_2 = \begin{bmatrix} 0{,}013 \\ -0{,}645 \\ 0{,}121 \\ -0{,}476 \\ -0{,}032 \\ -0{,}008 \\ 0{,}942 \\ 0{,}705 \end{bmatrix}.$$

Вычисляем вектор ξ_2^* по формуле (9.56):

$$\xi_2^{*\prime} = \xi_2' - c_{21}\xi_1',$$

где

$$c_{21} = \frac{\sum_{i=1}^{8} \xi_i' \xi_2'}{\sum_{i=1}^{8} (\xi_i')^2} = \frac{0{,}815225}{7{,}039852} = 0{,}115806.$$

В табл. 9.3 выписаны элементы векторов ξ_1^* и ξ_2^* и нормированных векторов w_1 и w_2, подсчитанные по формуле (9.58).

Таблица 9.3

ξ_1^*	ξ_2^*	w_1	w_2
0,101	0,001	0,038	0,001
0,432	−0,695	0,163	−0,498
0,555	0,057	0,210	0,041
−0,114	−0,463	−0,043	−0,332
1,918	−0,254	0,726	−0,182
−0,739	0,078	−0,280	0,056
−0,044	0,947	−0,017	0,678
1,515	0,530	0,573	0,380

Перейдем теперь к рассмотрению последнего этапа решения задачи — вычислению оценок коэффициентов модели.

Аналогично (9.32) введем обозначения

$$\begin{aligned} f_0^l &= 1, \\ f_1^l &= x_1^l, \\ &\cdots \\ f_k^l &= x_k^l, \\ f_{k+1}^l &= \psi_1^l, \\ &\cdots \\ f_{k+m}^l &= \psi_m^l. \end{aligned}$$

Модель (9.36) исследуемого процесса в новых обозначениях имеет вид

$$y = \sum_{i=0}^{k+m} a_i f_i. \tag{9.59}$$

Коэффициенты a_i с индексом $i = k+1, \ldots, k+m$ относятся к описанию дрейфа. Оценки $\hat{a}$ для модели вида (9.59), рассчитанные на основе метода наименьших квадратов по выборке, состоящей из N опытов, определяются выражением

$$\hat{a} = (F'F)^{-1} F'\tilde{Y},$$

где $\tilde{Y}$ — вектор-столбец результатов измерения выходной переменной, а F — матрица размера $N \times (m+1+k)$:

$$F = \begin{bmatrix} f_0^1 & f_1^1 \ldots f_{m+k}^1 \\ f_0^2 & f_1^2 \ldots f_{m+k}^2 \\ \cdot & \\ \cdot & \\ \cdot & \\ f_0^N & f_1^N \ldots f_{m+k}^N \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 \ldots x_k^1 & \psi_1^1 \ldots \psi_m^1 \\ 1 & x_1^2 \ldots x_k^2 & \psi_1^2 \ldots \psi_m^2 \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ 1 & x_1^N \ldots x_k^N & \psi_1^N \ldots \psi_m^N \end{bmatrix}.$$

Заметим, что ψ_i — нормированные ортогональные полиномы, причем нормирование выполняется по формуле (9.48), т. е. таким образом, что сумма квадратов значений полинома ψ_i, подсчитанная по всем N опытам, равна N. Теперь, имея в виду, что план X ортогонален и для него справедливо условие (9.38б), нетрудно убедиться в том, что

$$F'F = NI_{k+m+1}, \tag{9.60}$$

где I_{k+m+1} — единичная матрица размера $k+m+1$. Отсюда сле-

дует, что оценки $\hat{a}_l$ могут быть подсчитаны по формуле

$$\hat{a}_l = \frac{1}{N} F'\tilde{Y} = \frac{\sum_{j=1}^{N} f_l^j \tilde{y}^j}{N}. \qquad (9.61)$$

Остаточная сумма квадратов определяется выражением

$$S_R = \sum_{j=1}^{N} (\tilde{y}^j)^2 - N \sum_{l=0}^{k+m} \hat{a}_l^2, \qquad (9.62)$$

а остаточная дисперсия — выражением

$$s_R^2 = \frac{S_R}{N-(m+k+1)}. \qquad (9.63)$$

Оценки дисперсий всех коэффициентов одинаковы и вычисляются по формуле

$$s_l^2 = \frac{s_R^2}{N}, \quad l = 0, 1, \ldots, k+m. \qquad (9.64)$$

Если случайная ошибка эксперимента подчиняется нормальному закону распределения, найденное математическое описание может быть подвергнуто статистическому анализу с целью проверки значимости коэффициентов (проверка гипотезы о равенстве истинных значений коэффициентов нулю). Для этого в данном случае можно воспользоваться как t-критерием, так и F-критерием. Согласно разд. 3, при использовании t-критерия проверка значимости коэффициента a_l сводится к проверке условия

$$|\hat{a}_l| > t_{кр} \frac{s_R}{\sqrt{N}}, \qquad (9.65)$$

где $t_{кр}$ — критическое значение t-распределения при заданном уровне значимости α и числе степеней свободы $\varphi = N - (m+k+1)$. При использовании F-критерия для каждого коэффициента рассчитывается величина

$$F_l = \frac{\hat{a}_l^2}{s_l^2} = \frac{N\hat{a}_l^2}{s_R^2}. \qquad (9.66)$$

Выражение (9.66) является отношением оценок двух дисперсий. В числителе здесь записана дисперсия

$$\tilde{s}_l^2 = \frac{N\hat{a}_l^2}{\varphi_1}, \text{ где } \varphi_1 = 1, \qquad (9.67)$$

определяемая вкладом, вносимым l-м коэффициентом в остаточную сумму квадратов. Число степеней свободы для этой оценки равно 1. В знаменателе выражения (9.66) записана оста-

точная дисперсия s_R^2, которой соответствует число степеней свободы $\varphi_2 = N-(m+k+1)$. Обе дисперсии в выражении (9.66) при условии правильности предположения о виде модели и равенстве нулю истинного значения коэффициента a_l являются оценками одной и той же величины — дисперсии ошибки наблюдений. С помощью F-критерия здесь фактически проверяется гипотеза о равенстве генеральных значений дисперсий, соответствующих оценкам $\tilde{s}_l^2$ и s_R^2. Рассчитанное значение F_l сравнивается с критическим значением $F_{кр}$, выбираемым по табл. 15.4 и 15.5 при заданном уровне значимости α и числах степеней свободы φ_1 и φ_2. Если

$$F_l \geqslant F_{кр},$$

рассматриваемый коэффициент считается значимым. Заметим, что применение при анализе t- и F-критериев приводит к одним и тем же результатам.

Рассмотренный здесь статистический анализ может быть проведен как для коэффициентов, описывающих дрейф, так и для коэффициентов при независимых переменных.

Пример 9.3

Рассмотрим пример применения изложенного метода [9, 10]. Пусть требуется оценить коэффициенты при линейных членах в задаче с двумя независимыми переменными в условиях аддитивного временного дрейфа, описываемого полиномом степени $m=3$, причем для решения задачи предполагается через равные промежутки времени провести $N=12$ экспериментов.

Для составления плана эксперимента воспользуемся табл. 9.2 ортогональных полиномов Чебышева для $N=12$. План эксперимента будем искать в соответствии с формулой (9.49), где Ψ_m — матрица столбцов значений ортогональных нормированных полиномов степени $m+1$, $m+2, \ldots, N-1$. Очевидно, что размер этой матрицы в нашей задаче 12×8. Для использования формулы (9.49) необходимо найти случайную ортогональную матрицу W размера 8×2. Воспользуемся матрицей W, полученной в примере 9.2. Элементы этой матрицы приведены в табл. 9.3. Элемент x_l^j матрицы планирования X при этом следует вычислять по формуле

$$x_l^j = \sum_{r=1}^{N-m-1} \psi_{m+r}^j w_l^r = \sum_{r=1}^{8} \frac{\varphi_{l+r}^j \sqrt{12}}{\sum_{j=1}^{12} (\varphi_{l+a}^j)^2} w_l^r. \qquad (9.68)$$

В выражении (9.68) через φ_{m+r}^j обозначен элемент табл. 9.2, стоящий в j-й строке и $(m+r)$-м столбце. Элемент x_1^1 матрицы X

вычисляется, например, следующим образом:

$$x_1^1 = \frac{33\sqrt{12}}{\sqrt{8008}}\,0{,}038 + \frac{(-33)\sqrt{12}}{\sqrt{15912}}\,0{,}163 + \ldots + \frac{(-1)\,12}{\sqrt{705432}}\,0{,}573 = 0{,}396.$$

Полученная в результате расчетов матрица X и результаты измерения выходной переменной в точках плана приведены в табл. 9.4.

Таблица 9.4

Номер опыта l	x_1^l	x_2^l	$\tilde{y}^l$
1	0,4	−0,5	100,3
2	−1,1	1,1	84,9
3	0,9	0,1	98,5
4	−0,2	−1,6	99,3
5	0,2	1,7	83,1
6	−0,8	−1,4	87,4
7	1,8	−0,1	95,9
8	−1,6	1,0	65,5
9	−0,5	0,2	74,8
10	1,6	0,2	88,0
11	−0,9	−1,2	76,2
12	0,2	0,6	75,4

Следующий этап решения задачи — вычисление оценок коэффициентов модели по формуле (9.61). Для оценки коэффициента $\hat{a}_0$ применение выражения (9.61) дает

$$\hat{a}_0 = \frac{\sum_{l=1}^{12} \tilde{y}^l}{12} = 85{,}8.$$

Для коэффициентов при независимых переменных x_1 и x_2 получаем оценки

$$\hat{a}_1 = \frac{\sum_{l=1}^{12} x_1^l \tilde{y}^l}{12} = 6{,}7,$$

$$\hat{a}_2 = \frac{\sum_{l=1}^{12} x_2^l \tilde{y}^l}{12} = -3{,}7.$$

Аналогичным образом можно вычислить оценки коэффициентов $\hat{a}_3$, $\hat{a}_4$, $\hat{a}_5$ при полиномах, описывающих дрейф:

$$\hat{a}_3 = \frac{\sum_{l=1}^{12} \psi_1^l \tilde{y}^l}{12} = \frac{\sum_{l=1}^{12} \frac{\varphi_1^l \sqrt{12}}{\sqrt{572}} \tilde{y}^l}{12} = -7{,}16,$$

$$\hat{a}_4 = \frac{\sum_{l=1}^{12} \frac{\varphi_2^l \sqrt{12}}{\sqrt{120125}} \tilde{y}^l}{12} = 0{,}416,$$

$$\hat{a}_5 = \frac{\sum_{l=1}^{12} \frac{\varphi_3^l \sqrt{12}}{5148} \tilde{y}^l}{12} = 1{,}26.$$

Остаточную сумму квадратов находим по формуле (9.62):

$$S_R = 70{,}55.$$

Остаточная дисперсия в соответствии с выражением (9.63) равна

$$s_R^2 = \frac{S_R}{12 - (3 + 1 + 2)} = \frac{70{,}55}{6} = 11{,}76.$$

Проверяем гипотезу о значимости коэффициентов $\hat{a}_1$ и $\hat{a}_2$ с помощью F-критерия. Для этого рассчитываем

$$F_1 = \frac{12\hat{a}_1^2}{s_R^2} = \frac{12\,(6{,}7)^2}{11{,}76} = 45{,}3,$$

$$F_2 = \frac{12\hat{a}_2^2}{s_R^2} = \frac{12\,(3{,}7)^2}{11{,}76} = 14{,}1.$$

Выбрав уровень значимости $\alpha = 0{,}05$, находим по таблице F-распределения 15.4 для $\varphi_1 = 1$ и $\varphi_2 = 6$

$$F_{кр} = 6.$$

Так как F_1 и F_2 превосходят $F_{кр} = 6$, то коэффициенты при переменных x_1 и x_2 значимо отличаются от нуля. Заметим, что с оценками $\hat{a}_3$, $\hat{a}_4$, $\hat{a}_5$ связана сумма квадратов S, равная

$$S = 12 \sum_{l=3}^{5} \hat{a}_l^2 = 635{,}78.$$

Это значит, что если бы мы не учли дрейф, то остаточная сумма квадратов увеличилась бы на 635,78. При этом остаточная дисперсия была бы равна

$$(s_R')^2 = \frac{70{,}55 + 635{,}78}{12 - 3} = 78{,}48.$$

В этом случае дисперсия, связанная с временным дрейфом, оказывается включенной в дисперсию, характеризующую ошибку опыта. Естественно, что это приводит к значительному ухудшению точности оценки выходной переменной по полученной модели.

9.5. Адаптивный метод построения математической модели в условиях неаддитивного случайного дрейфа

9.5.1. Постановка задачи

Пусть модель исследуемого объекта имеет вид

$$y = a'(t)\, f[x(t)] = \sum_{l=0}^{k} a_l(t)\, f_l[x(t)], \tag{9.69}$$

где $a(t)$ — вектор-столбец размерности $k+1$ коэффициентов в момент времени t; $x(t)$ — вектор независимых переменных размерности n; $f[x(t)]$ — вектор-столбец функций от независимых переменных размерности $k+1$.

Положим, что $a(t)$ — эргодический стационарный случайный процесс с монотонно убывающей корреляционной функцией. Примем, кроме того, что значения функций независимых переменных также образуют случайный стационарный процесс, для которого справедлива гипотеза эргодичности [1]).

Допустим, что через равноотстоящие интервалы времени произведено N наблюдений за входными и выходной переменной объекта. Собранная информация представлена в виде матриц

$$F_N = \begin{bmatrix} f_0^1 & f_1^1 \ldots f_k^1 \\ f_0^2 & f_1^2 \ldots f_k^2 \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ f_0^N & f_1^N \ldots f_k^N \end{bmatrix} = \begin{bmatrix} f'(1) \\ f'(2) \\ \cdot \\ \cdot \\ \cdot \\ f'(N) \end{bmatrix}, \tag{9.70}$$

$$\tilde{Y}_N = \begin{bmatrix} \tilde{y}^1 \\ \tilde{y}^2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{y}^N \end{bmatrix} = \begin{bmatrix} \tilde{y}(1) \\ \tilde{y}(2) \\ \cdot \\ \cdot \\ \cdot \\ \tilde{y}(N) \end{bmatrix}. \tag{9.71}$$

[1]) Относительно требования эргодичности процессов $a(t)$ и $f[x(t)]$ см. примечание на стр. 309.

В формулах (9.70) и (9.71) верхний индекс определяет момент соответствующего наблюдения. Все процессы мы будем рассматривать только в дискретные моменты времени, кратные величине интервала между опытами. Эти моменты мы будем обозначать числами натурального ряда, полагая

$$t_i = i = 1, 2, 3, \ldots .$$

Как и всюду ранее, полагаем, что измерение выходной переменной в момент t_i осуществляется с некоторой ошибкой $e(t_i)$, так что измеренное значение выражается в виде

$$\tilde{y}(t_i) = \tilde{y}(i) = \bar{a}'(i) f[x(i)] + \tilde{e}(i), \tag{9.72}$$

где $\bar{a}(i)$ — действительные значения коэффициентов в момент времени t_i. Будем считать, что $e(i)$ — независимая случайная последовательность с нулевым математическим ожиданием, так что

$$E\{e(i)\, e(j)\} = \begin{cases} 0, & i \neq j, \\ \sigma_e^2, & i = j, \end{cases}$$
$$E\{e(i) f[x(j)]\} = E\{e(i) \bar{a}(j)\} = 0. \tag{9.73}$$

На основании собранной информации требуется найти оценку $\hat{a}(N+\alpha)$ вектора коэффициентов в момент времени $N+\alpha$, где $\alpha = 1, 2, \ldots$. Оценка $\hat{a}(N+\alpha)$ должна удовлетворять следующему условию оптимальности:

$$E\{[y(N+\alpha) - \hat{a}'(N+\alpha) f(N+\alpha)]^2\} = \min_{a(N+\alpha)} E\{[y(N+\alpha) - a'(N+\alpha) f(N+\alpha)]^2\}, \tag{9.74}$$

где $y(N+\alpha)$ — величина выходной переменной в момент времени $N+\alpha$.

Сформулированная здесь задача является задачей прогнозирования параметров дрейфующего объекта, причем прогноз осуществляется на время α, кратное интервалу между наблюдениями. Такого рода задачи возникают там, где существует временной сдвиг между воздействием x и полученным откликом y о его результате. Причиной такого сдвига могут быть, например, динамические свойства объекта или запаздывания, возникающие в случаях, когда невозможно мгновенно установить значение выходной переменной (например, в случаях, когда измерения связаны с лабораторными анализами).

9.5.2. Метод решения

Синтез оптимального в смысле условия (9.74) алгоритма для решения сформулированной выше задачи возможен при известных статистических свойствах процесса $a(t)$, например в случае знания корреляционной функции этого процесса. На практике такие сведения, как правило, отсутствуют. Излагаемый ниже метод ввиду этого не предполагает знания статистических свойств процесса $a(t)$.

Метод характеризуется следующими особенностями:

1. Построение оптимального в смысле условия (9.74) алгоритма решения сводится к выбору значения некоторого параметра γ алгоритма, причем структура алгоритма считается заданной.

2. Выбор оптимального значения параметра γ осуществляется на основе имеющихся в распоряжении исследователя экспериментальных данных.

Обычно такой способ решения задачи построения оптимального алгоритма называют адаптивным.

Метод решения состоит в следующем.

Рассчитываем последовательность оценок коэффициентов модели для моментов времени $l=\alpha, \alpha+1, \ldots, \alpha+N$ по рекуррентной формуле

$$\hat{a}(1+j+\alpha, \gamma)=\hat{a}(j+\alpha, \gamma)+ \\ +\frac{C_j f(j+1)}{f'(j+1) C_j f(j+1)+\frac{\gamma}{1-\gamma}}\left[\tilde{y}(j+1)-f'(j+1)\hat{a}(j+\alpha, \gamma)\right]. \quad (9.75)$$

Здесь C_j—матрица размера $(k+1)\times(k+1)$, которая рассчитывается рекуррентно по формуле

$$C_j=\frac{1}{\gamma}\left[C_{j-1}-\frac{C_{j-1} f(j) f'(j) C_{j-1}}{f'(j) C_j f(j)+\frac{\gamma}{1-\gamma}}\right]. \quad (9.76)$$

Начальные значения C_0 и $\hat{a}(\alpha, \gamma)$ определяем по формулам

$$C_0=N(F'_N F_N)^{-1}, \quad (9.77)$$

$$\hat{a}(\alpha, \gamma)=(F'_N F_N)^{-1} F'_N \tilde{Y}_N. \quad (9.78)$$

Значение параметра γ задается в интервале

$$\delta \leqslant \gamma \leqslant 1, \quad (9.79)$$

где δ—малая положительная величина.

Применение уравнений (9.75) и (9.76) для $j=0, 1, \ldots, N-1$ позволяет получить последовательность оценок

$$\hat{a}(1+\alpha, \gamma),\ \hat{a}(2+\alpha, \gamma), \ldots, \hat{a}(N+\alpha, \gamma).$$

При этом оценка $\hat{a}(N+\alpha, \gamma)$ минимизирует следующую величину:

$$Q^* = \gamma^N (\hat{a}(\alpha, \gamma) - a(N+\alpha, \gamma))' C_0^{-1} (\hat{a}(\alpha, \gamma) - a(N+\alpha, \gamma)) + (1-\gamma) \sum_{j=1}^{N} \gamma^{N-j} [\tilde{y}(j) - a'(N+\alpha, \gamma) f(j)]^2. \quad (9.80)$$

Легко убедиться в том, что при увеличении N первое слагаемое выражения (9.80) стремится к нулю (при $\gamma < 1$), а величина Q^* представляет собой при этом взвешенную сумму квадратов ошибок вычисления выходной переменной для N наблюдений. Весовая функция имеет вид

$$w(N, j, \gamma) = (1-\gamma)\gamma^{N-j}. \quad (9.81)$$

Выражение (9.75) определяет рекуррентный алгоритм метода наименьших квадратов для случая, когда весовая функция имеет вид (9.81) [12—14]. Для вычисления последовательности оценок коэффициентов вместо выражения (9.75) могут быть использованы и рекуррентные формулы иного вида (см., например, [4]).

Последовательность оценок коэффициентов и оценка $\hat{a}(N+\alpha,\gamma)$, полученные по формулам (9.75) и (9.76), зависят от значения γ. Параметр γ является в данном случае параметром алгоритма решения. Очевидно, что величину этого параметра следует выбрать таким образом, чтобы обеспечить максимальную точность оценки величины выходной переменной в момент времени $N+\alpha$.

Выбор оптимального значения γ^* параметра метода произведем по условию

$$S_\alpha(\gamma^*) = \min_{\gamma} S_\alpha(\gamma), \quad (9.82)$$

где

$$S_\alpha(\gamma) = \frac{1}{N-\alpha+1} \sum_{j=\alpha}^{N} [\tilde{y}(j) - \hat{a}'(j, \gamma) f(j)]^2. \quad (9.83)$$

Как следует из приведенной ниже теоремы, при достаточно больших N величина (9.83) с точностью до множителя, не зависящего от γ, совпадает с математическим ожиданием квадрата ошибки оценивания выходной переменной (9.74). Это значит, что при использовании такого способа выбора параметра γ при достаточно больших N значение γ^* будет совпадать с оптимальным значением с точки зрения критерия (9.74).

Упомянутая выше теорема формулируется следующим образом.

Теорема

Предположим, что

1) для всякого $\delta \leqslant \gamma \leqslant 1$ может быть найдено такое $m(\gamma)$, что разность $r(j) = \tilde{y}(j) - \hat{a}'(j, \gamma) f(j)$ при $j \geqslant m(\gamma)$ есть стационарный эргодический случайный процесс [1];

2) $\alpha > 0$;

3) ошибка $e(i)$ удовлетворяет условиям (9.73). Тогда с вероятностью единица имеет место равенство

$$\lim_{N \to \infty} S_\alpha(\gamma) = \lim_{N \to \infty} E\{[\tilde{y}(N+\alpha) - \hat{a}'(N+\alpha, \gamma) f(N+\alpha)]^2\} + \sigma_e^2. \tag{9.84}$$

Доказательство этой теоремы можно найти в работе [4].

Приведем теперь некоторые пояснения к выбору начальных значений C_0 и $\hat{a}(\alpha, \gamma)$. Условие 1 в формулировке теоремы требует, чтобы при использовании алгоритма (9.75) при любом значении γ после конечного числа итераций $m(\gamma)$ имел место установившийся режим слежения за изменением коэффициентов модели. Поскольку мы используем для получения оценки γ^* конечную выборку наблюдений, то целесообразно стремиться к тому, чтобы время $m(\gamma)$ было по возможности меньшим. Покажем, что достичь этого можно, выбирая начальные значения C_0 и $\hat{a}(\alpha, \gamma)$ в соответствии с (9.77) и (9.78). Нетрудно убедиться [15], что с помощью формулы (9.75) при начальных значениях C_0 и $\hat{a}(\alpha, \gamma)$ оценка $\hat{a}(j+\alpha, \gamma)$ может быть записана следующим образом:

$$\hat{a}(j+\alpha, \gamma) = \left\{\left[\frac{\gamma^j}{1-\gamma} C_0^{-1} + \sum_{i=1}^{j-1} \gamma^{j-i} f(i) f'(i)\right]^{-1} + f(j) f'(j)^{-1}\right\} \times$$

$$\times \left\{\left[\frac{\gamma^j}{1-\gamma} C_0^{-1} \hat{a}(\alpha, \gamma) + \sum_{i=1}^{j-1} \gamma^{j-i} \tilde{y}(i) f(i)\right] + \tilde{y}(j) f(j)\right\}. \tag{9.85}$$

При $j \to \infty$, когда процесс изменения коэффициентов, описываемый уравнением (9.85), установится, математические ожидания членов, заключенных в квадратные скобки, при $\gamma < 1$ будут

---

[1] Условие эргодичности позволяет с помощью одной реализации определить наилучшее значение γ^* для любой реализации процессов $a(i)$ и $f[x(i)]$. Если γ^* определяется для каждой реализации, то необходимость в эргодичности отпадает. В этом случае γ^* есть наилучшая оценка параметра с точки зрения среднего квадрата ошибки для данной реализации.

выражаться следующим образом:

$$E\left\{\lim_{j\to\infty}\left[\frac{\gamma^j}{1-\gamma}C_0^{-1}+\sum_{l=1}^{j-1}\gamma^{j-l}f(l)f'(l)\right]\right\}=\frac{E\{f(j)f'(j)\}}{1-\gamma}\gamma, \quad (9.86)$$

$$E\left\{\lim_{j\to\infty}\left[\frac{\gamma^j}{1-\gamma}C_0^{-1}\hat{a}(\alpha,\gamma)+\sum_{l=1}^{j-1}\gamma^{j-l}\tilde{y}(l)f(l)\right]\right\}=\frac{E\{\tilde{y}(j)f(j)\}}{1-\gamma}\gamma. \quad (9.87)$$

С целью ускорения установления процесса $\hat{a}(j+\alpha,\gamma)$ выберем C_0 и $\hat{a}(\alpha,\gamma)$ такими, чтобы уже при $j=1$ выражения в квадратных скобках (9.85) были бы равны их математическим ожиданиям при $j\to\infty$. Очевидно, что для этого необходимо положить

$$(C_0^{-1})^*=E\{f(j)f'(j)\}, \quad (9.88)$$

$$\hat{a}^*(\alpha,\gamma)=C_0^*E\{\tilde{y}(j)f(j)\}=E\{f(j)f'(j)\}E\{\tilde{y}(j)f(j)\}. \quad (9.89)$$

Оценками матриц $E\{f(j)f'(j)\}$ и $E\{\tilde{y}(j)f(j)\}$ по выборке из N наблюдений являются матрицы $(1/N)F'_NF_N$ и $(1/N)F'_N\tilde{Y}_N$. Отсюда следуют выражения (9.77) и (9.78) для начальных значений C_0 и $\hat{a}(\alpha,\gamma)$.

9.5.3. Описание алгоритма и примеры

Резюмируя сказанное выше, приведем здесь последовательность действий при реализации рассмотренного метода прогнозирования коэффициентов дрейфующего объекта. Итак, пусть модель объекта имеет вид

$$y=\sum_{l=0}^{k}a_l(l)f_l[x(l)].$$

Имеется выборка N наблюдений через равностоящие моменты времени входных и выходной переменных объекта, определяемая матрицами F_N и $\tilde{Y}_N$. Требуется найти оценку $\hat{a}(N+\alpha)$ коэффициентов модели в момент времени $N+\alpha$, где $\alpha=1, 2, \ldots$.

Для решения задачи необходимо проделать следующее:

1) по формулам (9.77) и (9.78) рассчитать начальные значения C_0 и $\hat{a}(\alpha,\gamma)$;

2) для каждого значения γ_r

$$\gamma_r=\delta+rh$$

из интервала $\delta\leqslant\gamma_r\leqslant 1$, где δ—малая положительная величина, h—заданный шаг, а r принимает значения 0, 1, 2, ..., рассчитать с помощью формул (9.75) и (9.76) последовательность $\hat{a}(1+\alpha,\gamma_r)$, $\hat{a}(2+\alpha,\gamma_r)$, ..., $\hat{a}(N+\alpha,\gamma_r)$. По результатам расчетов для каждого γ_r найти величину $S_\alpha(\gamma_r)$ по формуле (9.83);

3) определить значение γ^*, соответствующее минимальной величине $S_\alpha(\gamma)$. При этом следует иметь в виду что функция $S_\alpha(\gamma)$ может иметь несколько экстремальных точек;

4) провести вычисления по формулам (9.75) и (9.76) для $j=0, 1, \ldots, N-1$ и $\gamma=\gamma^*$. В результате получаем искомую оценку $\hat{a}(N+\alpha, \gamma^*)=\hat{a}(N+\alpha)$.

Пример 9.4

Пусть модель объекта имеет следующий вид:

$$y(t)=a_1(t)x_1(t)+a_2(t)x_2(t).$$

Задана выборка, содержащая $N=10$ наблюдений (табл. 9.5). Требуется найти оценки коэффициентов модели в момент времени $N+1$.

1. Пользуясь формулами (9.77) и (9.78), рассчитываем величины $\hat{a}_1(\alpha, \gamma)$, $\hat{a}_2(\alpha, \gamma)$ и C_0. Матрица F'_NF_N в данном случае имеет вид

$$F'_NF_N=\begin{bmatrix} \sum_{t=1}^{10} x_1^2(t) & \sum_{t=1}^{10} x_1(t)x_2(t) \\ \sum_{t=1}^{10} x_1(t)x_2(t) & \sum_{t=1}^{10} x_2^2(t) \end{bmatrix},$$

а обратная матрица

$$(F'_NF_N)^{-1}=\begin{bmatrix} 0{,}06317 & -0{,}01524 \\ -0{,}01524 & 0{,}0816 \end{bmatrix}.$$

По формуле (9.77) находим

$$C_0=\begin{bmatrix} 0{,}6317 & -0{,}1524 \\ -0{,}1524 & 0{,}816 \end{bmatrix}.$$

Таблица 9.5

t	$x_1(t)$	$x_2(x)$	$\tilde{y}(t)$
1	−0,74	0,662	−0,2598
2	1,667	1,667	0,3517
3	−1,83	−1,667	−0,3925
4	−0,778	−0,74	0,0512
5	1,945	−1,83	0,4703
6	−1,667	−0,74	0,5950
7	0,623	−0,778	−0,131
8	1,167	0,623	−0,5451
9	0,662	0,1945	−0,6581
10	−0,74	1,167	0,3616

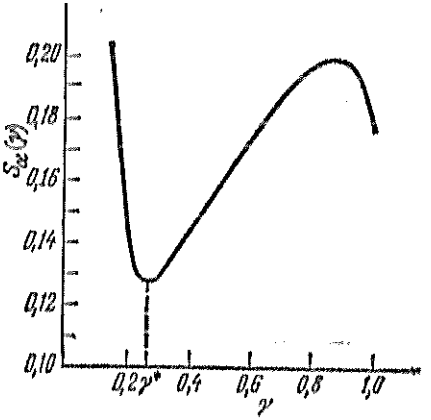
Матрица $F'_N\tilde{Y}_N$ имеет вид

$$F'_N\tilde{Y}_N = \begin{bmatrix} \sum_{t=1}^{10} x_1(t)\,\tilde{y}(t) \\ \sum_{t=1}^{10} x_2(t)\,\tilde{y}(t) \end{bmatrix} = \begin{bmatrix} -0{,}0411 \\ -0{,}2139 \end{bmatrix}.$$

Из уравнения (9.78) следует

$$\hat{a}_1(\alpha, \gamma) = 0{,}06317\,(-0{,}0411) + (0{,}01524)\,(-0{,}2139) = 0{,}0007,$$
$$\hat{a}_2(\alpha, \gamma) = (-0{,}01524)(-0{,}0411) + (0{,}0816)(-0{,}2139) = -0{,}0168.$$

2. По формулам (9.75) и (9.76) рассчитываем для $\alpha = 1$ последовательность оценок коэффициентов для значений γ из ин-



Фиг. 9.1. Зависимость $S_\alpha(\gamma)$ для примера 9.4.

тервала $0{,}1 \leqslant \gamma \leqslant 1$ с шагом $h = 0{,}1$. В диапазоне значений γ от 0,2 до 0,3 проводим вычисления с шагом 0,02. На основе результатов вычислений по формуле (9.83) находим величины $S_\alpha(\gamma)$. Полученная зависимость $S_\alpha(\gamma)$ показана на фиг. 9.1.

В качестве примера проведем вычисление последовательности оценок коэффициентов и значений $S_\alpha(\gamma)$ для $\gamma = 0{,}26$. Используя полученные выше значения $\hat{a}_1(\alpha, \gamma)$, $\hat{a}_2(\alpha, \gamma)$ и C_0, находим

$$\hat{a}_1(\alpha + 1;\ 0{,}26) = 0{,}0007 + \frac{[0{,}6317\times(-0{,}74) + (-0{,}1524)\times 0{,}662][-0{,}2598 - (-0{,}74\times 0{,}0007 - 0{,}0168\times 0{,}662)]}{(0{,}74)^2\times 0{,}6317 + (0{,}662)^2\times 0{,}816 + 2(-0{,}1524)\times(-0{,}74)\times 0{,}662 + \frac{0{,}26}{1-0{,}26}} =$$
$$= 0{,}1435,$$

$$\hat{a}_2(\alpha+1;\ 0{,}26) = -0{,}0168 + \frac{[(-0{,}1524)\times(-0{,}74)+(0{,}816\times 0{,}662)][-0{,}2598-(0{,}74\times 0{,}0007-0{,}0168\times 0{,}662)]}{(0{,}74)^2\times 0{,}6317+(0{,}662)^2\times 0{,}816+2\,(-0{,}1524)\times(-0{,}74)\times 0{,}662+\frac{0{,}26}{1-0{,}26}} = -0{,}1244.$$

Матрицу C_1 вычисляем по формуле (9.76):

$$C_1 = \begin{pmatrix} 2{,}08 & -1{,}22 \\ -1{,}22 & 2{,}04 \end{pmatrix}.$$

Аналогичным образом рассчитываем оценки $\hat{a}_1(\alpha+2;\ 0{,}26)$, $\hat{a}_2(\alpha+2;\ 0{,}26)$, C_2 и т. д. Результаты расчетов приведены в табл. 9.6.

Величину $S_\alpha(0{,}26)$ рассчитываем по формуле (9.83):

$$S_\alpha(0{,}26) = \frac{1}{10}\Big\{\Big[-0{,}2598-(-0{,}0007\times 0{,}74-0{,}0168\times 0{,}662)\Big]^2+[0{,}3517-(0{,}1435\times 1{,}667-0{,}1244\times 1{,}667)]^2+\ldots+[0{,}3616-(0{,}6489\times 0{,}74-0{,}1513\times 1{,}167)]^2\Big\} = 0{,}1276.$$

3. Из фиг. 9.1 видно, что минимум функции $S_\alpha(\gamma)$ имеет место при $\gamma = \gamma^* = 0{,}26$ и $S_\alpha(\gamma^*) = 0{,}1276$.

4. По формулам (9.75) и (9.76) рассчитываем для $\gamma^* = 0{,}26$ последовательность оценок коэффициентов для $j = 1, \ldots, N$ (см. табл. 9.6). Для момента времени $N+1$ получаем следующий результат:

$$\hat{a}_1(N+1,\ \gamma^*) = \hat{a}_1(N+1) = -0{,}664,$$
$$\hat{a}_2(N+1,\ \gamma^*) = \hat{a}_2(N+1) = -0{,}112.$$

Отметим, что величина $NS_\alpha(\gamma)$ при $\gamma = 1$ соответствует остаточной сумме квадратов, которая получается в случае, когда

Таблица 9.6

j	$\hat{a}_1(\alpha+j;\ 0{,}26)$	$\hat{a}_2(\alpha+j;\ 0{,}26)$
0	0,0007	—0,0168
1	0,1435	—0,1244
2	0,2374	—0,03197
3	0,2418	—0,03133
4	0,2385	—0,1362
5	0,1763	—0,06977
6	—0,1498	—0,3892
7	—0,2392	—0,158
8	—0,3237	—0,237
9	—0,6489	—0,1513
10	—0,6644	—0,1121

модель строится без учета дрейфа. В данном примере $NS_{\alpha}(1) = 1{,}8$ и, таким образом, применение изложенного здесь метода привело к уменьшению дисперсии оценки выходной переменной приблизительно на 30%.

Пример 9.5

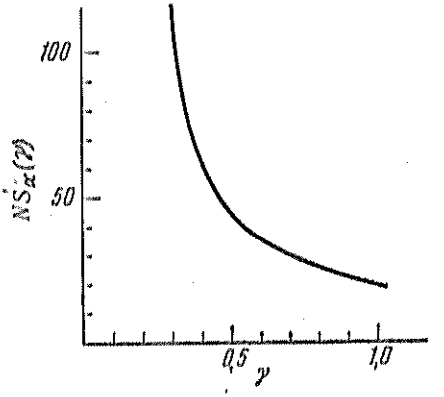
Исследовался процесс формирования и сушки ленты синтетического каучука. Формирование ленты осуществлялось с помощью лентоотливочной машины. Исходным продуктом для получения каучука является латекс. В аппаратах, установленных перед лентоотливочной машиной, латекс коагулировал и смешивался с масляной эмульсией. После лентоотливочной машины лента каучука поступала в сушилку. На выходе из сушилки лента должна содержать не выше определенного процента влаги. Увеличение производительности лентоотливочной машины требует более интенсивной подачи латекса. Однако при этом (при условии, что скорость лентоотливочной машины и режим работы сушилки не изменяются) возрастает толщина ленты и ухудшаются условия ее сушки. Поэтому для управления процессом важно знать допустимую толщину, при которой лента в тех или иных условиях может быть высушена до требуемых пределов. Было выяснено, что на скорость сушки оказывают влияние такие параметры, как дефо латекса (x_1), содержание масла в ленте каучука (x_2), pH в аппаратах коагуляции (x_3). Кроме того, скорость сушки существенным образом зависит от неконтролируемого состояния сита лентоотливочной машины, которое по мере работы забивается частицами латекса.

Была поставлена задача определить зависимость между допустимой толщиной (y) ленты и перечисленными параметрами x_1, x_2 и x_3 при условии нормального высыхания ленты. Для этой цели была получена выборка из 38 наблюдений, в которую были включены лишь те режимы, при которых производительность установки была близка к оптимальной. Точки выборки отстояли одна от другой приблизительно на одинаковые интервалы времени. Зависимость между толщиной ленты и перечисленными выше факторами x_1, x_2 и x_3 искалась в виде линейного уравнения. Предполагалось, что за время сбора данных состояние сита лентоотливочной машины не изменяется, т. е. что собранные данные соответствуют случаю отсутствия дрейфа. По результатам экспериментов методом регрессионного анализа была построена математическая модель в виде

$$y = 0{,}362x_1 + 0{,}35x_2 - 0{,}465x_3. \qquad (9.90)$$

В уравнении (9.90) все переменные представлены в нормированном виде. Остаточная сумма квадратов оказалась равной $S_R = 19{,}1$.

С помощью изложенного выше метода можно проверить правильность предположения об отсутствии дрейфа при проведении эксперимента. Для этой цели необходимо найти значение γ^*, минимизирующее $S_\alpha(\gamma)$, например, при $\alpha = 1$. Случаю отсутствия дрейфа будет соответствовать $\gamma^* = 1$. После проведения необхо-



Фиг. 9.2. Зависимость $NS_\alpha(\gamma)$ для примера 9.5.

димых вычислений была получена зависимость $S_\alpha(\gamma)$, приведенная на фиг. 9.2. Как видно из графика кривой, функция $S_\alpha(\gamma)$ монотонно убывает при увеличении γ и достигает минимального значения при $\gamma^* = 1$. Этот результат подтверждает предположение об отсутствии дрейфа на интервале сбора экспериментальных данных.

ЛИТЕРАТУРА

1. Большаков И. А., Гуткин Л. С., Левин Б. Р., Стратанович Р. Л., Математические основы современной радиоэлектроники, вып. 2, изд-во „Советское радио", М., 1968.
2. Фельдбаум А. А., Основы теории оптимальных автоматических систем, изд-во „Наука", М., 1966.
3. Маслов Е. П., Оценка параметров марковских объектов, *Автоматика и телемеханика*, № 1 (1964).
4. Бородюк В. П., Лецкий Э. К., Статистическое описание промышленных объектов, изд-во „Энергия", М., 1971.
5. Бородюк В. П., Проверка однородности статистических данных в регрессионном анализе, см. сб., „Проблемы планирования эксперимента", изд-во „Наука", М., 1969.
6. Box G. E. P., Multifactor designs of first oder, *Biometrica*, 39, № 1, 49 (1952).
7. Горский В. Г., Бродский В. З., О симплекс-планах первого порядка и связанных с ними планах второго порядка, см. сб. „Новые идеи в планировании эксперимента", изд-во „Наука", М., 1969.
8. Большев Л. Н., Смирнов Н. В., Таблицы математической статистики, изд-во „Наука", М., 1965.
9. Experimental Designs in Industry (ed. V. Chew), Wiley, New York; Chapman a. Hall, London, 1958.

10. Налимов В. В., Чернова Н. А., Статистические методы планирования экстремальных экспериментов, изд-во „Наука", М., 1965.
11. Круг Г. К., Лисенков А. Н., Планирование эксперимента в условиях непрерывного временного дрейфа, Труды МЭИ, вып. 67, М., 1966.
12. Calman R. E., Design of self-optimizing control System, *Trans. ASME* (Febr. 1958).
13. Перельман И. И., Текущий регрессионный анализ и его применение в некоторых задачах автоматического управления, *Изв. АН СССР*, ОТН, Энергетика и автоматика, № 2 (1960).
14. Lee R.C.K., Optimal estimation, identification and control, Res. monogr. No. 28, M.I.T. Press, Cambridge, Mass.; русский перевод: Ли Р., Оптимальные оценки, определение характеристик и управление, изд-во „Наука", М., 1966.
15. Gyürki I., Über Algorithmen zur Ermittlung von linearen Modellen, Informationsartikel C_1, IWK, S. 139.
16. Wernstedt J., Peter U., Zur Parameterschätzung statischer Systeme mittels der rekursiven Regression, *MSR*, № 3 (1971).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Ланцош К., Практические методы прикладного анализа, Физматгиз, М., 1961.
2. Маркова Е. В., Лисенков А. Н., Планирование эксперимента в условиях неоднородностей, изд-во „Наука", М., 1973.
3. Соколов Б. М., Учет временного дрейфа при планировании эксперимента, *Заводск. лаб.*, № 5 (1975).
4. Лисенков А. Н., *Заводск. лаб.*, № 1 (1975).
5. Bodkin R. G., A note on the standard error of prediction in the multivariate case, *Metroeconomica*, 22, № 2, 149—164 (1970).
6. Burnett J. D., Guthrie D., Estimation of stationary stochastic regression parameters, *J. Am. Statist. Ass.*, 65, № 332 (1970).

10.

Планирование эксперимента и оценивание параметров нелинейных моделей

10.1. Введение и основные понятия

В этом разделе рассматривается планирование эксперимента и оценивание параметров нелинейных моделей произвольного вида. Модель будем записывать в следующей форме:

$$y^t = f(x_1^t, \ldots, x_m^t, a_1, \ldots, a_p). \qquad (10.1)$$

Здесь x_j^t—m независимых переменных ($j = 1, m$; $t = 1, N$)[1]; y^t—значение зависимой переменной в точке x_j^t ($t = 1, N$); a_l ($l = 1, p$)—неизвестные параметры.

От модели (10.1) требуется лишь, чтобы в пределах некоторых областей в пространствах x и a она позволяла бы рассчитать значения y и чтобы возможно было по крайней мере численное дифференцирование y по x и a. Для того чтобы исключить простейшие частные случаи, мы предполагаем, что f нелинейна по крайней мере по a. Относительно x функция f может быть как линейной, так и нелинейной.

В практических задачах значения переменных, входящих в уравнение (10.1), не известны. Тем не менее с помощью метода наименьших квадратов удается найти соответствующие оценки $\hat{y}$, $\hat{x}$ и $\hat{a}$, для которых имеет место

$$\hat{y}^t = f(\hat{x}_1^t, \ldots, \hat{x}_m^t, \hat{a}_1, \ldots, \hat{a}_p), \qquad (10.2)$$

где $\hat{y}$, $\hat{x}$ и $\hat{a}$—расчетные значения, соответствующие истинным значениям $\bar{y}$, $\bar{x}$ и $\bar{a}$. Предполагая, что в эксперименте не возникает систематических ошибок, из значений $\hat{y}$, $\hat{x}$ и $\hat{a}$, оцененных по результатам N наблюдений, можно получить значения $\bar{y}$, $\bar{x}$ и $\bar{a}$, если N стремится к бесконечности, т. е. если в оценивании участвует сколь угодно большое число наблюдений. Это утверждение будет справедливо в том случае, если уравнение (10.1) адекватно отражает взаимосвязь между параметрами, независимыми и зависимыми переменными. Такая предпосылка не является ни в коей мере тривиальной и на практике часто нарушается. Однако имеются методы, позволяющие проверить адекватность модели.

---

[1] Запись $t = 1, N$ означает, что индекс t принимает целые значения в диапазоне от 1 до N.—*Прим. ред.*

Под оценками параметров, наилучшими в смысле метода наименьших квадратов, понимаются значения параметров $\hat{a}$, полученные путем минимизации по a суммы квадратов

$$S = \sum_{i=1}^{N} \left[\frac{1}{\sigma_{y^i}^2} (\tilde{y}^i - f(\tilde{x}^i, a))^2 + \sum_{j=1}^{m} \frac{1}{\sigma_{x_j^i}^2} (\tilde{x}_j^i - x_j^i)^2 \right]. \qquad (10.3)$$

В этой формуле $\tilde{y}^i$ и $\tilde{x}_j^i$ — результаты измерения y^i и x_j^i, а σ_{y^i} и $\sigma_{x_j^i}$ — соответствующие им среднеквадратические ошибки.

Планирование эксперимента, о котором в дальнейшем будет идти речь, непосредственно связано с методом наименьших квадратов, т. е. с минимизацией выражения (10.3).

Следует, однако, подчеркнуть, что параметры заданной модели (10.1) можно определить не только путем минимизации (10.3). Для хорошего согласования измеренных значений с рассчитанными вполне могут быть применены также и другие критерии. С точки зрения теории невозможно предпочесть один критерий другому. Тем не менее применяют почти исключительно метод наименьших квадратов, и одна из причин этого заключается в том, что на практике с помощью этого метода были получены хорошие результаты.

И все же в некоторых случаях оказывается полезным, или по крайней мере возможным, вместо критерия (10.3) минимизировать наибольшее отклонение измеренных значений от расчетных (чебышевская аппроксимация). В качестве примера применения этого метода мы рассмотрим ниже задачу балансировки.

После указания на математическую основу задач нелинейного оценивания и прежде чем перейти к более подробному изложению методов решения, сделаем несколько замечаний, касающихся практических аспектов планирования эксперимента в таких задачах. Простой пример, связанный с химической кинетикой, позволяет проиллюстрировать, что́ представляет собой планирование эксперимента для нелинейных моделей и на что следует обратить внимание.

Вещество A распадается с образованием вещества B, которое в свою очередь переходит в C:

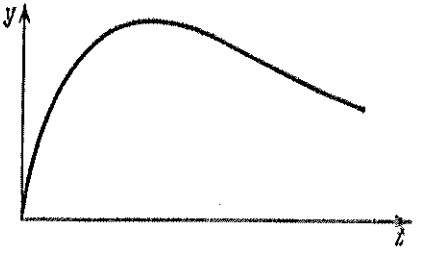
$$A \xrightarrow{a_1} B \xrightarrow{a_2} C.$$

Обе реакции являются необратимыми и имеют первый порядок. Тогда для концентрации вещества B справедливо выражение

$$y = \frac{a_1}{a_1 - a_2} (\exp(-a_2 t) - \exp(-a_1 t)),$$

где t — время реакции; y — концентрация вещества B; a_1 и a_2 — константы скорости реакции.

Изменение y во времени показано на фиг. 10.1. Задача оптимального планирования эксперимента состоит здесь в следующем: пусть измерения могут быть проведены в двух точках временной оси и по их результатам предполагается найти оценки параметров a_1 и a_2. Необходимо так выбрать точки измерения, чтобы обеспечить наибольшую точность оценок a_1 и a_2. Что понимается под точностью оценок, будет определено в дальнейшем. Здесь



Фиг. 10.1. Изменение концентрации вещества B.

мы пока ограничимся употреблением этого неопределенного понятия.

Результаты измерений содержат ошибки. Достаточно очевидно и легко проверяется путем расчетов, что, например, измерения в моменты времени $t = 6$ и $t = 7$ не позволяют получить надежной информации о значениях параметров a_1 и a_2. То же самое касается и пары значений $t = 1$ и $t = 2$. Поэтому задача оптимального планирования эксперимента заключается в таком выборе экспериментальных точек при заданной дисперсии ошибок наблюдений, при котором оценки параметров можно было бы определить с возможно большей точностью. В некоторых случаях заранее задается число экспериментов. Тогда соответствующим выбором расположения экспериментальных точек должна быть максимизирована статистическая точность оценок параметров. В других случаях статистическая точность оценок параметров задается до проведения эксперимента и должны быть определены число и расположение точек измерения.

Можно показать, что оптимальное планирование эксперимента в общем случае представляет собой нелинейную экстремальную задачу. В нашем примере при заданном числе экспериментальных точек должны быть определены моменты времени измерения, т. е. значения непрерывной переменной. В других случаях число измерений заранее не задается, так как не всегда можно утверждать, что при некотором заранее заданном числе измерений удается достигнуть требуемой статистической точности оценок параметров. Наконец, во всех практических задачах на возможное расположение экспериментальных точек наложены

ограничения; например, область планирования задается условиями взрывобезопасности для газовых смесей, температурными ограничениями для катализатора и т. д. Таким образом, приходится иметь дело с нелинейной экстремальной задачей, включающей непрерывные и дискретные переменные, которые к тому же могут быть ограничены. Каждый, кто хотя бы поверхностно знаком с методом решения экстремальных задач, знает, насколько большие затруднения может вызвать решение такой общей задачи. Поэтому планирование эксперимента (так же как и оценивание в нелинейных задачах) необходимо рассматривать не как оторванную от практики теоретическую задачу, а нужно подходить к нему с экономической точки зрения. В конечном итоге речь идет о минимизации затрат. Наиболее существенными составляющими затрат являются собственно затраты на эксперимент и затраты на обработку эксперимента и на его планирование (в обоих случаях важную роль играют затраты вычислительного времени). Должны быть минимизированы общие затраты на достижение определенной цели, например на обеспечение наперед заданной статистической точности оценки параметров.

Если затраты на проведение эксперимента являются незначительными, то во многих случаях можно отказаться от численной оптимизации при планировании эксперимента. Поскольку в этом случае затраты на вычисления существенно превышают затраты на проведение эксперимента, предпочтительнее провести на несколько экспериментов больше, а при выборе экспериментальных точек опираться на опыт и интуицию исследователя.

Вообще можно сказать, что во многих случаях планирование эксперимента по описанной ниже методике не вызывается необходимостью или же связано со слишком большими затратами, так как достаточно хорошие результаты получаются уже при учете некоторых общих соображений (см. разд. 10.3.3).

Принципиальная трудность планирования эксперимента для нелинейных моделей состоит в том, что определение оптимального плана предполагает знание истинных значений параметров (разд. 10.3.2). Это выглядит противоречием, но на самом деле является вполне естественным.

Для того чтобы эффективно планировать эксперимент, необходимо хорошо знать поведение объекта исследования в области планирования. Чем основательнее это знание, тем лучше удается планирование. Это справедливо для любых задач планирования, в том числе и для задач планирования эксперимента.

На практике с исследуемым объектом связываются определенные физико-технические представления. Одновременно с этим часто оказываются известными начальные значения для искомых параметров. Обычно вначале проводится и обрабатывается неко-

торое число опытов, выбираемых без трудоемких процедур планирования, с целью получения исходных оценок параметров, знание которых необходимо для эффективного планирования эксперимента. Кроме того, планирование эксперимента может быть повторено несколько раз для различных начальных значений параметров модели.

Мы выделим две группы методов планирования эксперимента: прямые и непрямые методы. Прямые методы представляют собой вычислительный процесс, с помощью которого при заданном числе экспериментов и известной оценке дисперсии ошибок наблюдений непосредственно вычисляют наилучшие в определенном смысле координаты точек измерений. В ранее упомянутом примере из химической кинетики с помощью прямых методов могут быть вычислены моменты времени, в которые должны производиться измерения. Следует подчеркнуть, что прямые методы, как правило, требуют значительного объема вычислений. Поэтому часто пользуются непрямыми методами.

С помощью непрямых методов планирования эксперимента не получают сами точки измерений, а на основе определенного критерия лишь проводится анализ ряда наперед заданных точек с целью выбора лучших из них для проведения измерений.

Применение непрямых методов требует активного участия исследователя в построении плана эксперимента. При этом исследователь может успешно использовать свой опыт, а иногда и свою интуицию. Учет наложенных на отдельные переменные ограничений оказывается здесь значительно менее трудоемким, чем в прямом планировании.

Применяя оба типа методов планирования эксперимента, необходимо помнить, что получаемые результаты будут справедливы лишь в случае, если строго выполняются все условия, которые часто неявным образом вводятся при планировании. Поэтому результаты планирования надо рассматривать лишь как рекомендации к проведению исследования. Вывод об эффективности планирования эксперимента можно получить лишь после проведения запланированных экспериментов и их статистической обработки.

Сделаем несколько замечаний относительно содержания и взаимной связи нижеследующих разделов. В разд. 10.2 поясняется метод наименьших квадратов, включая соответствующий статистический анализ. Однако теория здесь детально не излагается, а лишь перечисляются наиболее важные для наших целей формулы. Для более основательного знакомства с этим вопросом можно обратиться к отличной и доступно написанной книге Вольберга [1], чтение которой доставило автору большое удовольствие. В изложенном материале будут использоваться обозначения Вольберга, чтобы облегчить заинтересованным

читателям понимание опущенных здесь промежуточных выкладок.

Затем приводится программа на ФОРТРАНе, которая позволяет произвести статистический анализ результатов эксперимента и может быть использована при непрямом планировании.

Разд. 10.3 посвящен различным методам планирования эксперимента. Эти методы применяются здесь для решения простых примеров, позволяющих выяснить их достоинства и недостатки.

В разд. 10.4 обсуждаются вопросы обработки результатов экспериментов. Эта область настолько обширна, что можно представить лишь некоторые общие ее положения. Обработка экспериментальных данных проиллюстрирована на примере одной сложной модели. Одновременно описываются способы проверки адекватности заданной модели и исследуется влияние систематических ошибок. Наконец, на примере задачи балансировки показано, что обработка результатов эксперимента может производиться не только методом Гаусса, но и другими методами. Коротко обсуждаются преимущества и недостатки чебышевской аппроксимации.

10.2. Метод наименьших квадратов

10.2.1. Теоретическое введение

Метод наименьших квадратов основан на минимизации суммы взвешенных квадратов разностей между экспериментальными и расчетными значениями:

$$\min_{a,\, x_j^l} S = \min_{a,\, x_j^l} \sum_{l=1}^{N} \left[\frac{1}{\sigma_{y^l}^2} (\tilde{y}^l - f(\tilde{x}^l,\, a))^2 + \sum_{j=1}^{m} \frac{1}{\sigma_{x_j^l}^2} (\tilde{x}_j^l - x_j^l)^2 \right].$$

В общем случае задача нахождения оценок параметров a_l является нелинейной, поэтому расчет оценок производится по итерационной процедуре, начиная с некоторых заданных значений оценок $\hat{a}_l$. Существует несколько различных способов организации итераций. Один из них приведен в книге [1]. Обозначая

$$\Lambda_k = \hat{a}_k - a_k \quad (k = 1,\, p), \tag{10.4}$$

получаем основную формулу для итераций:

$$\Lambda_k = \sum_{j=1}^{p} C_{kj} V_j. \tag{10.5}$$

Здесь введены следующие обозначения:

$$C_{kj} = \sum_{l=1}^{N} F^l_{a_k} F^l_{a_j} / L_l,$$

$$V_j = \sum_{l=1}^{N} F^l_{a_j} F^l_0 / L_l,$$

$$L_l = \sigma^2_{y^l} + \sum_{j=1}^{m} (F^l_{x_j} \sigma_{x_j^l})^2,$$

$$F^l_{x_j} = \left.\frac{\partial F}{\partial x_j}\right|_{\substack{x=x^l \\ a=\hat{a}}},$$

$$F^l_{a_k} = \left.\frac{\partial F}{\partial a_k}\right|_{\substack{x=x^l \\ a=\hat{a}}},$$

$$F = y - f(x_1, \ldots, x_m, a_1, \ldots, a_p),$$

$$F^l_0 = \tilde{y}^l - f(\tilde{x}^l_1, \ldots, \tilde{x}^l_m, \hat{a}_1, \ldots, \hat{a}_p).$$

Итерационная формула (10.5) получена путем разложения функции $f(x, a)$ в ряд Тейлора по a с отбрасыванием всех членов, кроме первого, содержащего A_k. Это приближение для $f(x, a)$ подставляется в S. Приравнивая нулю первые частные производные от S по A_k, получаем линейную систему уравнений (10.5).

В точке экстремума справедливо равенство

$$A_k = 0$$

или

$$\sum_{k=1}^{p} C^{-1}_{lk} \sum_{l=1}^{N} L_l^{-1} F^l_{a_k} F^l_0 = 0, \tag{10.6}$$

где F^l_0 вычисляются уже при значениях $\hat{a}^*_1, \hat{a}^*_2, \ldots, \hat{a}^*_p$ параметров, полученных в результате итераций:

$$F^l_0 = \tilde{y}^l - f(\tilde{x}^l_1, \ldots, \tilde{x}^l_m, \hat{a}^*_1, \ldots, \hat{a}^*_p). \tag{10.7}$$

Практическое применение итерационной процедуры (10.5) показывает, что во многих случаях она не сходится. Поэтому применим другой, часто более эффективный метод, предложенный Марквардтом [2]. Уравнение (10.6) в точке экстремума остается справедливым также и для этого метода.

Для полноты изложения коротко охарактеризуем метод Марквардта. Он представляет собой метод минимизации суммы квадратов относительно некоторого параметра:

$$\min_a \sum_{l=1}^{N} [h_l(a)]^2.$$

В нашей задаче упрощенно полагаем

$$h_l(a) = \tilde{y}^l - f(\tilde{x}^l, a).$$

Разложим теперь $f(x, a)$ по a в ряд Тейлора, который ограничим членами первого порядка:

$$f(\tilde{x}^l, a) = f(\tilde{x}^l, \hat{a}) + \sum_{j=1}^{p} \left.\frac{\partial f(\tilde{x}^l, a)}{\partial a_j}\right|_{a=\hat{a}} v_j.$$

Используя это приближение для $f(\tilde{x}^l, a)$, получаем

$$\min_{v} \sum_{l=1}^{N} \left(\tilde{y}^l - f(\tilde{x}^l, \hat{a}) - \sum_{j=1}^{p} \left.\frac{\partial f(\tilde{x}^l, a)}{\partial a_j}\right|_{a=\hat{a}} v_j \right)^2.$$

Для минимизации этой суммы квадратов по v приравниваем нулю первые частные производные по v_k и получаем линейную систему уравнений

$$A\hat{v} = b,$$

в которой использованы обозначения

$$A_{lj} = \sum_{k=1}^{N} \left.\frac{\partial f(\tilde{x}^k, a)}{\partial a_l}\right|_{a=\hat{a}} \left.\frac{\partial f(\tilde{x}^k, a)}{\partial a_j}\right|_{a=\hat{a}},$$

$$b_j = -\sum_{l=1}^{N} \left.\frac{\partial f(\tilde{x}^l, a)}{\partial a_j}\right|_{a=\hat{a}} (\tilde{y}^l - f(\tilde{x}^l, \hat{a})).$$

Сущность метода Марквардта состоит в том, чтобы не решать непосредственно линейную систему уравнений относительно $\hat{v}$, а модифицировать матрицу A добавлением величины λ в главную диагональ:

$$A_{lj} \longrightarrow A_{lj} + \lambda\delta_{lj}.$$

Разумеется, предварительно следует соответствующим образом пронормировать A и b. На каждой итерации процедуры система уравнений решается для различных λ. Сравниваются соответствующие значения целевой функции и выбирается наилучшее из них. Укажем здесь без доказательства, что различным выбором λ можно получить интерполяцию между методом Ньютона и обычным градиентным методом. При больших λ получаем шаг $\hat{v}$ малой величины приблизительно в направлении градиента. Введение λ позволяет решить многие проблемы, которые не могут быть решены другими способами, например исходя из равенства (10.5).

Рассмотрим теперь, насколько точными получаются оценки параметров, найденные с помощью метода наименьших квадра-

тов, или, говоря точнее, какова их статистическая достоверность.

Измеренные и неизвестные истинные значения отличаются друг от друга на величину ошибок e_{y^l}, $e_{x_j^l}$:

$$\begin{aligned} \tilde{y}^l &= \bar{y}^l + e_{y^l}, \\ \tilde{x}_j^l &= \bar{x}_j^l + e_{x_j^l}. \end{aligned} \tag{10.8}$$

Требуется оценить точность σ_{a_k} оценок $\hat{a}_k$ параметров a_k. Величина σ_{a_k} определяется следующим образом:

$$\sigma_{a_k} = (E[(\hat{a}_k - \bar{a}_k)^2])^{1/2} \quad (k = 1, p), \tag{10.9}$$

где E — математическое ожидание по множеству наблюдений.

Наша цель состоит в том, чтобы определить взаимосвязь между σ_{a_k} и ошибками e_{y^l} и $e_{x_j^l}$. Для истинных значений $\bar{y}$, $\bar{x}$ и $\bar{a}$ справедливо

$$\bar{y}^l - f(\bar{x}_1^l, \ldots, \bar{x}_m^l, \bar{a}_1, \ldots, \bar{a}_p) = 0. \tag{10.10}$$

Вычитая уравнение (10.10) из (10.7) и подставляя равенства (10.8), получаем после отбрасывания членов порядка выше первого в разложении разности $f(x, a) - f(\bar{x}, \bar{a})$ в ряд Тейлора

$$F_0^l = e_{y^l} - \sum_{j=1}^{m} \frac{\partial f(x, \hat{a})}{\partial x_j}\bigg|_{x=x^l} e_{x_j^l} - \sum_{k=1}^{p} \frac{\partial f(x^l, a)}{\partial a_k}\bigg|_{a=\hat{a}} (\hat{a}_k - \bar{a}_k). \tag{10.11}$$

Здесь линеаризация производилась как по отклонениям наиболее вероятных значений параметров от истинных, т. е. по $(\hat{a}_k - \bar{a}_k)$, так и по отклонениям наиболее вероятных значений зависимой переменной от ее истинных значений, т. е. по $\tilde{y}^l - \bar{y}^l = e_{y^l}$. Если ввести предположение о некоррелированности ошибок в различных точках измерений, которое оказывается в большинстве случаев справедливым, получаем с помощью чисто алгебраических преобразований два соотношения

$$E[(\hat{a}_l - \bar{a}_l)^2] = E[(L^{-1}e^2)]\, C_{ll}^{-1} \tag{10.12}$$

и

$$\rho_{lk}\sigma_{a_l}\sigma_{a_k} = E[(\hat{a}_l - \bar{a}_l)(\hat{a}_k - \bar{a}_k)] = E[(L^{-1}e^2)]\, C_{lk}^{-1}. \tag{10.13}$$

Здесь $E[(L^{-1}e^2)]$ — математическое ожидание величины

$$L_l^{-1}\left(e_{y^l} + \sum_{j=1}^{m} F_{x_j}^l e_{x_j^l}\right)^2$$

для всех l.

Назовем ковариацией оценок параметров $\hat{a}_l$ и $\hat{a}_k$ произведение $\rho_{lk}\sigma_{a_l}\sigma_{a_k}$, в котором ρ_{lk}—коэффициент корреляции между $\hat{a}_l$ и $\hat{a}_k$.

Для величины $E[(L^{-1}e^2)]$ приближенно можно записать

$$E[(L^{-1}e^2)] \approx \frac{S}{N-p},$$

где S—взвешенная сумма квадратов отклонений в точке экстремума, а $(N-p)$—разность числа наблюдений и числа параметров. Для S имеем выражение

$$S = \sum_{l=1}^{N} \frac{1}{L_l} (F_0^l)^2.$$

Отсюда получаем формулы для расчета дисперсии и ковариации оценок параметров:

$$\sigma^2_{a_k} \approx \frac{S}{N-p} C^{-1}_{kk}, \tag{10.14}$$

$$\rho_{lk}\sigma_{a_k}\sigma_{a_l} \approx \frac{S}{N-p} C^{-1}_{lk} \quad (l \neq k). \tag{10.15}$$

Равенства (10.14) и (10.15) представляют собой искомую зависимость статистической достоверности параметров от ошибок измерений. На этих уравнениях основывается как статистический анализ экспериментальных данных, так и оптимальное планирование эксперимента. Конечно, точные значения σ_{a_k} не могут быть рассчитаны по небольшому числу экспериментов. Однако по этим данным можно получить их оценки. Обозначая эти оценки дисперсий через $s^2_{a_k}$, получаем аналогично (10.14) и (10.15) уравнения

$$s^2_{a_k} = \frac{S}{N-p} C^{-1}_{kk}, \tag{10.16}$$

$$s_{lk} = r_{lk} s_{a_l} s_{a_k} = \frac{S}{N-p} C^{-1}_{lk}. \tag{10.17}$$

Используя оценки дисперсий, можно найти доверительные интервалы для отдельных параметров. При этом используется t-распределение [1]. Значения $t_{(\alpha/2,\, N-p)}$ приведены в табл. 15.3. Здесь $(N-p)$—число степеней свободы, α—вероятность события, состоящего в том, что в 100α из 100 случаев параметры находятся вне доверительного интервала. Справедливы нера-

[1] Использование здесь t-распределения предполагает нормальность распределения ошибок наблюдений. Значение $t_{(\alpha/2,\, N-p)}$ выше обозначалось через $t_{кр}$ при уровне значимости α и числе степеней свободы $\varphi = N-p$.—*Прим. ред.*

вества

$$\hat{a}_k - s_{a_k} t_{(\alpha/2,\ N-p)} < \bar{a}_k < \hat{a}_k + s_{a_k} t_{(\alpha/2,\ N-p)}. \qquad (10.18)$$

При $(N-p)$ степенях свободы истинные (неизвестные) значения параметров $\bar{a}_k$ находятся между $\hat{a}_k - s_{a_k} t_{(\alpha/2,\ N-p)}$ и $\hat{a}_k + s_{a_k} t_{(\alpha/2,\ N-p)}$ со статистической достоверностью $(1-\alpha)$.

Проиллюстрируем сказанное следующим примером. Пусть число степеней свободы равно 10. Тогда с 90%-ной достоверностью истинные значения параметров $\bar{a}_k$ находятся внутри интервала

$$\hat{a}_k - s_{a_k} \cdot 1{,}812 < \bar{a}_k < \hat{a}_k + s_{a_k} \cdot 1{,}812;$$

90%-ной достоверности соответствует $\alpha/2 = 0{,}05$.

Исходя из неравенства (10.18), можно сделать вывод о достигнутой точности оценок параметров. Не представляет особых трудностей сделать заключение и о точности самой модели $\hat{f} = f(x, \hat{a})$.

Если пренебречь членами более высоких порядков, то для дисперсии оценки функции f при любых x справедливо выражение[1]

$$\sigma_f^2 = \sum_{j=1}^{p} \sum_{k=1}^{p} \left(\frac{\partial f}{\partial a_j}\right) \left(\frac{\partial f}{\partial a_k}\right) \rho_{jk} \sigma_{a_j} \sigma_{a_k}. \qquad (10.19)$$

Вводя величины F_{a_j}, получаем

$$\sigma_f^2 = \sum_{j=1}^{p} \sum_{k=1}^{p} F_{a_j} F_{a_k} \rho_{jk} \sigma_{a_j} \sigma_{a_k}. \qquad (10.20)$$

Поскольку истинные значения σ_j, ρ_{jk} и σ_{a_k} неизвестны, мы будем использовать их оценки s_j, r_{jk} и s_{a_k}:

$$s_f^2 = \sum_{j=1}^{p} \sum_{k=1}^{p} F_{a_j} F_{a_k} r_{jk} s_{a_j} s_{a_k}. \qquad (10.21)$$

Подставляя сюда (10.16) и (10.17), окончательно получаем выражение для дисперсии оценки функции f в некоторой произвольной точке x:

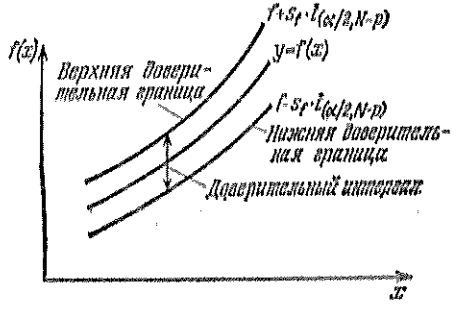
$$s_f^2 = \frac{S}{N-p} \sum_{j=1}^{p} \sum_{k=1}^{p} F_{a_j} F_{a_k} C_{jk}^{-1}. \qquad (10.22)$$

[1] В предыдущих разделах книги дисперсия оценки зависимой переменной обозначалась через $\sigma_{\hat{y}}^2$. — *Прим. ред.*

По аналогии с неравенством (10.18) можно задать доверительные границы для функции f:

$$f(x, \hat{a}) - s_f t_{(\alpha/2, N-p)} \leqslant f(x, \bar{a}) \leqslant f(x, \hat{a}) + s_f t_{(\alpha/2, N-p)}. \quad (10.23)$$

Введение доверительных границ оказывается особенно полезным при выделении наблюдений с грубыми ошибками измерений. В случае когда f зависит лишь от одной переменной, легко представить доверительные границы в графическом виде (фиг. 10.2).



Фиг. 10.2. Аппроксимирующая кривая и ее доверительные границы.

До сих пор при нахождении статистических оценок параметров речь шла лишь о методе наименьших квадратов. Однако имеются и другие статистические методы оценивания параметров. При использовании метода наименьших квадратов следует, строго говоря, различать его взвешенный и невзвешенный варианты. Обычно не требуется знать распределение $p(e)$ ошибок наблюдений зависимой переменной

$$e = \tilde{y} - f(x, \bar{a}).$$

Если же известно значение математического ожидания квадрата ошибок наблюдений

$$E\{ee'\},$$

то можно использовать взвешенный вариант с весами

$$W = E\{(ee')^{-1}\}.$$

Найденные таким способом оценки параметров имеют наименьшую дисперсию.

Остановимся кратко еще на двух методах нахождения параметров. Первый из них — это так называемый метод максимального правдоподобия. Для его использования необходимо знать вид распределения $p(e)$. Если параметры этого распределения не известны, они могут рассчитываться одновременно с оцен-

ками параметров функции f. При этом требуется максимизировать функцию правдоподобия $L(a)$, которая обычно определяется как условная плотность вероятности $p(y|a)$, связывающая значения выхода модели y с ее параметрами a[1]:

$$\max_a L(a),$$
$$L(a) = p(y|a).$$

Если ошибки наблюдений некоррелированны и подчиняются нормальному закону распределения с нулевым средним, то метод максимального правдоподобия совпадает с методом наименьших квадратов. В качестве весов здесь используются обратные величины дисперсий ошибок наблюдений.

Второй метод нашел широкое применение не только при оценивании параметров, но и в теории статистических решений. Речь идет о байесовском методе. По сравнению с уже упомянутыми методами он требует большей априорной информации об объекте. Помимо $p(e)$, требуется иметь также априорную функцию распределения $p(a)$ для параметров a. Ее можно получить либо из ранее проведенных наблюдений, либо из определенных физических соображений. Сущность байесовского метода состоит в определении апостериорной условной функции распределения $p(a|y)$. Этот расчет производится по правилу

$$p(a|y) = \frac{p(y|a)\,p(a)}{p(y)}.$$

Если эти функции известны, с их помощью можно рассчитать оценки параметров a. Это может быть сделано несколькими способами. Укажем три из них.

1. Расчет наиболее вероятных значений параметров a^*. Эти значения получаются из условия

$$\max_a p(a|y).$$

2. Определение значений параметров с минимальной дисперсией. При этом необходимо минимизировать по a^* интеграл

$$\int_{-\infty}^{+\infty} \|a - a^*\|^2 p(a|y)\,da.$$

3. Определение параметров из условия минимизации максимальной ошибки $\|a - a^*\|$. В этом случае a^* — медиана распределения $p(a|y)$.

---

[1] Здесь y — вектор результатов наблюдений зависимой переменной в N опытах. — *Прим. ред.*

Байесовский метод может быть также успешно применен при выборе одной из нескольких возможных моделей. Один из примеров такого применения можно найти в [9].

10.2.2. Программа PREDAN на языке ФОРТРАН

Ниже дается описание программы на языке ФОРТРАН, которая позволяет произвести статистический анализ результатов обработки экспериментальных данных, а также оценить эффективность спланированных экспериментов.

При проведении статистического анализа предполагается, что значения параметров найдены методом наименьших квадратов. Разумеется, можно совместить расчет параметров со статистическим анализом, однако это связано с некоторыми трудностями. Довольно часто оказывается, что итерационный процесс расчета оценок параметров не сходится или плохо сходится. Такие случаи едва ли можно легко выявить, и поэтому имеет смысл разделить процессы нахождения оценок параметров и статистического анализа.

С математической точки зрения программа является не слишком сложной. Ее основу составляют уравнения (10.18) и (10.23). Интерес представляет структура программы. Деление ее на различные подпрограммы оказывается довольно очевидным. Однако невозможно выбрать структуру данных и вариантов планов, одинаково хорошую для всех случаев. Здесь необходимо исходить из ограниченного класса задач. В некоторых ситуациях приведенная здесь программа должна быть несколько видоизменена. Например, одна из таких ситуаций может возникнуть при необходимости в ходе планирования оценить вклад одного дополнительного эксперимента. Расчет заново всей обращаемой матрицы оказывается слишком трудоемким. Более рациональным здесь будет хранение этой матрицы в памяти.

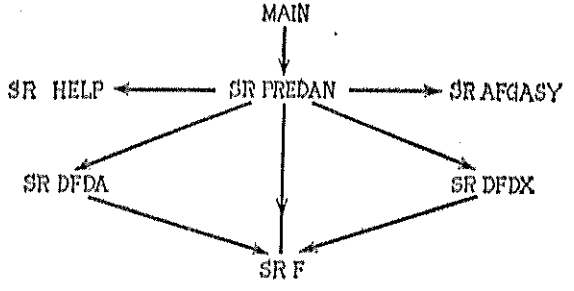
Программа применима при следующей постановке задачи. Пусть задана математическая модель $f(x, a)$, в которой x — независимые переменные и a — оцениваемые параметры.

В тех случаях, когда требуется произвести статистический анализ, задаются оценки $\hat{a}$, которые минимизируют сумму квадратов (10.3) по a. Если же необходимо построить некоторый план эксперимента, а измеренные значения $\tilde{y}^l$ еще отсутствуют, то для решения задачи все равно должны быть заданы оценки параметров a.

Различные возможности программы лучше всего иллюстрируются примерами, приведенными в разд. 10.3.2 и 10.4.2. На фиг. 10.3 изображена схема системы подпрограмм, содержащихся в программе. Выбранная структура этой системы является весьма гибкой. Подпрограмма PREDAN используется во всех

вариантах программы; AFGASY—стандартная программа обращения симметричной матрицы. Все остальные подпрограммы являются проблемно ориентированными.

В F рассчитывается значение функции $f(x, a)$ при заданных значениях x и a. В DFDA или DFDX вычисляются производные функции $f(x, a)$ соответственно по a или x. Если имеются аналитические выражения для этих производных, то F не будет вызываться. HELP—вспомогательная программа, с помощью



Фиг. 10.3. Взаимодействие системы подпрограмм с программой PREDAN: SR=SUBROUTINE

которой пользователь может выполнять произвольные операции. Часть данных для PREDAN задается в MAIN.

Блок-схема подпрограммы PREDAN показана на фиг. 10.4. Текст на ФОРТРАНе приведен ниже.

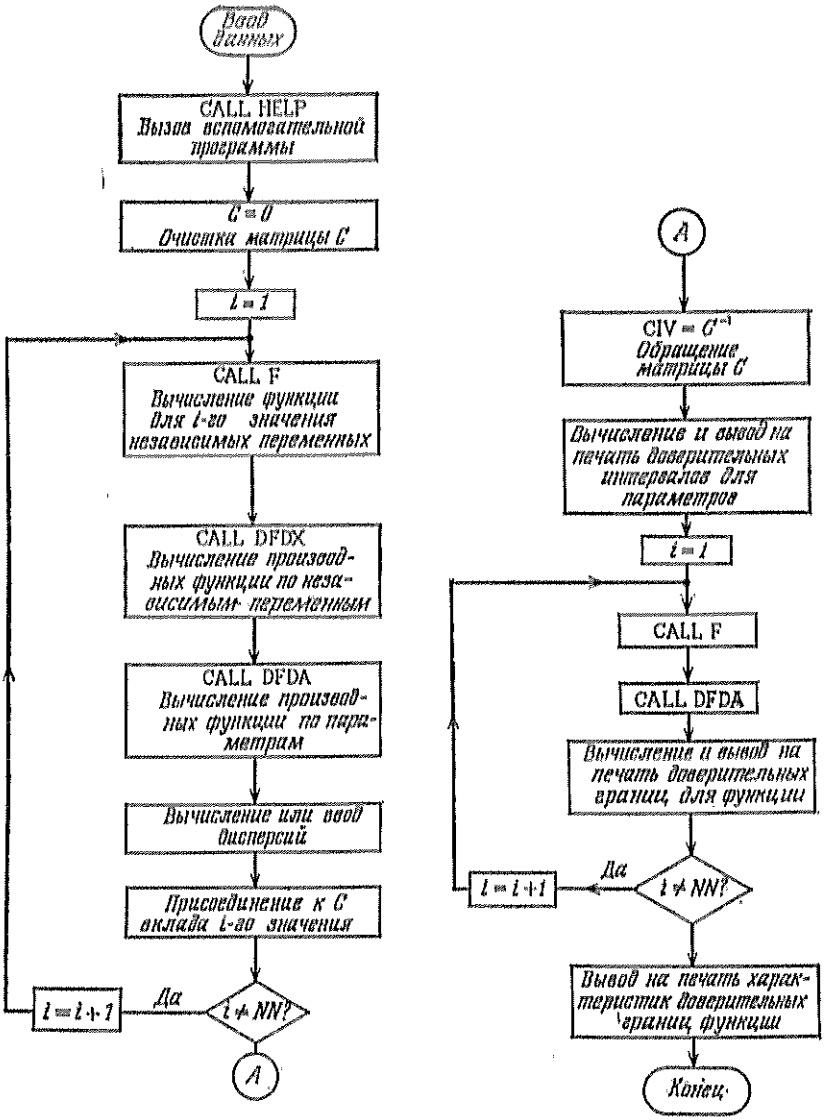
Как для статистического анализа, так и для анализа планов эксперимента необходимо иметь значения среднеквадратичных отклонений $s_{x_j'}$ и $s_{y'}$. В общем случае эти величины зависят от значений переменных:

$$s_{x_j'} = g(\tilde{x}_j'),$$
$$s_{y'} = h(\tilde{y}').$$

На практике эти общие случаи довольно редки, поэтому в программе, кроме этого варианта, предусмотрена возможность использования трех стандартных типов распределений ошибок наблюдений. Выбор того или иного из них определяется заданием одного параметра в исходных данных. К стандартным здесь отнесены следующие типы распределений:

а) постоянная точность

$$s_{x'} = K_{ex},$$
$$s_{y'} = K_{ey};$$



Фиг. 10.4. Блок-схема подпрограммы PREDAN.

б) постоянная относительная точность

$$s_{xl} = K_{fx} |\tilde{x}^l|,$$
$$s_{yl} = K_{fy} |\tilde{y}^l|;$$

в) распределение результатов наблюдений пуассоновского типа

$$s_{xl} = K_{sx} (\tilde{x}^l)^{1/2},$$
$$s_{yl} = K_{sy} (\tilde{y}^l)^{1/2}.$$

Дадим теперь описание исходных данных для PREDAN. Обозначения переменных совпадают с идентификаторами, используемыми в программе. Часть исходных данных передается из MAIN в PREDAN через формальные параметры. Это следующие величины:

1. $((X_{lj},\ i = 1,\ M),\ j = 1,\ NN)$ — Значения независимых переменных
2. NN — Число экспериментов
3. $(YM_l,\ i = 1,\ NN)$ — Измеренные значения выходной переменной (требуются только для статистического анализа)
4. $(YSIG_l,\ i = 1,\ NN)$ — Дисперсии измеренных значений y (только для статистического анализа и LSA = 2)
5. LSA — Управляющий параметр
   - LSA $\leqslant 0$ — Режим планирования эксперимента
   - LSA > 0 — Режим статистического анализа; в зависимости от величины LSA выбирается тот или иной способ задания распределения ошибок измеренных значений y
     - LSA = 1 — В соответствии с величиной IK выбирается одно из трех стандартных распределений
     - LSA = 2 — Распределение задается через YSIG
     - LSA = 3 — В качестве дисперсии будут рассматриваться средние квадраты разностей между измеренными и вычисленными значениями y

Все остальные величины вводятся в самой программе PREDAN:

1. NP, M, IK, NFK, IDR

NP — Число параметров a модели

M — Число независимых переменных

IK — Параметр, управляющий расчетом дисперсий:

IK = 1 — постоянная точность

$s_{x_j^l} = \mathrm{SXK}_j$, $s_{y^l} = \mathrm{SYK}$

IK = 2 — постоянная относительная точность

$s_{x_j^l} = \mathrm{SXK}_j |\tilde{x}_j^l|$, $s_{y^l} = \mathrm{SYK} |\tilde{y}^l|$

IK = 3 — пуассоновское распределение

$s_{x_j^l} = \mathrm{SXK}_j (\tilde{x}_j^l)^{1/2}$, $s_{y^l} = \mathrm{SYK} (\tilde{y}^l)^{1/2}$

NFK — Если доверительные границы для функции $f(x, a)$ необходимо знать не только в точках ранее введенной выборки, то можно задать в NFK другие точки XF, для которых следует напечатать соответствующие доверительные границы. В противном случае задается NFK = 0

IDR — IDR = 0, если доверительные границы надо напечатать для каждой точки из X или XF, иначе IDR = 0

2. $(\mathrm{A}_i,\ i = 1,\ \mathrm{NP})$ — Начальные (при планировании эксперимента) или вычисленные (при статистическом анализе) значения параметров

3. $(\mathrm{DELA}_i,\ i = 1,\ \mathrm{NP})$ — Относительные приращения при расчете производных по параметрам

4. $(\mathrm{DELX}_i,\ i = 1,\ \mathrm{M})$ — Относительные приращения при расчете производных по независимым переменным

5. SYK, $(\mathrm{SXK}_j,\ j = 1,\ \mathrm{M})$ — Параметры дисперсий (см. IK)

6. $((\mathrm{XF}_{ij},\ i = 1,\ \mathrm{M}),\ j = 1,\ \mathrm{NFK})$ — Задается только при NFK $\neq 0$

Программирование проблемно ориентированных подпрограмм производится по следующим правилам:

```
      SUBROUTINE F(X, Y, A, I, M, NN)
      DIMENSION X(1), A(1)
C Y—значение функции f(x, a) в i-й точке для независимой
C переменной X и параметра A
      RETURN
      END
      SUBROUTINE DFDX(X, Y, A, FX, I, M, NN)
      DIMENSION X(1), A(1), FX(1)
C FX(j)—значение производной функции f(x, a) в i-й точке по
C j-й независимой переменной, вычисленное при значении век-
C тора параметров A и значениях независимых переменных X
      RETURN
      END
      SUBROUTINE DFDA(X, Y, A, FA, I, M, NN)
      DIMENSION X(1), A(1), FA(1)
C FA(j)—значение производной функции f(x, a) в i-й точке
C по j-му параметру, вычисленное при значении вектора па-
C раметров A и значениях независимых переменных X
      RETURN
      END
```

Применение подпрограммы PREDAN будет продемонстрировано в следующих разделах.

Перед началом расчетов печатаются все исходные данные, что позволяет их проконтролировать. Так как при обращении матриц могут возникнуть затруднения вычислительного характера, то на печать выводятся обращаемая и обращенная матрицы (C и CIV). Далее следует печать доверительных интервалов для параметров. В режиме статистического анализа выдается таблица, которая позволяет оценить достигнутое качество аппроксимации. Для этого производится сопоставление измеренного и вычисленного значений и печатаются заданное стандартное отклонение, вычисленные абсолютное и относительное отклонения. Для визуального контроля указывается знак разности между измеренным и вычисленным значениями. Звездочками в остальной части таблицы отмечены измеренные значения, лежащие вне данного доверительного интервала. Такая таблица приведена для примера в разд. 10.4.2. Среди результатов приводятся также данные о ширине доверительных интервалов. Поскольку интерес представляет общая характеристика качества аппроксимации, а не доверительные границы в отдельных точках, то на печать выводятся четыре величины, позволяющие сделать выводы об этом качестве.

Следует особо подчеркнуть, что осмысленному использованию программы PREDAN способствует эффективная организация ввода и вывода данных. Данные выводятся таким образом, чтобы можно было „с первого взгляда" получить представление обо всех наиболее существенных величинах и тенденциях.

10.3. Планирование эксперимента

10.3.1. Прямой метод

Планирование эксперимента с помощью прямого, так же как и с помощью непрямого, метода тесно связано с методом наименьших квадратов, а следовательно, и с минимизацией (10.3). В соответствии с прямым методом условия проведения эксперимента оптимальны, когда определитель

$$|C^{-1}| \tag{10.24}$$

имеет наименьшую величину[1]. Это условие имеет простой геометрический смысл. При заданной статистической надежности все допустимые значения параметров находятся в некотором гиперэллипсоиде. Определитель $|C^{-1}|$ пропорционален объему такого гиперэллипсоида. Минимизация определителя матрицы C^{-1} обеспечивает минимум объема гиперэллипсоида для любой заданной статистической надежности. Если мы заинтересованы в одинаковой точности всех оценок параметров, то гиперэллипсоид по возможности должен иметь форму, близкую к сферической. Существуют, однако, такие функции, для которых гиперэллипсоид сильно вытянут (имеет форму сигары) и не может быть превращен в сферу даже с помощью специальных планов эксперимента. Сигароподобная форма всегда указывает на то, что одни параметры оцениваются сравнительно хуже других. Впрочем, это можно заметить также по соотношению собственных значений матрицы C^{-1}. Если отношения собственных значений близки к единице, то параметры оцениваются с одинаковой точностью; в противном случае в точности этих оценок имеются существенные различия.

Прямое планирование эксперимента состоит из следующих этапов. Задают число N планируемых экспериментов. Пусть число независимых переменных равно M. Тогда определитель (10.24) должен быть минимизирован по N векторам независимых переменных x. Таким образом, нужно определить $N \cdot M$ оптимальных значений. Разумеется, должны быть известны

[1]) В этом разделе рассматривается только один критерий планирования, а именно критерий D-оптимальности. Более подробно суть этого критерия поясняется в разд. 6. — *Прим. ред.*

начальные значения. При минимизации оказывается, что для расчета определителя $|C^{-1}|$ необходимо знать как дисперсии x, так и их измеренные значения. Оценки этих величин можно получить почти всегда. Более проблематичной оказывается необходимость задания начальных значений параметров a. Это осложнение уже обсуждалось во введении. Оно вытекает из нелинейности функции $f(x, a)$ по параметрам a.

Ситуация, в которой задана нелинейная функция f и отсутствует достаточно точное представление о значениях параметров a, свидетельствует о плохой постановке задачи. На практике приходится иметь дело с одной из следующих ситуаций. Иногда удается установить примерные величины параметров из теоретических соображений, например равновесные или кинетические константы. С другой стороны, ранее могли рассматриваться объекты подобного типа, так что выводы можно сделать по аналогии. Наконец, может встретиться ситуация, в которой начальные значения параметров можно приближенно установить, исходя из измеренных значений. При этом выбирают такие области в пространстве независимых переменных, в которых существенным оказывается влияние лишь одного или нескольких параметров. Для этих параметров с помощью измеренных значений получают уравнение или систему уравнений, из которых можно вычислить искомые начальные оценки. Нужно заметить, что на практике не встречается задач, в которых требуется планировать эксперимент, не имея представления о поведении функции $f(x, a)$ и тем самым о ее приближенных значениях. Поэтому практически гарантируется получение данных, требуемых для нахождения оценок параметров с помощью этого третьего метода.

До сих пор речь шла только о критерии планирования (10.24). Заметим, однако, что в практических задачах на независимые переменные x всегда наложены ограничения типа неравенств. В простейших случаях это постоянные верхние и нижние границы:

$$\underline{x_l} \leqslant x_l \leqslant \overline{x_l}.$$

Так, степень превращения вещества в химической реакции не может быть меньше нуля или больше единицы; температура в реакторе не должна превышать некоторую верхнюю границу, так как иначе начинает разрушаться катализатор и т. д.

Если эксперимент планируется на сложных объектах, которые, например, содержат обратные связи или циклы, то граничные условия могут оказаться значительно более сложными. Задание их в явном виде может даже стать вообще невозможным. Тогда возникает необходимость учета при поиске экстремума нелинейных и неразрешимых граничных условий типа

равенств или неравенств. Решение экстремальной задачи из-за этого сильно затрудняется и может оказаться практически невыполнимым. Если же этих осложняющих обстоятельств нет, поиск экстремума не представляет больших трудностей.

Программно-техническая реализация расчетов проста. Мы предполагаем, что у пользователя есть стандартные программы решения экстремальных задач. Далее, помимо управляющей программы, в основном предназначенной для ввода исходных данных, необходимо иметь программу, упоминавшуюся при описании PREDAN и выполняющую расчет производных функции $f(x, a)$ по параметрам. Расчет целевой функции планирования, т. е. определителя (10.24), удобнее всего производить с помощью стандартной программы обращения матрицы. Значение определителя получается в ней в качестве дополнительного результата.

Для иллюстрации изложенного метода нами были проведены расчеты на ЭВМ ICL 4130 применительно к примеру, данному во введении и ранее уже решенному Боксом и Лукасом [3]. Для поиска экстремума применялся метод Пауэлла [4]. Кратко приведем результаты, которые совпадают с полученными Боксом и Лукасом другим методом.

Задана модель

$$f(x, a) = \frac{a_1}{a_1 - a_2}[\exp(-a_2 t) - \exp(-a_1 t)].$$

Для искомых параметров a_1 и a_2 известны начальные приближения

$$\hat{a}_1 = 0{,}7, \quad \hat{a}_2 = 0{,}2.$$

Надо произвести два измерения: одно в точке $t = x^1$, второе — в точке $t = x^2$; x^1 и x^2 нужно найти так, чтобы параметры могли быть определены с возможно большей точностью. После поиска экстремума получаем

$$x^1_{\text{опт}} = 1{,}23, \quad x^2_{\text{опт}} = 6{,}86.$$

Задача здесь значительно облегчена тем, что число экспериментов равно числу искомых параметров. Это обстоятельство не является существенным, поскольку целью была лишь демонстрация метода. Однако уже в этом простейшем примере проявляется одна из трудностей. Зависимость определителя от плана эксперимента здесь имеет два локальных экстремума.

У Бокса и Лукаса можно найти также пример оптимального планирования эксперимента для двух независимых переменных. Для решения этого примера мы применим непрямой метод, а для сравнения приведем результаты, полученные с помощью прямого метода.

Степень превращения вещества, распадающегося в химической реакции первого порядка, описывается температурной зависимостью

$$y = \exp\left[-a_1 x_1 \exp\left(-a_2 x_2\right)\right],$$

где a_1 и a_2 — параметры, начальные приближения которых равны $\hat{a}_1 = 1$ и $\hat{a}_2 = 16\,000$; x_1 — текущее время реакции; x_2 — отклонение обратной температуры от обратной начальной температуры T_0:

$$x_2 = T^{-1} - T_0^{-1}.$$

Для температуры заданы ограничения

$$380 \leqslant T = (x_2 + 0{,}0025)^{-1} \leqslant 420.$$

Величина T_0 выбрана равной $T_0 = 400$.

Задача состоит в расчете двух экспериментальных точек, т. е. в определении времени измерения x_1 и соответствующей преобразованной температуры x_2. При этом предполагается, что установка найденных значений x_1 и x_2 производится без ошибок.

Экстремальная задача здесь может быть решена аналитически, в результате чего получаются оптимальные условия проведения эксперимента

$$T = 380 \text{ и } x_1 = 8{,}21,$$
$$T = 420 \text{ и } x_1 = 0{,}15.$$

Приведенные факты позволяют оценить область применения прямых методов. Здесь нужно обратить внимание на следующие обстоятельства:

1. Учет сложных граничных условий для независимых переменных затруднен.
2. Использование методов поиска экстремума может потребовать больших затрат вычислительного времени. Это делает планирование эксперимента дорогостоящим.
3. Может существовать много локальных экстремумов.
4. Прямые методы не позволяют без дополнительных статистических расчетов получить непосредственно применимую количественную оценку точности параметров.
5. Полученные экспериментальные точки при использовании критерия (10.24) являются оптимальными лишь в смысле точности параметров. О точности оценки самой функции f не делается никаких выводов. Преимущество же этого метода состоит в том, что поиск оптимальных условий проведения эксперимента производится автоматически.

Прямой метод часто оказывается легко применимым. Однако из-за больших затрат вычислительного времени и упомянутой неуверенности в качестве результата можно рекомендовать лишь ограниченное применение этого метода. Непрямой метод показал себя более надежным.

10.3.2. Непрямой метод

Мы выбрали термин „непрямой метод" (indirekte Methode) для того, чтобы отчетливо выразить его отличие от прямого метода. Непрямой метод соответствует „методу предсказания" (prediction theory) у Вольберга. Непрямой метод не позволяет автоматически вычислять оптимальные условия проведения эксперимента. Наоборот, программе „предлагается" большое число осмысленных условий проведения эксперимента, из которых выбираются оптимальные. При этом получают количественную оценку ожидаемой точности параметров. Кроме того, понятие „оптимальный" может здесь определяться более индивидуально. Помимо доверительных интервалов для параметров, можно рассмотреть также доверительные пределы для функции $f(x, a)$.

Непрямой метод реализуется с помощью уже описанной программы PREDAN. Метод основывается на формулах (10.18) и (10.23).

На практике задача решается следующим образом. Сначала пишутся и раздельно отлаживаются все необходимые программы. Кроме того, находятся начальные приближения параметров. Затем выбирается совокупность условий проведения эксперимента. Она состоит из NN значений независимых переменных x и оценок дисперсий s_{xj} и s_{yj}. Значения x задаются так, чтобы учитывались возможные граничные условия. Это выполняется проще, чем в случае прямого планирования эксперимента, где должен гарантироваться автоматический учет граничных условий для любых значений x. На основании исходных данных PREDAN вычисляет ожидаемые доверительные интервалы для параметров и функции $f(x, a)$.

Применим PREDAN для решения приведенного в предыдущем разделе примера с реакцией первого порядка непрямым методом. Здесь нельзя сделать количественного вывода о точности параметров, так как число экспериментов равно числу параметров (число степеней свободы для вычисления оценки дисперсии наблюдений равно нулю).

Программе было предложено 7 различных пар точек измерений. Поскольку здесь рассматривался контрольный пример, то эти точки были выбраны вблизи оптимальных точек эксперимента. Функция $f(x, a)$ в рассматриваемой области является монотон-

Таблица 10.1

Результаты непрямого планирования эксперимента для реакции первого порядка

$x_1^{(1)}$ $x_1^{(2)}$	$x_2^{(1)} \times 10^3$ $x_2^{(2)} \times 10^3$	s_{a_1}	s_{a_2}
8,21 0,15	0,1316 —0,119	1,925	15340,0
10,0 1,0	0,1316 —0,119	64,38	489236,0
2,0 0,2	0,1316 —0,119	2,90	23800,0
8,3 0,1	0,1316 —0,119	2,002	15905,0
8,1 0,2	0,1316 —0,119	1,977	15722,0
8,1 0,05	0,1316 —0,119	2,54	19856,0
8,5 0,5	0,1316 —0,119	4,67	35788,0

ной относительно температуры, и потому в этом частном случае температура в экспериментах могла выбираться лишь на нижней или верхней границах. В табл. 10.1 исходные данные приведены в первых двух столбцах. В третьем и четвертом столбцах указаны „доверительные интервалы" для обоих параметров [1].

Из данных таблицы можно сделать вывод о том, что среди представленных вариантов оптимальным является первый. Это совпадает с результатом, полученным прямым методом. Четвертый и пятый варианты дают близкие результаты, остальные оказываются гораздо хуже.

В реальных задачах, для которых числа степеней свободы обычно больше нуля, при статистической обработке, помимо доверительных интервалов для параметров, представляют инте-

[1] В третьем и четвертом столбцах табл. 10.1 приведены значения s_{a_1} и s_{a_2}, пропорциональные величинам доверительных интервалов для коэффициентов модели. — *Прим. ред.*

рее также доверительные границы для функции. При этом можно ограничиться четырьмя „глобальными" характеристиками: средними и максимальными абсолютной и относительной ошибками, или рассмотреть доверительные интервалы в заданных точках XF.

Если нужно исследовать большое число вариантов, рекомендуется изменить вывод результатов решения по программе PREDAN. Можно исключить из этого вывода некоторые величины и автоматизировать выбор оптимального варианта. Одновременно следует очень внимательно подойти к заданию исходных данных, чтобы избежать излишних затрат вычислительного времени.

Если сравнить непрямой метод планирования с прямым, то окажется, что непрямой метод уступает прямому в одном отношении: он ничего не „улаживает" сам. С его помощью лишь делаются количественные заключения о доверительных интервалах для параметров и функции в заданных точках измерения. В остальном он оказывается равноценным прямому методу или даже более предпочтительным. Особенно хорошо в нем учитывается опыт экспериментатора; кроме того, легко выполняются граничные условия для независимых переменных. Преимуществом является также возможность оптимизации плана эксперимента по доверительным интервалам для параметров или для функции $f(x, a)$, или по тем и другим.

Весьма полезным свойством здесь оказывается то, что можно сделать заключение о правильности спланированного эксперимента или о достижимости с его помощью желаемых результатов. Такое заключение имеет чрезвычайно большое экономическое значение, особенно при экспериментах на промышленных установках.

На практике типичными оказываются следующие задачи. Необходимо построить математическую модель технологической установки. К этой модели предъявляются определенные требования; например, чтобы ее параметры с наперед заданной статистической достоверностью находились в определенных границах или чтобы модель с определенной статистической достоверностью находилась в некотором толерантном интервале. Эксперименты планируются на основании некоторого опыта. При этом необходимо реализовать и обработать по меньшей мере столько же экспериментов, сколько имеется параметров в модели. Остается открытым вопрос о том, будут ли практически применимыми результаты, полученные после проведения и обработки эксперимента. Непрямой метод планирования дает возможность оценить ожидаемую точность параметров и модели. Если она оказывается неудовлетворительной, надо спланировать новый эксперимент. При этом может также выясниться, что существующая точность измерений является недостаточной.

10.3.3. Правила планирования эксперимента без вычислений

В этом разделе мы дадим некоторые рекомендации, которые помогут спланировать эксперимент, не производя громоздких вычислений и не прибегая к услугам вычислительной машины. На практике редко удается использовать сразу все эти рекомендации, однако всегда надо стараться учитывать возможно большее их число.

В принципе следует стремиться выбирать область планирования настолько большой, чтобы модель применялась не для экстраполяции, а для интерполяции по участвовавшим в ее построении измеренным значениям. Только в этом случае можно быть уверенным в правильности глобальной структуры модели. При экстраполяции далеко за границы исходной области планирования модель может оказаться совершенно неверной. В пределах выбранной области планирования и на ее границах экспериментальные точки надо распределять по возможности равномерно, но при этом следует учесть следующие ниже замечания.

Это общее правило справедливо лишь в тех случаях, когда нет достаточной уверенности в правильности выбранной модели или когда не удается провести описанное выше предварительное исследование. В иных случаях надо учитывать, что некоторым подобластям в пределах всей области планирования должно отдаваться предпочтение с тем, чтобы обеспечить одинаковую точность определения параметров.

Продемонстрируем этот способ планирования эксперимента на примере необратимой реакции первого порядка

$$y = \frac{a_1}{a_1 - a_2}(\exp(-a_2 t) - \exp(-a_1 t)).$$

Предположим, что эксперименты проводятся в области, для которой справедливы условия

$$a_2 t \ll 1, \quad a_1 t \ll 1,$$

т. е., иначе говоря, при малых значениях t. В этом случае y можно приближенно разложить в ряд и представить в виде

$$y = \frac{a_1}{a_1 - a_2}(1 - a_2 t - (1 - a_1 t)) = \frac{a_1}{a_1 - a_2}(a_1 t - a_2 t) = a_1 t.$$

Следовательно, при малых t можно определить лишь параметр a_1. Интуитивно это легко понять, так как вещество B, прежде чем оно сможет распасться, должно сначала образоваться. Хотя рассмотренный пример является в некотором смысле экстремальным, он позволяет сделать важный вывод. Эксперименты надо проводить в тех точках, в которых определяемые параметры

оказывают существенное влияние на поведение функции $f(x, a)$. В более общем виде можно сказать так: следует проводить измерения там, где что-то „происходит". Соответствующие области легко находятся графически. Для этого изображают функцию $f(x, a)$ в виде проекций, зависящих от отдельных независимых переменных. В качестве значений параметров берут их начальные приближения. Области, наиболее перспективные для проведения измерений, находятся в окрестности максимумов, минимумов или точек перегиба.

Необходимо стремиться к многократному повторению отдельных экспериментов. Во-первых, это позволяет проверить воспроизводимость эксперимента. Только в этом случае имеет смысл статистическая обработка его результатов. Во-вторых, проведение многократных измерений при одних и тех же условиях эксперимента позволяет вычислить дисперсии ошибок наблюдений. Величины, обратные этим дисперсиям, используются в качестве весов при обработке данных с помощью метода наименьших квадратов. Кроме того, сделать вывод о статистической точности параметров можно, лишь зная эти дисперсии. Расчет дисперсий на основании многократных измерений, безусловно, предпочтительнее их субъективной оценки. Число измерений должно быть как можно бóльшим; при этом увеличивается число степеней свободы. Множество измерений в одной экспериментальной точке следует заменить их средним значением, которое содержит всю необходимую информацию. При этом одновременно уменьшается вычислительное время при статистической обработке результатов эксперимента.

Особого внимания заслуживает случай, когда модель задана в виде системы обыкновенных дифференциальных уравнений. Здесь имеется два различных способа планирования заданного числа экспериментов. Во-первых, можно реализовать небольшое число последовательностей измерений так, чтобы каждая из них содержала большое число точек проведения измерений и была распределена по всей траектории. Во-вторых, можно реализовать большое число последовательностей измерений, состоящих лишь из начальных и конечных точек траектории. В первом случае в обработке участвует вся траектория, и его следует, безусловно, предпочесть второму. Только он позволяет определить параметры с достаточной точностью, поскольку при использовании лишь начальных и конечных значений траектории ее изменение между этими точками определяется неудовлетворительно и, главное, неоднозначно.

Количественные выводы о необходимой точности измерений могут быть сделаны лишь после обработки на вычислительной машине (например, с помощью программы PREDAN). Помимо требования обеспечения наибольшей возможной точности (оно

не может быть выполнено по экономическим причинам), можно руководствоваться правилом, заключающимся в том, что точность измерений должна быть тем выше, чем более „пологой" является функция f в некоторой области.

10.4. Обработка экспериментальных данных

10.4.1. Метод наименьших квадратов

Целью обработки экспериментальных данных является отыскание математической модели, которая как можно лучше соответствовала бы этим данным. Эта цель нуждается в уточнении, на что уже было обращено внимание во введении.

Если в качестве критерия различия измеренных и вычисленных значений выбрать сумму квадратов отклонений между ними, то мы получим метод наименьших квадратов. Чисто формально этот метод приводит к минимизации следующей целевой функции:

$$Q = \sum_{l=1}^{N} [h_l(a)]^2,$$

где

$$h_l(a) \approx \tilde{y}^l - f(\tilde{x}^l, a),$$

что, впрочем, не является существенным для минимизации Q. Если же критерием различия является критерий Чебышева, то должна быть минимизирована по a следующая величина:

$$Q_{\text{Ч}} = \max_{1 \leqslant l \leqslant N} |h_l(a)|.$$

Тем самым минимизируется максимальное различие между измеренными и вычисленными значениями. Этот метод будет продемонстрирован на примере в разд. 10.4.6.

Метод наименьших квадратов является хорошо разработанным методом. Имеются эффективные алгоритмы минимизации Q, и, кроме того, могут быть сделаны выводы о статистической точности найденных параметров. При решении минимаксной задачи чебышевской аппроксимации для произвольной нелинейной функции $h_l(a)$ приходится сталкиваться с большими трудностями. Нам не известны способы оценки статистической точности параметров для этого метода. Поэтому мы в дальнейшем ограничимся рассмотрением обработки экспериментальных данных с помощью метода наименьших квадратов.

Для этого метода существенной оказывается классификация, связанная с видом функции $f(x, a)$. При этом специфическая задача возникает в случае, когда речь заходит о так называемой проблеме начальных значений. Чтобы пояснить это, рассмотрим следующий пример:

Пример 10.1

Заданы измеренные значения $\tilde{y}^i$ $(i = 1, N)$ в точках $\tilde{x}^i$. Требуется использовать модель вида

$$y(x, a) = \exp(-ax)$$

для выравнивания наблюдений.

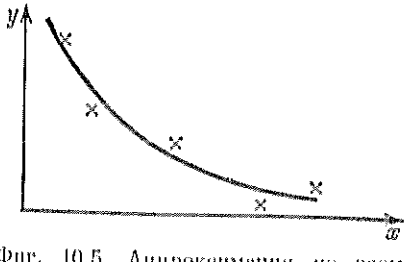
Пример 10.2

Измеренные значения те же, однако модель имеет форму системы дифференциальных уравнений, состоящей в данном случае из одного уравнения:

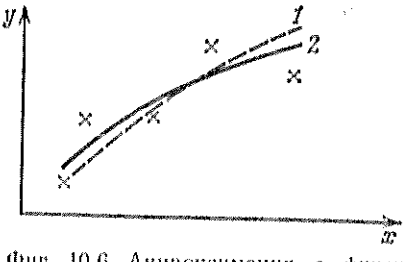
$$\frac{dy}{dx} = z(y, x, a).$$

При этом предполагается, что z — заданная функция параметров и что данное дифференциальное уравнение не может быть проинтегрировано аналитически. Надо особенно подчеркнуть, что измеряется y, а не производная от y.

Минимизация, требуемая в методе наименьших квадратов, выполняется путем вариации параметра a. Для некоторого вектора параметров a вычисляются значения функции $h_i(a)$, а затем Q. В процессе поиска экстремума добиваются, чтобы функция „равномерно" приближала измеренные значения во всех точках (фиг. 10.5).



Фиг. 10.5. Аппроксимация по всем измеренным значениям.



Фиг. 10.6. Аппроксимация с фиксированными (*1*) и варьируемыми (*2*) начальными значениями.

Несколько сложнее обстоит дело, когда модель аналогична рассмотренной в примере 10.2. Расчет значения функции при заданном a в этом случае обычно выполняется путем численного интегрирования дифференциального уравнения, причем задача рассматривается как задача с начальными значениями. Начальным значением для численного интегрирования здесь является результат первого измерения. Отсюда тотчас следует, что эта начальная точка не может более варьироваться путем

изменения параметров в ходе поиска экстремума. Поэтому может встретиться случай, изображенный на фиг. 10.6. Пунктирная линия здесь соответствует фиксированному начальному значению и получена путем оптимального изменения параметров. Если же, напротив, допустить варьирование начального значения, то получится сплошная кривая, которая, совершенно очевидно, в целом является более хорошим приближением.

Казалось, выход из этой трудной ситуации мог бы состоять в том, чтобы включить все начальные значения в процессе поиска экстремума в качестве дополнительных параметров. Однако на практике этот способ вряд ли может быть реализован. Во-первых, при этом резко возрастает число параметров. Во-вторых, начальные значения совсем иначе входят в целевую функцию, чем другие параметры. Вследствие этого чувствительность целевой функции к изменениям параметров также оказывается весьма различной. Из-за этого достижение экстремума становится проблематичным. Мы провели численное исследование такого случая на простом примере. Поиск закончился преждевременно, и результаты оказались далекими от наилучших.

Этим поясняется, какую особую роль играют модели, представленные в виде аналитически не интегрируемых дифференциальных уравнений. Такая особая роль не обнаруживается, если рассматривать и решать систему дифференциальных уравнений как краевую задачу. Поясним это с помощью примера 10.2.

Обычно задача решается путем минимизации Q по a. При этом получается $f(x^l, \hat{a})$ как решение проблемы с начальными значениями (первая точка измерений равна начальному значению для интегрирования). Возникающие при этом неприятности были продемонстрированы с помощью фиг. 10.6.

Чтобы не придавать особого значения первой точке измерения, поступают следующим образом. Вновь минимизируют Q по a, изменив, однако, способ вычисления $f(x^l, a)$. Для этого решают не проблему с начальными значениями, а (в общем случае переопределенную) краевую задачу. В некотором смысле это приводит к оптимизации по начальным значениям. При заданных параметрах a для y должны быть найдены такие начальные условия, чтобы отклонение расчетных значений от измеренных было минимальным. Получаемые таким способом кривые, вообще говоря, не проходят через экспериментально найденные начальные значения. Такой способ соответствует вложенным один в другой поискам экстремума: внешний поиск — по параметрам, внутренний — по начальным значениям. Общую формулировку такой задачи и соответствующее ее решение получают с помощью метода квазилинеаризации Ли [5]. При этом определение оценок параметров превращается в решение краевой задачи. На-

ряду с заданной системой дифференциальных уравнений

$$\frac{dy}{dx} = z(y, x, a)$$

рассматривается система

$$\frac{da}{dx} = 0.$$

Требуется найти такое решение системы, чтобы среднеквадратичное отклонение вычисленных значений от измеренных было минимальным.

Для практической обработки результатов экспериментов важно также знать, какой метод позволяет наиболее эффективно осуществить поиск экстремума Q. Принципиально здесь пригоден любой метод нелинейной оптимизации. Однако практика показывает, что методы, не учитывающие особенностей структуры целевой функции, оказываются не очень эффективными. К ним относятся, например, градиентный метод, метод Пауэлла [4] или Флетчера—Пауэлла [6]. Значительно быстрее получается результат с помощью методов, учитывающих структуру суммы в целевой функции. В качестве примеров укажем на методы Марквардта [2] и Пауэлла [7]. На практике особенно хорошие результаты были нами получены с помощью метода Марквардта.

В связи с применением методов поиска экстремума для минимизации Q возникает целый ряд проблем, которые мы не можем здесь обсуждать или хотя бы только затронуть. К ним относятся: определение начальных приближений для параметров, учет ограничений при поиске экстремума, проблематика численного расчета производных по параметрам и особенно, вопросы сходимости. Проблема сходимости играет решающую роль при применении квазилинеаризации. Здесь есть еще над чем поработать, чтобы найти пути и средства для достижения сходимости в различных ситуациях.

В практике обработки экспериментальных данных часто идут следующим путем. Делается попытка так преобразовать первоначально определяемые параметры, чтобы видоизмененная задача минимизации оказалась более простой для решения. Поясним это на примере. Пусть модель представлена в виде

$$f(x, a) = \frac{x_1}{a_1 + a_2 x_1 x_2 + a_3 x_2}.$$

Определение параметров a_i приводит здесь к проблеме нелинейной оптимизации. С помощью преобразования удается получить решение, выраженное в виде системы линейных уравнений. Для этого приведенное выше уравнение тождественно преобразуют

к виду

$$f(x, a)(a_1 + a_2 x_1 x_2 + a_3 x_2) = x_1.$$

Теперь для расчета параметров минимизируют по a величину

$$\sum_{l=1}^{N} [\tilde{y}^l (a_1 + a_2 \tilde{x}_1^l \tilde{x}_2^l + a_3 \tilde{x}_2^l) - \tilde{x}_1^l]^2.$$

Поскольку измеренные значения $\tilde{y}^l$ и $\tilde{x}_j^l$ являются здесь постоянными, минимизация приводит к линейной системе уравнений, правда, не для самих a. Тем не менее оценки параметров с помощью простых операций могут быть найдены из решения этой системы. Заметим, что в обоих случаях применяется одна и та же модель, но разные критерии оценивания, что, вообще говоря, может привести к несовпадению результатов.

Подобным же образом поступают в случае, когда исходная кривая многократными преобразованиями превращается „почти в прямую" относительно новых параметров. Это позволяет рассчитать параметры методом линейного регрессионного анализа. Часто этого удается достичь, применяя экспоненциальные или логарифмические функции.

Формальное использование таких манипуляций может привести к грубым ошибкам. Если сравнить полученные таким способом параметры с теми, которые были найдены при решении исходной задачи, можно заметить большую разницу. Если бы не было измерительных ошибок, результаты должны были бы совпасть. При нормальных ошибках измерений параметры, полученные с помощью преобразований, могут даже иметь неправильные знаки [8]. Поэтому рекомендуется либо вообще не применять такие преобразования, либо полученные таким образом параметры использовать лишь в качестве начальных значений при последующей обработке исходной задачи методом наименьших квадратов.

До сих пор мы говорили об обработке экспериментальных данных только в смысле минимизации Q. На самом деле это лишь первый этап обработки. После него необходимо еще произвести статистический анализ для определения точности оценок параметров. Наконец, надо исследовать адекватность модели. Это особенно важно, когда не только ищутся числовые значения параметров, но одновременно нужно решить, адекватно ли процесс описывается выбранной моделью.

В следующем разделе будет на реальном примере продемонстрирована обработка эксперимента со статистическим анализом результатов и проверкой адекватности модели.

10.4.2. Определение структурных параметров

Мы не будем останавливаться на физическом существе рассматриваемой задачи, так как пример выбран более или менее произвольно, интерес представляет лишь метод его решения. Рассматриваемая задача возникает при исследовании протонного резонанса. Изменяя напряженность поля x, получают экспериментальные кривые, которые могут иметь один или два максимума. При некоторых упрощенных предположениях теоретически можно получить следующее выражение для таких кривых:

$$f(x) = af_1(x) + (1-a) f_2(x).$$

Таким образом, теоретическая кривая представляет собой линейную комбинацию двух основных функций f_1 и f_2, где

$$f_1(x) = \frac{x}{b^3 \sqrt{2\pi}} \exp\left(-\frac{x^2}{2b^2}\right)$$

и

$$f_2(x) = (4\beta^3\sqrt{6\pi})^{-1} \int_{-2\alpha}^{\alpha} \left(\frac{x-h}{\alpha}\right)\left(1-\frac{h}{\alpha}\right)^{1/2} \exp\left(-\frac{(x-h)^2}{2\beta^2}\right) dh +$$

$$+ (4\beta^3\sqrt{6\pi})^{-1} \int_{-\alpha}^{2\alpha} \left(\frac{x-h}{\alpha}\right)\left(1+\frac{h}{\alpha}\right)^{-1/2} \exp\left(-\frac{(x-h)^2}{2\beta^2}\right) dh.$$

В теоретическом выражении для $f(x)$ есть четыре неизвестных параметра, которые нужно определить из эксперимента: a, b, α, β.

При вычислении f_2 интегралы могут быть преобразованы следующим образом:

$$f_2(x) = -\frac{1}{2\beta^2\sqrt{2\pi}}\left[t_1 \exp\left(-\frac{t_1^2}{2}\right) + t_4 \exp\left(-\frac{t_4^2}{2}\right)\right] -$$

$$- \frac{\sqrt{\beta/\alpha}}{2\beta^2\sqrt{6\pi}} \int_{t_1}^{t_3} \sqrt{t_3 - t}\,(1-t^2) \exp\left(-\frac{t^2}{2}\right) dt +$$

$$+ \frac{\sqrt{\beta/\alpha}}{2\beta^2\sqrt{6\pi}} \int_{t_2}^{t_4} \sqrt{t - t_2}\,(1-t^2) \exp\left(-\frac{t^2}{2}\right) dt,$$

где

$$t_1 = \frac{-2\alpha - x}{\beta}, \quad t_2 = \frac{-\alpha - x}{\beta},$$

$$t_3 = \frac{\alpha - x}{\beta}, \quad t_4 = \frac{2\alpha - x}{\beta}.$$

Экспериментальным путем для эквидистантных значений x (около 30 точек) были измерены величины $\tilde{y}^i$. Измерения в каждой точке

повторялись многократно, так что $\tilde{y}^l$ представляют собой средние значения параллельных наблюдений. Это позволило также вычислить дисперсию v^l для $\tilde{y}^l$ в каждой точке $\tilde{x}^l$:

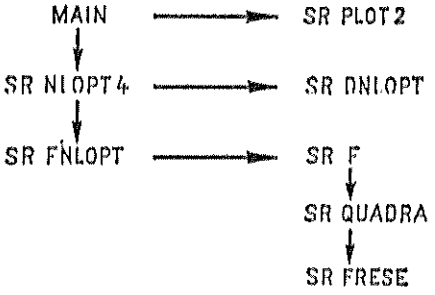
$$v^l = \frac{1}{k-1} \sum_{l=1}^{k} (\tilde{y}^{lj} - \tilde{y}^l)^2,$$

где $\tilde{y}^{lj}$ — отдельные измерения, а $\tilde{y}^l$ — рассчитанное по ним среднее значение (полученное при фиксированном $\tilde{x}^l$). Расчет параметров выполнялся путем минимизации целевой функции

$$Q = \sum_{l=1}^{N} \frac{1}{v^l} (\tilde{y}^l - f(\tilde{x}^l))^2$$

методом Марквардта.

Для вычисления значений $f(x)$ применялось численное интегрирование. Обработка данных является в данном случае довольно сложной проблемой, так как оцениваемые параметры



Фиг. 10.7. Система программ для расчета структурных параметров.

входят в пределы определенных интегралов. Параметры рассчитывались на ЭВМ ICL 4130 с помощью системы программ, написанных на ФОРТРАНе. Структура этой системы показана на фиг. 10.7.

Подпрограммы NLOPT 4 (нелинейная оптимизация методом Марквардта) и QUADRA (численное интегрирование) являются стандартными программами. Все остальные программы были разработаны как проблемно-ориентированные. В MAIN экспериментальные данные считываются и выводятся на контрольную печать. Исходные данные целесообразно печатать в том же формате, в котором они представлены на носителе информации: так легче производить их контроль. Помимо экспериментальных данных, в MAIN вводятся также управляющие параметры и начальные приближения для поиска экстремума. Затем управление передается подпрограмме NLOPT 4. Расчет минимизируемой це-

левой функции выполняется в FNLOPT. Для каждого значения x с помощью подпрограммы F должно вычисляться значение $f(x)$. Для этого производится численное интегрирование в QUADRA. Подынтегральное выражение запрограммировано в FRESE. Подпрограмма DNLOPT при получении соответствующего управления позволяет рассчитать производные целевой функции по параметрам. Поскольку в нашем случае этого нельзя сделать аналитически, DNLOPT не будет вызываться.

Во время поиска экстремума в NLOPT 4 промежуточные результаты будут выдаваться на печать. Вначале печатаются управляющие параметры. Затем для каждой итерации следует строка, в которую входят значение целевой функции, несущественный для пользователей параметр метода (λ), число вычислений значений функции на этой итерации (не считая образования производных) и значения параметров.

После окончания поиска экстремума управление вновь передается в MAIN, где происходит выдача результатов. Она начинается с печати заголовка для последовательности измерений, за которым следуют оптимальные оценки параметров. Затем в виде таблицы приводятся значения независимых переменных x, измеренные значения $\tilde{y}$, вычисленные оценки значений $\hat{y}$, абсолютные и относительные ошибки.

Для быстрой оценки результатов удобна графическая форма их представления. Для этого применяется стандартная программа для быстрой печати. Выдача кривой через графопостроитель оказывается по сравнению с этим способом более медленной, хотя качество изображения значительно лучше. По экономическим причинам мы предпочли быструю печать.

Слева, рядом с графическим изображением, еще раз приводятся: текущий номер, измеренное значение и рассчитанное значение. На картинке измеренное значение представляется цифрой „1“, рассчитанное значение цифрой „2“. Если оба значения попадают в один и тот же интервал дискретизации, печатается звездочка. Масштаб картинки выбирается оптимальным самой стандартной подпрограммой.

Графическое представление облегчает визуальную оценку данных и, кроме того, дает возможность быстро составить себе представление о полученных результатах. На фиг. 10.8 приведены полные результаты анализа кривой [1]).

С помощью отдельной программы был произведен статистический анализ полученных результатов. При этом, естественно, применялись уже имевшиеся подпрограммы F, QUADRA и FRESE.

---

[1]) Пояснения к распечаткам программ на фиг. 10.8, 10.9, 10.11 и 10.12 см. в конце раздела (стр. 374). — *Прим. ред.*

```
0.5227E-02      0.2738E-05
0.8591E-02      0.3687E-05
0.9984E-02      0.3441E-05
0.1145E-01      0.4077E-05
0.1207E-01      0.4456E-05
0.1267E-01      0.3054E-05
0.1334E-01      0.3087E-05
0.1412E-01      0.2903E-05
0.1470E-01      0.2939E-05
0.1501E-01      0.3477E-05
0.1455E-01      0.1745E-05
0.1316E-01      0.6265E-06
0.1188E-01      0.5976E-06
0.1023E-01      0.4231E-06
0.8279E-02      0.3421E-06
0.6469E-02      0.6026E-06
0.5017E-02      0.1114E-05
0.4304E-02      0.7020E-06
0.3449E-02      0.7043E-06
0.2758E-02      0.5440E-06
0.2030E-02      0.4283E-06
0.1464E-02      0.3543E-06
0.9595E-03      0.2562E-06
0.6655E-03      0.2774E-06
0.4124E-03      0.2231E-06
0.1988E-03      0.1654E-06
0.1267E-03      0.1124E-06
0.6335E-04      0.2809E-07
0.1584E-04      0.1756E-08
```

Фиг. 10.8. Распечатка результатов примера разд. 10.4.2 (I).

```
    VERFAHREN NACH ** MARQUARDT **

INTEGER PARAMETER        4        20        0        5        2        29        2        1
  REAL PARAMETER   0.100E-03  0.100E-05  0.500E-03  0.500E+00  0.100E+01  0.100E-03  0.000E+00
    SCHAETZWERTE   0.4500000E+01  0.2000000E+01  0.1800000E+01  0.2000000E+00
0.2290241819E+02  -0.69315E+00  0.4500000E+01  0.2000000E+01  0.1800000E+01  0.2000000E+00
0.3633359009E+01  0.500E+00  3  0.4457708E+01  0.1793609E+01  0.1915132E+01  0.2186074E+00
0.1685972893E+01  0.182E+00  8  0.4290254E+01  0.1852667E+01  0.1847476E+01  0.1911372E+00
0.1508296071E+01  0.612E-04 10  0.4230109E+01  0.1880008E+01  0.1812176E+01  0.1790834E+00
0.1499984025E+01  0.225E-04  2  0.4212186E+01  0.1889296E+01  0.1803659E+01  0.1759205E+00
0.1499768350E+01  0.828E-05  2  0.4210268E+01  0.1890230E+01  0.1802641E+01  0.1755496E+00
0.1499748004E+01  0.305E-05  2  0.4209824E+01  0.1890409E+01  0.1802389E+01  0.1754665E+00
0.1499745121E+01  0.305E-05  2  0.4209726E+01  0.1890445E+01  0.1802331E+01  0.1754481E+00

KEINE VERBESSERUNG DES FUNKTIONSWERTES
 0.1499745121E+01   0.30475E-05  0.4209726E+01  0.1890445E+01  0.1802331E+01  0.1754481E+00
```

```
TESTB
A=      1.754E-01      B=      1.802E+00      ALPHA=      4.210E+00      BETA=      1.890E+00

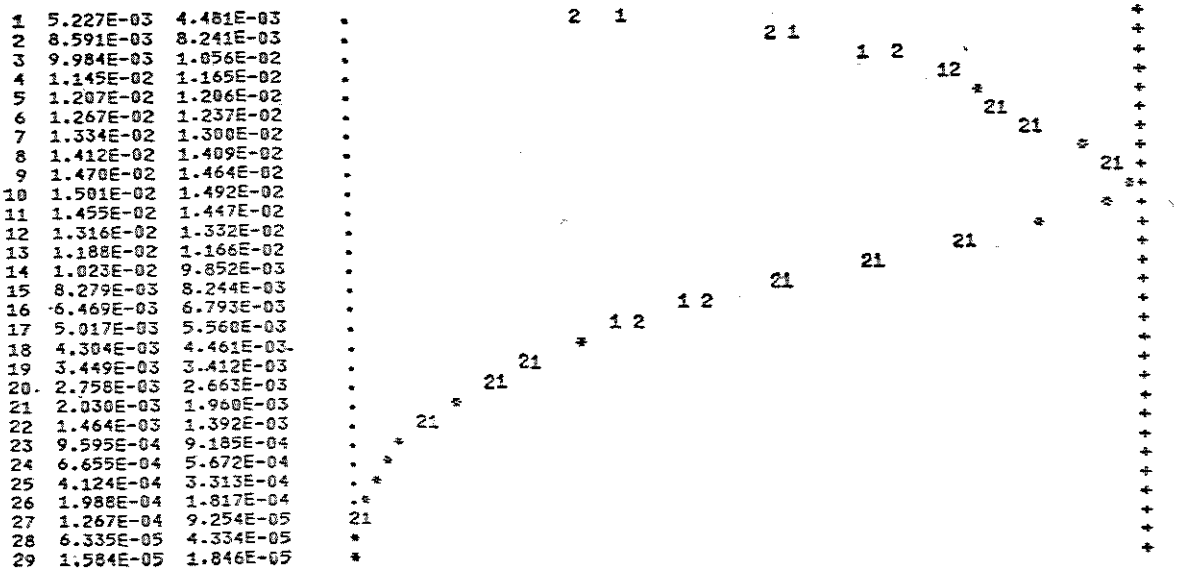
  H(I)          G(I)          F(I)
 0.5020     5.227E-03     4.461E-03     7.461E-04     1.427E-01
 1.0040     8.591E-03     8.241E-03     3.498E-04     4.072E-02
 1.5060     9.984E-03     1.056E-02    -5.792E-04    -5.802E-02
 2.0080     1.145E-02     1.165E-02    -1.965E-04    -1.716E-02
 2.5100     1.207E-02     1.206E-02     5.661E-06     4.690E-04
 3.0120     1.267E-02     1.237E-02     2.977E-04     2.350E-02
 3.5140     1.334E-02     1.300E-02     3.443E-04     2.581E-02
 4.0160     1.412E-02     1.409E-02     2.932E-05     2.077E-03
 4.5180     1.470E-02     1.464E-02     5.984E-05     4.070E-03
 5.0200     1.501E-02     1.492E-02     9.214E-05     6.139E-03
 5.5220     1.455E-02     1.447E-02     8.343E-05     5.734E-03
 6.0240     1.316E-02     1.332E-02    -1.612E-04    -1.225E-02
 6.5260     1.188E-02     1.166E-02     2.170E-04     1.827E-02
 7.0280     1.023E-02     9.852E-03     3.778E-04     3.693E-02
 7.5300     8.279E-03     8.244E-03     3.477E-05     4.199E-03
 8.0320     6.469E-03     6.793E-03    -3.238E-04    -5.005E-02
 8.5340     5.017E-03     5.560E-03    -5.431E-04    -1.082E-01
 9.0360     4.304E-03     4.461E-03    -1.569E-04    -3.646E-02
 9.5380     3.449E-03     3.412E-03     3.669E-05     1.064E-02
10.0400     2.758E-03     2.663E-03     9.548E-05     3.462E-02
10.5420     2.030E-03     1.960E-03     7.000E-05     3.448E-02
11.0440     1.464E-03     1.392E-03     7.196E-05     4.915E-02
11.5460     9.595E-04     9.185E-04     4.103E-05     4.276E-02
12.0480     6.655E-04     5.672E-04     9.830E-05     1.477E-01
12.5500     4.124E-04     3.313E-04     8.108E-05     1.966E-01
13.0520     1.986E-04     1.817E-04     1.706E-05     8.581E-02
13.5540     1.267E-04     9.254E-05     3.416E-05     2.696E-01
14.0560     6.335E-05     4.334E-05     2.001E-05     3.159E-01
14.5580     1.584E-05     1.846E-05    -2.620E-06    -1.654E-01
```

Фиг. 10.8 (продолжение).

GEMESSENE UND BERECHNETE WERTE

1	5.227E-03	4.481E-03	.
2	8.591E-03	8.241E-03	.
3	9.984E-03	1.056E-02	.
4	1.145E-02	1.165E-02	.
5	1.207E-02	1.206E-02	.
6	1.267E-02	1.237E-02	.
7	1.334E-02	1.300E-02	.
8	1.412E-02	1.409E-02	.
9	1.470E-02	1.464E-02	.
10	1.501E-02	1.492E-02	.
11	1.455E-02	1.447E-02	.
12	1.316E-02	1.332E-02	.
13	1.188E-02	1.166E-02	.
14	1.023E-02	9.852E-03	.
15	8.279E-03	8.244E-03	.
16	6.469E-03	6.793E-03	.
17	5.017E-03	5.560E-03	.
18	4.304E-03	4.461E-03	.
19	3.449E-03	3.412E-03	.
20	2.758E-03	2.663E-03	.
21	2.030E-03	1.960E-03	.
22	1.464E-03	1.392E-03	.
23	9.595E-04	9.185E-04	.
24	6.655E-04	5.672E-04	.
25	4.124E-04	3.313E-04	.
26	1.988E-04	1.817E-04	.
27	1.267E-04	9.254E-05	21
28	6.335E-05	4.334E-05	*
29	1.584E-05	1.846E-05	*

END



Фиг. 10.8 (продолжение).

Распечатка результатов представлена на фиг. 10.9. Она начинается с выдачи исходных данных, сопровождаемых текстом и поэтому не требующих дополнительных комментариев. Ниже суммы квадратов ошибок для контроля печатаются обращаемая и обращенная матрицы. Основные результаты печатаются в виде таблицы. Сначала даются доверительные интервалы для параметров. Параметры приводятся в последовательности α, β, b, a. Если выбрать 90%-ный доверительный интервал, справедливо следующее утверждение. В 90% всех случаев истинные значения параметров находятся в следующих границах:

$$\alpha = 4{,}2097 \pm 0{,}1039,$$
$$\beta = 1{,}8904 \pm 0{,}0714,$$
$$b = 1{,}8013 \pm 0{,}0825,$$
$$a = 1{,}7545 \pm 0{,}0208.$$

Границы доверительных интервалов определяются из уравнения (10.18).

Следующая за этим таблица на фиг. 10.9 имеет особое значение для ответа на вопрос об адекватности модели. Каждая строка соответствует одному измеренному значению. Кроме измеренного значения, на печать выводятся: вычисленное значение, среднеквадратичное отклонение, полученные абсолютное и относительное отклонения измеренного значения от рассчитанного и знак этих отклонений. Звездочка в одном из семи последних столбцов показывает, что измеренное значение находится вне доверительного интервала, вычисленного с соответствующей статистической достоверностью. Так, второе измеренное значение еще находится в 99%-ном интервале, но уже вне 98%-ного интервала.

Эта таблица позволяет сделать важные выводы об адекватности модели. Двумя наиболее существенными характеристиками для контроля адекватности являются доверительный интервал в каждой точке измерения и случайность остаточных отклонений.

Если измеренное значение лежит вне соответствующего доверительного интервала, это может быть вызвано тремя причинами. Во-первых, может иметь место ошибка в измеренном значении из-за сбоя при считывании, передаче данных и т. п. Во-вторых, может оказаться, что дисперсии независимых и зависимой переменных заданы слишком малыми. Наконец, расхождение может быть вызвано неадекватностью модели.

Наиболее надежное заключение об адекватности модели делается на основании изучения остаточных отклонений измеренных значений от вычисленных. Если модель адекватна, эти отклонения оказываются чисто случайными величинами. Напротив, если в их распределении имеются систематические состав-

```
PREDICTION ANALYSIS

      1 UNABHAENGIGE VARIABLE

      4 PARAMETER

     29 SAETZE UNABHAENGIGER VARIABLER

UNABHAENGIGE VERAENDERLICHE

 5.0200E-01  1.0040E+00  1.5060E+00  2.0080E+00  2.5100E+00  3.0120E+00  3.5140E+00  4.0160E+00  4.5180E+00  5.0200E+00
 5.5220E+00  6.0240E+00  6.5260E+00  7.0280E+00  7.5300E+00  8.0320E+00  8.5340E+00  9.0360E+00  9.5380E+00  1.0040E+01
 1.0542E+01  1.1044E+01  1.1546E+01  1.2048E+01  1.2550E+01  1.3052E+01  1.3554E+01  1.4056E+01  1.4558E+01

PARAMETER

 4.2097E+00  1.8904E+00  1.8013E+00  1.7545E-01

VERTEILUNG DER MESSFEHLER -IK= 1

          SYK=0.0000E+00 SXK=1.0000E-02

STREUUNG AUS MESSWERTEN BERECHNET

MESSWERTE

 5.2270E-03  8.5910E-03  9.9840E-03  1.1450E-02  1.2070E-02  1.2670E-02  1.3340E-02  1.4120E-02  1.4700E-02  1.5010E-02
 1.4550E-02  1.3160E-02  1.1880E-02  1.1230E-02  8.2790E-03  6.4690E-03  5.0170E-03  4.3040E-03  3.4490E-03  2.7580E-03
 2.0300E-03  1.4640E-03  9.5950E-04  6.5550E-04  4.1240E-04  1.9880E-04  1.2670E-04  6.3350E-05  1.5840E-05
```

Фиг. 10.9. Результаты статистического анализа.

PARAMETERZUWAECHSE BEI ABLEITUNGSBILDUNG

5.0000E-04 5.0000E-04 5.0000E-04 5.0000E-06

ZUWACHS UNABHAENGIGER VARIABLER BEI ABLEITUNGSBILDUNG

1.0000E-03

FEHLERQUADRATSUMME 1.497828E+00

0.15606E+03	0.62159E+02	0.96406E+01	-0.57304E+03	0.62159E+02	0.22743E+03	-0.15307E+03	0.66804E+03
0.96406E+01	-0.15307E+03	0.24617E+03	-0.11107E+04	-0.57304E+03	0.66804E+03	-0.11107E+04	0.80148E+04
0.32115E-01	-0.15601E-01	0.14064E-01	0.55455E-02	-0.15601E-01	0.15150E-01	-0.18668E-02	-0.26369E-02
0.14064E-01	-0.18668E-02	0.20257E-01	0.39684E-02	0.55455E-02	-0.26369E-02	0.39684E-02	0.12910E-02

PARAMETER	99%	98%	95%	90%	80%	70%	60%
4.2097E+00	1.695312E-01	1.511608E-01	1.253083E-01	1.038964E-01	8.005133E-02	6.435738E-02	5.206986E-02
1.8904E+00	1.164397E-01	1.038222E-01	8.606591E-02	7.135950E-02	5.498191E-02	4.420278E-02	3.576331E-02
1.8013E+00	1.346419E-01	1.200521E-01	9.952002E-02	8.251466E-02	6.357687E-02	5.111271E-02	4.135395E-02
1.7545E-01	3.399087E-02	3.030761E-02	2.512421E-02	2.083115E-02	1.605023E-02	1.290360E-02	1.043996E-02

GUETE DES ANGLEICHS DER LEAST-SQUARE-ANALYSE

* - MESSWERT AUSSERHALB DES ENTSPRECHENDEN KONFIDENZBANDES

MESSWERT	BER.WERT	STANDARDABW.	ABS.ABW.	REL.ABW.	SIGN.ABW.	99%	98%	95%	90%	80%	70%	60%
5.22700E-03	4.49050E-03	2.738E-06	7.37E-04	1.41E-01	+	*	*	*	*	*	*	*
8.59100E-03	8.25693E-03	3.687E-06	3.34E-04	3.89E-02	+							*
9.98400E-03	1.05798E-02	3.441E-06	-5.96E-04	-5.97E-02	-					*	*	*
1.14500E-02	1.16593E-02	4.077E-06	-2.09E-04	-1.83E-02	-							
1.20700E-02	1.20712E-02	4.456E-06	-1.22E-06	-1.01E-04	-							
1.26700E-02	1.23735E-02	3.054E-06	2.97E-04	2.34E-02	+							*
1.33400E-02	1.29931E-02	3.087E-06	3.47E-04	2.60E-02	+						*	*
1.41200E-02	1.40866E-02	2.903E-06	3.34E-05	2.36E-03	+							
1.47000E-02	1.46363E-02	2.939E-06	6.37E-05	4.33E-03	+							
1.50100E-02	1.49150E-02	3.477E-06	9.50E-05	6.33E-03	+							
1.45500E-02	1.44647E-02	1.745E-06	8.53E-05	5.86E-03	+							
1.31600E-02	1.33201E-02	6.265E-07	-1.60E-04	-1.22E-02	-						*	*
1.18800E-02	1.16624E-02	5.978E-07	2.18E-04	1.83E-02	+			*	*	*	*	*
1.02300E-02	9.85182E-03	4.231E-07	3.78E-04	3.70E-02	+	*	*	*	*	*	*	*
8.27900E-03	8.24400E-03	3.421E-07	3.50E-05	4.23E-03	+							
6.46900E-03	6.79263E-03	6.026E-07	-3.24E-04	-5.00E-02	-	*	*	*	*	*	*	*
5.01700E-03	5.55996E-03	1.114E-06	-5.43E-04	-1.08E-01	-	*	*	*	*	*	*	*
4.30400E-03	4.46083E-03	7.020E-07	-1.57E-04	-3.64E-02	-	*	*	*	*	*	*	*
3.44900E-03	3.41224E-03	7.043E-07	3.68E-05	1.07E-02	+							
2.75800E-03	2.66244E-03	5.440E-07	9.56E-05	3.46E-02	+						*	*
2.03000E-03	1.95991E-03	4.283E-07	7.01E-05	3.45E-02	+							*
1.46400E-03	1.39195E-03	3.543E-07	7.20E-05	4.92E-02	+						*	*
9.59500E-04	9.18393E-04	2.562E-07	4.11E-05	4.28E-02	+							
6.65500E-04	5.67133E-04	2.774E-07	9.84E-05	1.48E-01	+			*	*	*	*	*
4.12400E-04	3.31271E-04	2.231E-07	8.11E-05	1.97E-01	+	*	*	*	*	*	*	*
1.98800E-04	1.81710E-04	1.654E-07	1.71E-05	8.60E-02	+							*
1.26700E-04	9.25175E-05	1.124E-07	3.42E-05	2.70E-01	+	*	*	*	*	*	*	*
6.33500E-05	4.33263E-05	2.809E-08	2.00E-05	3.16E-01	+	*	*	*	*	*	*	*
1.58400E-05	1.84542E-05	1.756E-09	-2.61E-06	-1.65E-01								*

KONFIDENZBAENDER (BEZOGEN AUF MESSWERTE)

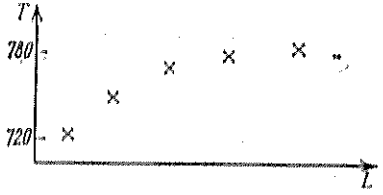
	99%	98%	95%	90%	80%	70%	60%
MITTL.ABS.FEHLER	3.9026E-04	3.4797E-04	2.8846E-04	2.3917E-04	1.8428E-04	1.4815E-04	1.1986E-04
MITTL.REL.FEHLER	11.46%	10.22%	8.47%	7.02%	5.41%	4.35%	3.52%
MAX.ABS.FEHLER	1.0052E-03	8.9626E-04	7.4298E-04	6.1602E-04	4.7464E-04	3.8159E-04	3.0873E-04
MAX.REL.FEHLER	44.93%	40.07%	33.21%	27.54%	21.22%	17.06%	13.80%

END

Фиг. 10.9 (продолжение).

ляющие, модель выбрана неверно. В шестом столбце таблицы дополнительно печатаются знаки остаточных отклонений.

При печатании результатов в конце приводится также несколько характеристик доверительных интервалов. Эти величины представляют наибольший интерес при планировании эксперимента, так как в этом случае построение доверительного интервала в каждой планируемой точке часто сопряжено с излишними затратами вычислительного времени.



Фиг. 10.10. Продольный температурный профиль реактора.

К числовым значениям относительных величин надо подходить с некоторой осторожностью. Во многих случаях эти результаты демонстрируют точность, которой нет на самом деле.

Рассмотрим один из подобных примеров. Необходимо построить модель химического реактора. Для этого, кроме всего прочего, был экспериментально построен продольный температурный профиль (фиг. 10.10). Если имеется математическая модель, то ошибку, отнесенную к измеренному значению, можно записать в виде

$$\delta_1 = \left| \frac{T_{\text{эксп}} - T_{\text{расч}}}{T_{\text{эксп}}} \right|.$$

Если же ради наглядности отнести ошибку к средней температуре $T = 750°$, получаем

$$\delta_1 = \left| \frac{T_{\text{эксп}} - T_{\text{расч}}}{750} \right|.$$

Вычисленная таким способом ошибка, естественно, зависит от уровня измеряемых температур. Значительно больший смысл имело бы такое определение относительной ошибки:

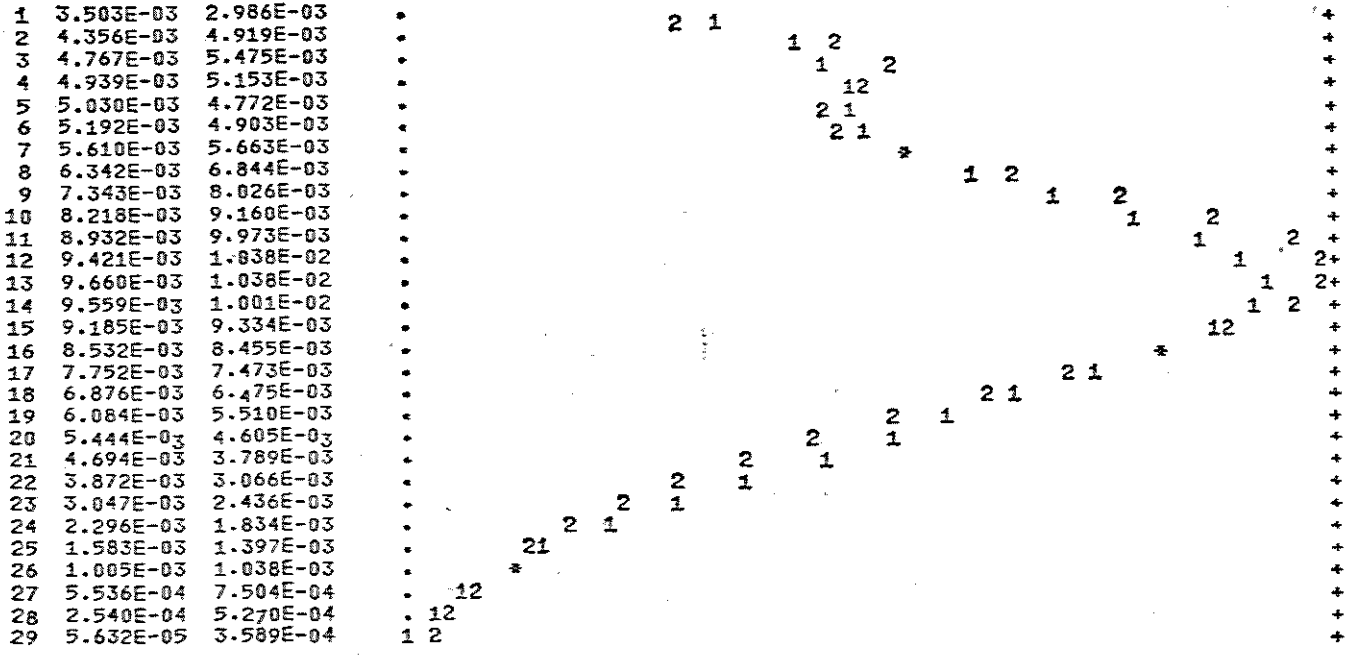
$$\delta_2 = \left| \frac{T_{\text{эксп}} - T_{\text{расч}}}{\Delta T_{\text{макс}}} \right|,$$

где $\Delta T_{\text{макс}}$ — ширина интервала, в котором вообще может изменяться температура в данном эксперименте. В приведенном выше примере $\Delta T_{\text{макс}} = 60°$.

При $|T_{\text{эксп}} - T_{\text{расч}}| = 5°$ мы при использовании того и другого определения относительной ошибки получим соответственно величины 0,0067 и 0,083. Разница очевидна, причем первая

.GEMESSENE UND BERECHNETE WERTE

1	3.503E-03	2.986E-03
2	4.356E-03	4.919E-03
3	4.767E-03	5.475E-03
4	4.939E-03	5.153E-03
5	5.030E-03	4.772E-03
6	5.192E-03	4.903E-03
7	5.610E-03	5.663E-03
8	6.342E-03	6.844E-03
9	7.343E-03	8.026E-03
10	8.218E-03	9.160E-03
11	8.932E-03	9.973E-03
12	9.421E-03	1.038E-02
13	9.660E-03	1.038E-02
14	9.559E-03	1.001E-02
15	9.185E-03	9.334E-03
16	8.532E-03	8.455E-03
17	7.752E-03	7.473E-03
18	6.876E-03	6.475E-03
19	6.084E-03	5.510E-03
20	5.444E-03	4.605E-03
21	4.694E-03	3.789E-03
22	3.872E-03	3.066E-03
23	3.047E-03	2.436E-03
24	2.296E-03	1.834E-03
25	1.583E-03	1.397E-03
26	1.005E-03	1.038E-03
27	5.536E-04	7.504E-04
28	2.540E-04	5.270E-04
29	5.632E-05	3.589E-04



Фиг. 10.11. Результаты примера структурного анализа (II).

величина создает видимость несуществующей на самом деле точности.

Вероятно, самой трудной частью статистической обработки экспериментальных данных является принятие решения о том, могут ли выбранная модель и числовые значения оценок параметров считаться правильными или нет. Обсуждавшиеся здесь тесты предоставляют определенную свободу действий. Каждый тест дает лишь некоторое указание. Решающее слово остается за тем, кто выполняет статистическую обработку. Это решение существенным образом зависит от цели построения модели. В нашем примере отчетливо видна определенная систематическая составляющая в знаках остаточных отклонений. Следовательно, модель не может быть названа полностью адекватной. То же самое касается и случая, когда отклонения превышают величины, соответствующие дисперсиям зависимой переменной. С другой стороны, доверительные интервалы для параметров оказываются здесь относительно узкими.

Об адекватности модели трудно судить по данным одной выборки, даже если при этом речь идет о средних значениях в каждой экспериментальной точке. В рассмотренной задаче исследовалась даже не одна кривая, а целое их семейство. Кривые соответствовали разным веществам, а также разным значениям параметров. Результаты, показанные на фиг. 10.11 и 10.12, свидетельствуют о том, что модель является недостаточно полной, чтобы правильно объяснить имеющуюся совокупность измерений. Таким образом, здесь следует отрицательно ответить на вопрос об адекватности модели.

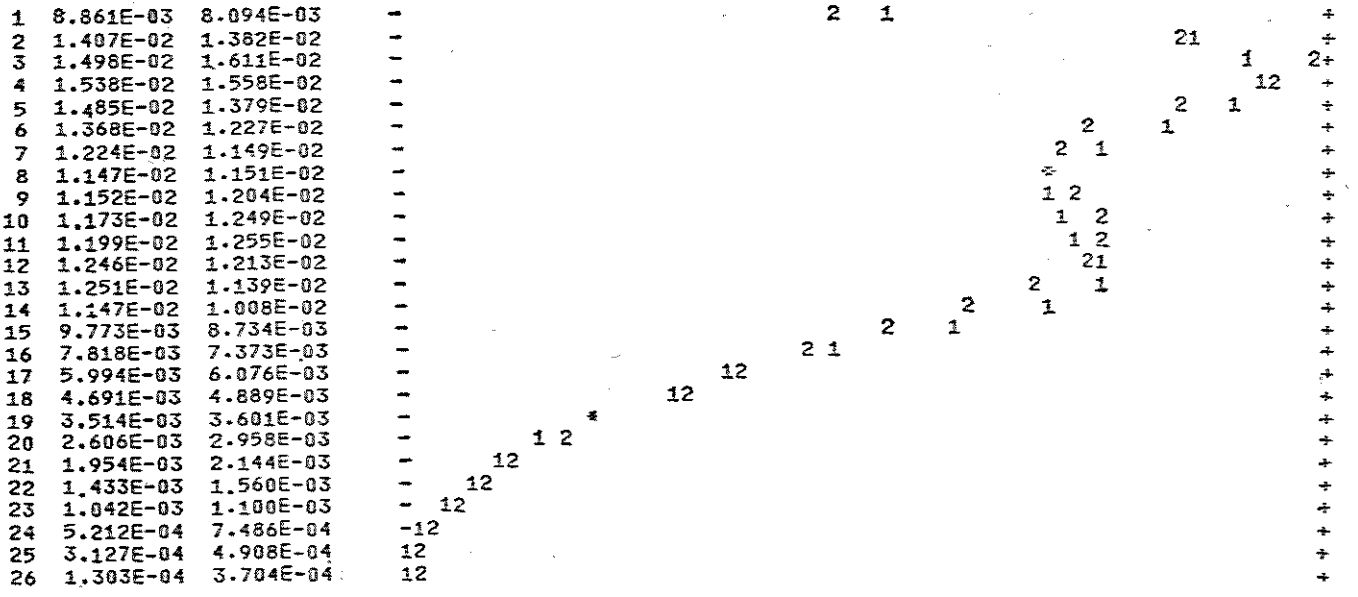
10.4.3. Адекватность модели

Последний этап обработки результатов эксперимента состоит в проверке адекватности модели, или, говоря иначе, в установлении качества полученного приближения. Важнейшие положения этого этапа были рассмотрены при анализе результатов примера в предыдущем разделе. Здесь мы приведем лишь некоторые дополнительные соображения.

В некоторых случаях (к сожалению, не в большинстве) характеристики экспериментальных ошибок s_{yl} и s_{xl}^{j} известны с достаточно большой достоверностью. Для таких случаев разработан тест проверки адекватности, основанный на величине остаточной суммы квадратов S. Сумма S имеет χ^2-распределение [1], которое зависит только от числа степеней свободы.

[1] Величина S здесь является взвешенной суммой квадратов остаточных отклонений (см. разд. 10.2.1) с числом степеней свободы $N-p$, где N — число опытов, p — число оцениваемых параметров. Для случая, когда в S не входят веса наблюдений [см., например, (3.89)], χ^2-распределению подчинена величина S/σ^2, где σ^2 — дисперсия ошибок наблюдений. — *Прим. ред.*

GEMESSENE UND BERECHNETE WERTE



Фиг. 10.12. Результаты примера структурного анализа (III).

```
C     PROGRAM PREDICTION ANALYSIS
CC
C            NACH  JOHN  R.  H O L B E R G
      SUBROUTINE PREDAN(X,NN,YM,YSIG,LSA)
      COMMON/PRED/A,DELA,DELX
C
      DIMENSION X(5,50),XF(5,50),C(15,15),CIV(15,15),FX(5),FA(15),
     *XX(5),SY(50),SX(5,50),A(15),DELA(15),DELX(15),SXK(5),
     *TB(19,8),DX(7),DXR(7),DXMAX(7),DXRMAX(7),DDR(7),YM(10),YSIG(1),
     *IDS(7)
      DATA TU/
     *63.657,9.925,5.841,4.604,4.032,3.707,3.499,3.355,
     *3.250,3.169,2.947,2.845,2.787,2.750,2.576,
     *31.821,6.965,4.541,3.747,3.365,3.143,2.998,2.896,
     *2.821,2.764,2.602,2.528,2.485,2.457,2.326,
     *12.706,4.303,3.182,2.776,2.571,2.447,2.365,2.306,
     *2.262,2.228,2.131,2.086,2.060,2.042,1.960,
     *6.314,2.920,2.353,2.132,2.015,1.943,1.895,1.860,
     *1.833,1.812,1.753,1.725,1.708,1.697,1.645,
     *3.078,1.886,1.638,1.533,1.476,1.440,1.415,1.397,
     *1.383,1.372,1.341,1.325,1.316,1.310,1.282,
     *1.963,1.386,1.250,1.190,1.156,1.134,1.119,1.108,
     *1.100,1.093,1.074,1.064,1.058,1.055,1.036,
     *1.376,1.061,.978,.941,.920,.906,.896,.889,
     *.883,.879,.866,.860,.856,.854,.842,
     *1.00,.816,.765,.741,.727,.718,.711,.706,
     *.703,.700,.691,.687,.684,.683,.674/
C
C     AKTUELLE MAXIMALE DIMENSIONIERUNG
CC
CC          5 UNABHAENGIGE VARIABLE
CC
CC         15 PARAMETER
CC
CC         50 MESSWERTE
C
      IN=3
      IO=2
C
      READ (IN,1) NP,M,IK,NFK,IDR
    1 FORMAT(10I5)
C
C     NP = ANZAHL DER PARAMETER
CC
CC     M = ANZAHL DER UNABHAENGIGEN VERAENDERLICHEN
CC
CC     IK = STEUERPARAMETER FUER STREUUNG (1,2,3)
CC
CC    NFK = KONFIDENZBAENDER WERDEN FUER NFK EINZULESENDE WERTE
CC          DER UNABHAENGIGEN VARIABLEN  XF  BERECHNET
CC
CC     IDR = FUER IDR   .   WERDEN KONFIDENZBAENDER FUER JEDEN EIN-
CC          ZELNEN WE   DER UNABHAENGIGEN VERAENDERLICHEN AUSGEGEBEN
CC
CC          FUER IDR.EQ.0  WERDEN NUR MITTEL- UND MAXIMALWERTE UEBER
CC          GANZEN BEREICH AUSGEGEBEN
C
      READ (IN,2) (A(I),I=1,NP)
    2 FORMAT(8E0.0)
C
CC     A = LEAST SQUARES PARAMETERS
C
      READ (IN,2) (DELA(I),I=1,NP)
      READ (IN,2) (DELX(I),I=1,M)
C
CC     DELA = DELX  = ZUWAECHSE BEI ABLEITUNGSBILDUNG
C
```

Фиг. 10.12 (продолжение).

```
      READ (1N,2) SYK,(SXK(I),I=1,M)

      SYK = SXK  = PARAMETER FUER FEHLERVERTEILUNG

      IF(NFK)10,15,10
   10 DO 12 J=1,NFK
   12 READ (1N,2) (XF(I,J),I=1,M)
   15 CONTINUE
      DO14I=1,NP
      DO14J=1,NP
   14 C(I,J)=.0
      IF(NN)99,99,20
   20 CALL HELP
      NF=NN-NP
      IF(NF-10)16,16,17
   16 NI=NF
      GO TO 21
   17 IF(NF-35)19,18,18
   18 NI=15
      GO TO 21
   19 NI=(NF-10)/5+10
   21 CONTINUE
      WRITE (10,3) M,NP,NN
    3 FORMAT(1H1,21H PREDICTION ANALYSIS  ///
     *I10,23H UNABHAENGIGE VARIABLE  //
     *I10,11H PARAMETER   //
     *I10,32H SAETZE UNABHAENGIGER VARIABLER   //)
      WRITE (10,4) ((X(I,J),I=1,M),J=1,NN)
    4 FORMAT(29H UNABHAENGIGE VERAENDERLICHE //(10(1PE12.4)))
      WRITE (10,6) (A(I),I=1,NP)
    6 FORMAT(/11H PARAMETER //(10(1PE12.4)))
      WRITE (10,7) IK,SYK,(SXK(I),I=1,M)
    7 FORMAT(/31H VERTEILUNG DER MESSFEHLER -IK=,I2//
     *10X,5H SYK=,1PE10.4,5H SXK=,3(1PE10.4))
      IF(IK1)606,608,606
  606 WRITE (2,607) (YM(I),I=1,NN)
  607 FORMAT(/35H STREUUNG AUS MESSWERTEN BERECHNET //
     *11H MESSWERTE //(10(1PE12.4)))
  608 WRITE (2,8) (DELA(I),I=1,NP)
    8 FORMAT(/42H PARAMETERZUWAECHSE BEI ABLEITUNGSBILDUNG
     *//(10(1PE12.4)))
      WRITE (2,9) (DELX(I),I=1,M)
    9 FORMAT(/37H ZUWACHS UNABHAENGIGER VARIABLER BEI
     *10HABLEITUNGSBILDUNG //(10(1PE12.4)))
      S=.0
      DO 40 I=1,NN
      DO 775 J=1,M
  775 XX(J)=X(J,I)
      CALL F(XX,Y,A,I,M,NN)
      YH=Y
      CALL DFDX(XX,Y,A,FX,I,M,NN)
      Y=YH
      CALL DFDA(XX,Y,A,FA,I,M,NN)
      IF(LSA-2)26,780,25
  780 AL=YSIG(I)**2
      GO TO 30
   25 AL=ABS(YH-YM(I))**2
      GO TO 30
   26 CONTINUE
      IF(IK-2)27,28,29
   27 AL=SYK**2
      GO TO 30
   28 AL=(SYK*YH)**2
      GO TO 30
   29 AL=SYK**2*ABS(YH)
   30 DO 35 J=1,M
      IF(IK-2)31,32,33
   31 AL=AL+(FX(J)*SXK(J))**2
```

Фиг. 10.12 (продолжение).

```
      GO TO 35
   32 AL=AL+(FX(J)*SXK(J)*X(J,I))**2
      GO TO 35
   33 AL=AL+(FX(J)*SXK(J))**2*ABS(X(J,I))
   35 CONTINUE
      IF(LSA.GT.0) S=S+(YM(I)-YH)**2/AL
      DO 40 J=1,NP
      DO 40 K=1,NP
   40 C(J,K)=C(J,K)+FA(J)*FA(K)/AL
      DO 45 I=1,NP
      DO 45 J=1,NP
   45 CIV(I,J)=.0
      DO 46 I=1,NP
   46 CIV(I,I)=1.
      IF(LSA)785,785,783
  783 WRITE (2,784) S
  784 FORMAT(//20H FEHLERQUADRATSUMME  ,1PE15.6)
  785 CONTINUE
      WRITE (2,48) ((C(I,J),J=1,NP),I=1,NP)
   48 FORMAT(8E15.5)
      CALL AFGASY(C,CIV,15,15,NP,NP,2,1)
      WRITE (2,48) ((CIV(I,J),J=1,NP),I=1,NP)
      WRITE (10,47)
   47 FORMAT(//10H PARAMETER,8X,3H99%,12X,3H98%,12X,3H95%,12X,3H90%,12X,
     *3H80%,.
     *12X,3H70%,12X,3H60%)
      DO 53 J=1,NP
      SIGA=CIV(J,J)
      IF(LSA)782,782,781
  781 SIGA=SIGA*S/NI
  782 SIGA=SQRT(SIGA)
      DO 52 I=1,7
      IF(NI)480,480,49
  480 DX(I)=SIGA
      GO TO 52
   49 CONTINUE
      DX(I)=SIGA*TB(NI,I)
   52 DXR(I)=DX(I)/ABS(A(J))*100.
  601 FORMAT(//)
   53 WRITE (10,54) A(J),(DX(I),I=1,7)
   54 FORMAT(/1X,1PE10.4,2X,7(1X,1PE13.6,1X))
      DO 60 I=1,7
      DX(I)=.0
       DXR(I)=.0
      DXMAX(I)=.0
   60 DXRMAX(I)=.0
      WRITE (2,601)
      IF(LSA)790,790,788
  788 WRITE (2,789)
  789 FORMAT(/46H GUETE DES ANGLEICHS DER LEAST-SQUARE-ANALYSE //
     *10X,43H * = MESSWERT AUSSERHALB DES ENTSPRECHENDEN,
     *17H KONFIDENZBANDES //
     *3X,8HMESSWERT,7X,8HBER.WERT,4X,12HSTANDARDABW.,2X,8HABS.ABW.,
     *3X,8HREL.ABW.,12H  SIGN,ABW. ,5H 99% ,5H 98% ,5H 95% ,
     *5H 90% ,5H 80% ,5H 70% ,5H 60%  /)
      GO TO 604
  790 CONTINUE
      IF(IDR)603,604,603
  603 WRITE (2,602)
  602 FORMAT(/18H KONFIDENZBAENDER
     */4X,1HY,9X,3H90%,7X,3H95%,7X,3H80%,8X,1HX)
  604 CONTINUE
      IF(NFK)61,62,61
   61 NN=NFK
   62 DO 83 I=1,NN
      IF(NEK)65,63,65
   63 DO 64 J=1,M
   64 XX(J)=X(J,I)
```

Фиг. 10.12 (продолжение).

```
      GO TO 70
   65 DO 66 J=1,M
   66 XX(J)=XF(J,I)
   70 CALL F(XX,Y,A,I,M,NN)
      YH=Y
      CALL DFDA(XX,Y,A,FA,I,M,NN)
      SIGQ=.0
      DO 72 L=1,NP
      DO 72 K=1,NP
   72 SIGQ=SIGQ+FA(L)*FA(K)*CIV(L,K)
      IF(LSA.GT.0) SIGQ=SIGQ*S/NI
      SIGQ=SQRT(SIGQ)
      DO 82 J=1,7
      IF(NI)73,73,74
   73 SIGJ=SIGQ
      GO TO 750
   74 CONTINUE
      SIGJ=SIGQ*TB(NI,J)
  750 CONTINUE
      DDR(J)=SIGJ
   76 FORMAT(1PE10.3,3(1X,1PE9.1),3X,5(1PE10.3,3X))
   77 DX(J)=DX(J)+SIGJ
      DXR(J)=DXR(J)+SIGJ/ABS(YH)*100.
      IF(SIGJ-DXMAX(J))80,80,78
   78 DXMAX(J)=SIGJ
   80 YREL=SIGJ/ABS(YH)*100.
      IF(YREL-DXRMAX(J))82,82,81
   81 DXRMAX(J)=YREL
   82 CONTINUE
      IF(LSA)801,801,791
  791 ISI=4H-
      IF(YM(I).GT.YH) ISI=4H
      DO 792 J=1,7
      IDS(J)=1H*
      IF(YH-DDR(J).LT.YM(I).AND.YM(I).LT.YH+DDR(J)) IDS(J)=1H
  792 CONTINUE
      IF(LSA-2)795,793,794
  793 AL=YSIG(I)**2
      GO TO 799
  794 AL=ABS(YM(I)-YH)**2
      GO TO 799
  795 IF(IK-2)796,797,798
  796 AL=SYK**2
      GO TO 799
  797 AL=(SYK*YH)**2
      GO TO 799
  798 AL=(SYK)**2*ABS(YH)
  799 YA=YM(I)-YH
      YAR=YA/YM(I)
      WRITE (2,800) YM(I),YH,AL,YA,YAR,ISI,IDS
  800 FORMAT(2(1PE13.5,2X),1PE10.3,2X,2(1PE10.2,1X),4X,
     *A4,4X,7(2X,A1,2X))
      GO TO 83
  801 CONTINUE
      IF(IDR)75,83,75
   75 WRITE (10,76) YH,DDR(1),DDR(3),DDR(5),(XX(K),K=1,M)
   83 CONTINUE
      DO 85 J=1,7
      DX(J)=DX(J)/NN
   85 DXR(J)=DXR(J)/NN
      IF(NFK)88,86,88
   86 WRITE (10,87)
   87 FORMAT(//44H KONFIDENZBAENDER ( BEZOGEN AUF MESSWERTE ) )
      GO TO 90
   88 WRITE (10,89)
   89 FORMAT(//48H KONFIDENZBAENDER ( FUER VORGEGEBENE XF-WERTE ) )
   90 WRITE (10,91)
   91 FORMAT(26X,3H99%,11X,3H98%,11X,3H95%,11X,
```

Фиг. 10.12 (продолжение).

```
     *3H90%,11X,3H80%,11X,3H70%,11X,3H60%)
      WRITE (IO,92) DX
   92 FORMAT(18H MITTL.ABS.FEHLER ,2X,7(2X,1PE10.4,2X))
      WRITE (IO,93) DXR
   93 FORMAT(18H MITTL.REL.FEHLER ,2X,7(4X,F6.2,1H%,3X)
      WRITE (IO,94) DXMAX
   94 FORMAT(2X,16H MAX.ABS.FEHLER ,2X,7(2X,1PE10.4,2X)
      WRITE (IO,95) DXRMAX
   95 FORMAT(2X,16H MAX.REL.FEHLER ,2X,7(4X,F6.2,1H%,3X)
   99 CONTINUE
      RETURN
      END
```

Фиг. 10.12 (продолжение).

Распределение χ^2 приведено в табл. 15.2. Покажем на примере, как надо применять это распределение. Спрашивается, например, какое значение S превышается в 5% всех случаев при числе степеней свободы, равном 15. Из табл. 15.2 находим значение $\chi^2_{кр}$, равное 24,996. Другими словами, это означает, что при выбранном уровне значимости 0,05 и числе степеней свободы 15 рассчитанное значение должно быть не больше 24,996. Если получено существенно большее значение, то это может означать наличие ошибки. Во-первых, она может заключаться в пере- или недооценке точности эксперимента. Во-вторых, причиной могут быть отдельные ошибочные измерения. Наконец, в-третьих, этой ошибкой может быть неудачный выбор модели, т. е. ее неадекватность.

Важной проверкой качества достигнутого приближения является его сравнение с теоретическими или другими экспериментальными результатами. Часто из теории бывает известно, в каких границах должны находиться параметры. Если эти границы нарушаются, причиной этого может оказаться плохой выбор функции $f(x, a)$.

Ошибка может быть сделана и в вычислительной программе, правда, только в период ее освоения. Поэтому предварительно необходимо произвести ее тщательную проверку на задачах с известными результатами.

Два важнейших способа проверки адекватности уже были рассмотрены в предыдущем разделе: отсутствие систематической составляющей в остаточных отклонениях и доверительные интервалы в каждой точке. Если хотя бы один из перечисленных тестов дает отрицательный ответ, необходимо произвести тщательный анализ, чтобы выяснить причину такого результата. Тесты в принципе дают лишь некоторые указания, но сами по себе еще ничего не решают. Решение может приниматься только человеком, и оно в значительной степени зависит от его опыта, знаний и целей.

10.4.4. Устранение систематических ошибок

Как метод наименьших квадратов, так и описанный метод планирования эксперимента основываются на предположении о статистической природе ошибок измерения независимых и зависимых переменных. Говоря точнее, предполагается, что при усреднении по очень большому числу экспериментов получаются истинные значения $\bar{x}$ и $\bar{y}$; точно так же считается, что при статистической обработке очень большого числа экспериментов оценки параметров a находятся сколь угодно близко к истинным значениям $\bar{a}$.

Если при проведении эксперимента возникает систематическая ошибка, то истинные значения параметров получить не удается. Различают два вида систематических ошибок:

1. Ошибки, из-за которых измеренные значения переменных систематически отклоняются от их истинных значений. В качестве примера можно привести измерение температуры. Точность измерений составляет $\pm$ 1°C. Однако при настройке измерительный прибор был установлен неправильно, так что все измерения температур оказываются завышенными на 5°C.

2. Ошибки, из-за которых оценки параметров отклоняются от их истинных значений даже в случае, когда средние измеренных значений сходятся к их истинным значениям. Примером может служить статистическая обработка результатов эксперимента, если модель имеет вид

$$y = a_1 \exp(bx) + a_2.$$

Необходимо определить параметры a_1 и a_2. Константа b предполагается известной. Однако если значение b взято неверно, то даже при точных измерениях нельзя получить истинные значения параметров. Если имеются систематические ошибки, равенство (10.16) оказывается несправедливым, поскольку при его выводе ошибки в различных точках измерения предполагались некоррелированными.

Покажем, как воздействуют систематические ошибки на величину $\sigma^2_{a_k}$. Предположим, что имеется q различных источников систематических ошибок. Пусть $e^l_{a_k}$ — неточность параметров a_k, обусловленная l-й систематической ошибкой.

Из закона переноса ошибок следует

$$\sigma^2_{a_k} = E\{s^2_{a_k}\} + \sum_{l=1}^{q} E\{(e^l_{a_k})^2\}.$$

Остался открытым вопрос о способе вычисления $e^l_{a_k}$:

$$e^l_{a_k} \approx \frac{\partial a_k}{\partial (l\text{-й источник ошибок})} \times \text{Оценка величины } l\text{-й ошибки.}$$

В примере с ошибочно заданной константой b получаем следующее выражение для $e^i_{a_k}$:

$$e^i_{a_k} \approx \frac{\partial a_k}{\partial b} \Delta b.$$

Из этого примера видно, что в общем случае для нахождения $e^i_{a_k}$ требуется численный расчет соответствующих производных. Выполнить его оказывается не слишком сложно, так как, за исключением простейших случаев, расчеты все равно производятся на вычислительной машине. Заметим, что для получения производных необходимо повторно применить метод наименьших квадратов. Для этого можно слегка изменить данные, соответствующие рассматриваемому источнику ошибок, и заново их обработать. На эту обработку уходит небольшое вычислительное время, поскольку имеются очень хорошие начальные значения для параметров. В результате получаются значения параметров, отличающиеся от найденных первоначально на величину Δa_k. Отсюда находим производные

$$\frac{\partial a_k}{\partial (i\text{-й источник ошибок})} \approx \frac{\Delta a_k}{\text{Изменение } i\text{-го источника ошибок}}$$

10.4.5. Чебышевская аппроксимация

В разд. 10.1 уже отмечалось, что вместо метода наименьших квадратов для обработки результатов эксперимента может применяться, например, чебышевская аппроксимация. Мы собираемся кратко описать этот метод, его достоинства и недостатки.

Как и раньше, задаются измеренные значения $\tilde{y}^i$ $(i = 1, N)$, веса g_i $(i = 1, N)$ и вычисленные значения $\hat{y}^i = f(\tilde{x}^i, \hat{a})$. В то время как оценки параметров a при использовании метода наименьших квадратов определяются из условия

$$\min_a \sum_{i=1}^{N} g_i (\tilde{y}^i - y^i)^2,$$

в чебышевской аппроксимации требуется обеспечить

$$\min_a \max_{1 \leqslant i \leqslant N} |g_i (\tilde{y}^i - y^i)|.$$

Таким образом, параметры должны рассчитываться так, чтобы минимизировать максимум взвешенного модуля разности между измеренным и вычисленным значениями.

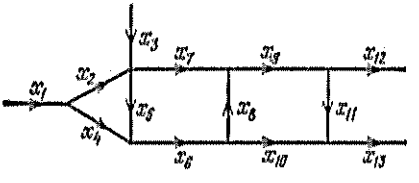
Совершенно очевидны два недостатка чебышевской аппроксимации. Первый из них вполне понятен: если велики дисперсии отдельных измеренных значений, то результаты окажутся неверными. В то время как метод наименьших квадратов из-за усреднения

локализует последствия отдельных грубых ошибок, результаты чебышевской аппроксимации могут быть искажены даже одним ошибочным измерением. Второй недостаток связан с вычислительными аспектами решения так называемой минимаксной задачи. Для решения задачи методом наименьших квадратов разработаны хорошие алгоритмы, чего нельзя сказать о чебышевской аппроксимации, по крайней мере в общей постановке.

Если функция $f(x, a)$ линейна, чебышевская аппроксимация превращается в линейную оптимизационную задачу. Такой пример будет приведен в следующем разделе. Если же $f(x, a)$ нелинейна по a, то применяются другие алгоритмы. Можно применить, например, следующую итерационную процедуру. Разлагают f в ряд Тейлора по a и ограничиваются линейными членами. Пусть имеется начальное приближение для параметров a. Введем приращение параметров Δa. Для этих приращений чебышевская аппроксимация приводит к линейной задаче оптимизации. Для ее решения необходимо задать верхнее и нижнее ограничения на Δa. Это нужно прежде всего потому, что линейное приближение справедливо лишь в некоторой ограниченной области. После решения линейной оптимизационной задачи получается значение целевой функции, т. е. максимальное взвешенное отклонение для $a + \Delta a$. Это значение сравнивается с целевой функцией, рассчитанной для a, и если последняя окажется меньше, необходимо уменьшить ограничения и вновь найти экстремум. В противном случае a заменяется на $a + \Delta a$ и вся процедура повторяется и т. д. Естественно, должно быть предусмотрено правило останова.

10.4.6. Задача балансировки

Задача балансировки заключается в следующем. Задается некоторая сеть (фиг. 10.13). Направленные линии называются балансовыми линиями, а точки разветвлений — узлами. Каждой



Фиг. 10.13. Балансируемая сеть для контрольного примера.

линии соответствует измеренное значение $\tilde{x}$. Мера точности измерения $\tilde{x}_i$ обозначается p_i. Если бы все измеренные значения были точными, баланс должен был бы выполняться в каждом

узле. Для реальных измеренных значений это не так. Задача балансировки состоит в вычислении таких значений x_i, которые точно обеспечивают балансы и по возможности хорошо соответствуют измеренным значениям. Иногда выдвигают еще требование, чтобы x_i были неотрицательными. Для LN линий и KN узлов справедливы условия

$$\sum_{j=1}^{LN} A_{ij}x_j = 0, \quad i = 1, \ KN,$$
$$x_j \geqslant 0, \quad j = 1, \ LN. \tag{10.25}$$

В качестве критерия обычно выбирается минимум среднего квадрата отклонения измеренного значения от вычисленного:

$$\min_x \left\{ Q_1 = \sum_{i=1}^{LN} \frac{(\tilde{x}_i - x_i)^2}{p_i} \right\}.$$

Можно применить также чебышевскую аппроксимацию

$$\min_x \max_{1 \leqslant i \leqslant LN} \left| \frac{\tilde{x}_i}{\sqrt{p_i}} - \frac{x_i}{\sqrt{p_i}} \right|.$$

Здесь x_i выступают в роли оцениваемых параметров, а независимых переменных нет вообще.

Для решения минимаксной задачи введем функцию $\eta_i(x)$, которую определим следующим образом:

$$\eta_i(x) = \frac{x_i}{\sqrt{p_i}} - \frac{\tilde{x}_i}{\sqrt{p_i}}. \tag{10.26}$$

Абсолютные величины $\eta_i(x)$ должны быть сделаны по возможности малыми. Введем $(LN+1)$-й параметр x_{LN+1}, который требуется минимизировать:

$$\min_x x_{LN+1} \tag{10.27}$$

при условиях

$$|\eta_i(x)| \leqslant x_{LN+1} \quad (i = 1, \ LN).$$

Эти условия могут быть преобразованы в форму

$$\left.\begin{aligned} -\eta_i(x) + x_{LN+1} \geqslant 0, \\ \eta_i(x) + x_{LN+1} \geqslant 0 \end{aligned}\right\} \quad i = 1, \ LN. \tag{10.28}$$

Таким образом, в линейном случае чебышевская аппроксимация приводит к линейной экстремальной задаче, определяемой с помощью уравнений (10.25)—(10.28). Поскольку на любой большой ЭВМ имеются программы линейного программирования, применение чебышевской аппроксимации в линейных задачах не представляет трудностей.

Таблица 10.2

Результаты решения задачи балансировки двумя методами: G — методом наименьших квадратов и T — методом чебышевской аппроксимации

Измерение 1			Измерение 2			Измерение 3		
x	результаты		x	результаты		x	результаты	
	G	T		G	T		G	T
28,0	13,32	17,5	24,0	20,61	21,5	20,8	20,84	20,84
1,0	1,36	0,0	10,0	11,84	12,5	9,0	9,46	9,42
18,0	3,69	7,5	11,0	9,45	8,5	7,5	7,99	7,92
8,0	11,96	17,5	8,5	8,77	9,0	11,0	11,38	11,42
0,5	4,09	7,5	9,0	7,42	7,5	4,9	4,82	4,66
16,0	16,05	25,0	14,0	16,19	16,5	16,5	16,20	16,08
1,0	0,96	0,0	14,5	13,87	13,5	12,8	12,63	12,68
14,0	10,31	12,5	15,0	13,75	13,0	7,8	7,8	7,8
2,0	11,26	12,5	24,0	27,62	26,5	20,9	20,43	20,48
2,0	5,75	12,5	1,0	2,44	3,5	8,7	8,4	8,28
5,0	3,17	0,0	28,0	27,06	26,5	2,0	1,97	1,76
3,0	8,09	12,5	2,0	0,56	0,0	18,3	18,45	18,72
2,0	8,92	12,5	30,0	29,5	30,0	10,2	10,38	10,04

Таблица 10.3

Характеристики результатов балансировки

	Измерения 1		Измерения 2		Измерения 3	
	G	T	G	T	G	T
$\sum_i (x_i - \tilde{x}_i)^2 / p_i$	639,6	891	44,99	51,25	1,13	1,43
$\max_i \lvert (x_i - \tilde{x}_i) / \sqrt{p_i} \rvert$	14,68	10,5	3,62	2,5	0,49	0,34

Интересно сравнить на одной и той же задаче результаты применения метода наименьших квадратов и чебышевской аппроксимации. Для этого был рассмотрен пример, показанный на фиг. 10.13. Расчеты проводились для трех вариантов задачи. Первая совокупность измерений привела к совершенно неверным балансам, для второй совокупности величины отклонений оказались еще довольно значительными, в то время как третья совокупность содержала измерения с „приемлемыми" ошибками. Результаты представлены в табл. 10.2. Для сравнения в табл. 10.3 приведены значения целевых функций для обоих методов. Для наглядности результатов меры точности измерения везде задавались равными $p_i \equiv 1$.

Различия в результатах очевидны. Они увеличиваются по мере ухудшения измеренных значений. Принять решение о том, какие результаты следует применять, оказывается весьма затруднительным: отчасти такое решение является произвольным. С другой стороны, из самой постановки задачи должно становиться ясным, следует ли производить усреднение, несмотря на отдельные большие отклонения измеренных значений от вычисленных, или же их следует исключить. Если на эти вопросы нельзя дать четкого ответа, нужно выбрать метод в зависимости от имеющихся в распоряжении программ. Если имеется только одна такая программа, то используют ее; в противном случае выбирают ту программу, которая требует наименьших затрат вычислительного времени.

Словарь терминов и сокращений к фиг. 10.8, 10.9 и 10.12

Abs. = absolut — абсолютный.

Abw. = Abweichung — отклонение.

Abs. Abw. = absolute Abweichung — абсолютное отклонение.

Rel. Abw. = relative Abweichung — относительное отклонение.

Sign. Abw. = Vorzeichen der Abweichung — знак отклонения.

Standardabw. = Standardabweichung — стандарт, среднеквадратичное отклонение.

Dimensionierung — размерность.

Fehler — ошибка.

Max. abs. Fehler — наибольшая абсолютная ошибка.

Max. rel. Fehler — наибольшая относительная ошибка.

Messfehler — измерительная ошибка.

Mittl. abs. Fehler — средняя абсолютная ошибка.

Mittl. rel. Fehler — средняя относительная ошибка.

Fehlerquadratsumme — сумма квадратов ошибок.

Konfidenzbaender — доверительные границы.

Steuerparameter — константа управления.

Streuung — рассеяние, дисперсия.

Variable — переменная.

Veraenderliche, Variable — переменная.

Unabhaengige Variable — независимая переменная.

Saetze unabhaengiger Variabler — векторы независимых переменных.

Verteilung — распределение.
Wert — значение.
ber. Wert = berechneter Wert — рассчитанное значение.
gemessener Wert = Messwert — измеренное значение.
Schaetzwert — оценка, оценочное значение.

ЛИТЕРАТУРА

1. Wolberg J. R., Prediction Analysis, van Nostrand, Princeton, 1967.
2. Marquardt D. W., *J. SIAM Control*, **11**, 431 (1963).
3. Box G. E. P., Lucas H. L., *Biometrika*, **46**, 77 (1959).
4. Powell M. J. D., *Computer J.*, **7**, 155 (1964).
5. Lee E. S., *IEC Fundamentals*, **7**, 152 (1968).
6. Fletcher R., Powell M. J. D., *Computer J.*, **6**, 163 (1963).
7. Powell M. J. D., *Computer J.*, **7**, 303 (1965).
8. Kitrell J. R., Mezaki R., Watson C. C., *British Chem. Eng.*, **11**, 15 (1966).
9. Hslang Th., Reilly P. M., *Canadian J. Chem. Eng.*, **49**, 865 (1971).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Ермаков С. М., Мелас В. Б., Об одном подходе к планированию эксперимента при нелинейной параметризации, *Заводск. лаб.*, № 10 (1973).
2. Федоров В. В., Теория оптимального эксперимента, изд-во „Наука", М., 1971.
3. Планирование и автоматизация эксперимента в научных исследованиях, сб. под редакцией Г. К. Круга, изд-во „Советское радио", М., 1974.
4. Горский В. Г., Статистические задачи при обработке кинетических экспериментов, см. сб. „Вопросы кибернетики. Некоторые вопросы планирования эксперимента", Научный совет по комплексной проблеме „Кибернетика", М., 1972.
5. Писаренко В. Н., Погорелов А. Г., Планирование кинетических исследований, изд-во „Наука", М., 1969.
6. Himmelblau D., Process Analysis by Statistical Methods, Wiley, New York, 1970; русский перевод: Химмельблау Д., Анализ процессов статистическими методами, изд-во „Мир", М., 1973.
7. Himmelblau D., Applied Nonlinear Programming, McGraw-Hill, New York, 1972; русский перевод: Химмельблау Д., Прикладное нелинейное программирование, изд-во „Мир", М., 1975.
8. Клепиков Н. П., Соколов С. Н., Анализ и планирование экспериментов методом максимума правдоподобия, Физматгиз, М., 1964.
9. Островский Г. М., Волин Ю. М., Методы оптимизации химических реакторов, изд-во „Химия", М., 1967.
10. Бояринов А. И., Кафаров В. В., Методы оптимизации в химической технологии, изд-во „Химия", М., 1975.

11.

ПОСЛЕДОВАТЕЛЬНЫЕ МЕТОДЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

11.1. Последовательные методы в экспериментальных исследованиях

При проведении экспериментальных исследований часто оказывается полезным производить обработку данных по мере их поступления. Если в ходе эксперимента в некоторый момент времени i исследователь располагает результатами обработки уже полученных экспериментальных данных, то он может использовать их для выработки предварительных рекомендаций по совершенствованию управления объектом, для проверки правильности исходных предпосылок проводимого эксперимента, наилучшей организации последующих опытов и т. п. [1, 2].

Пусть, например, с помощью эксперимента строится математическое описание некоторого объекта, причем по мере поступления данных пересчитываются оценки коэффициентов $\hat{a}$, ковариационная матрица оценок $\mathrm{cov}\{\hat{a}\}$ и остаточная дисперсия s^2_R. Тогда, располагая в некоторый момент i перечисленными величинами, рассчитанными по имеющимся к этому моменту данным, можно проверить правильность принятого предположения о виде модели и в случае необходимости изменить вид модели. Кроме того, можно оценить точность оценок коэффициентов, достигнутую с помощью уже проделанных опытов, и прекратить дальнейшие эксперименты, если требуемое качество математической модели обеспечено. Тем самым появляется возможность сократить время и стоимость проводимого исследования.

Ниже мы везде будем предполагать, что эксперимент носит дискретный характер, т. е. что опыты осуществляются в некоторые дискретные моменты i, причем величина i принимает значения $i = 1, 2, \ldots$. Уточнение интересующих экспериментатора величин (таких, например, как оценок коэффициентов, их ковариационной матрицы и остаточной дисперсии) осуществляется после всякого опыта, или, в общем случае, через каждые m опытов, где $m = 1, 2, \ldots$. Такой подход к построению математической модели приводит к ряду особенностей вычислительной процедуры и процедуры планирования эксперимента. Метод решения задачи построения математического описания при этом называют последовательным.

Вычислительные алгоритмы при последовательном способе построения модели целесообразно строить по рекуррентной форме.

При этом значения оценок в момент $(i+m)$ вычисляются с помощью выражения, в которое входят значения оценок в момент времени i и данные, поступившие в моменты $(i+1)$, $(i+2)$, ..., $(i+m)$. Такая форма вычисления часто оказывается удобной как с точки зрения экономии памяти вычислительной машины (так как вместо всех предыдущих данных сохраняются лишь результаты их обработки), так и с точки зрения времени вычислений (или количества вычислительных операций).

При применении последовательного способа решения задачи исследователь во всякий момент i попадает в ситуацию, когда ему оказываются известными некоторые априорные сведения об оцениваемых на основе экспериментальных данных параметрах. Эти априорные сведения получаются путем обработки результатов ранее проведенных опытов, причем к этим опытам могут быть отнесены и эксперименты других исследователей. Очевидно, что при проведении данного исследования привлечение этих априорных сведений может оказаться полезным. В общем случае рекуррентное построение вычислительного алгоритма „увязывает" априорные сведения с результатами последующих экспериментов.

При последовательном способе построения модели задача планирования последующих экспериментов также должна решаться с учетом имеющейся априорной информации. Такая постановка задачи отличается от традиционной постановки задачи планирования.

В этом разделе рассматривается последовательная методика решения задачи построения математической модели объекта в постановке, изложенной выше в разд. 3. В разд. 11.2 описаны вычислительные алгоритмы, используемые при последовательном построении моделей. В разд. 11.3 рассмотрена задача последовательного планирования эксперимента. В разд. 11.4 излагается правило принятия решений об остановке эксперимента при достижении требуемой точности математической модели. Наконец, в разд. 11.5 последовательное планирование применено к задаче построения моделей, нелинейных по параметрам.

11.2. Рекуррентные алгоритмы вычисления

В этом разделе будут получены рекуррентные алгоритмы вычисления оценок коэффициентов уравнения регрессии, их ковариационной матрицы и остаточной суммы квадратов для случая, когда поиск оценок осуществляется на основе метода наименьших квадратов.

11.2.1. Постановка задачи

Здесь рассматривается та же задача, что и в разд. 3.2. Предполагается, что модель исследуемого объекта имеет вид

$$y(x, a) = \sum_{l=0}^{k} a_l f_l(x) = a' f(x). \tag{11.1}$$

Обозначения в (11.1) те же, что и в (3.74).

Пусть на основе выборки объема N методом наименьших квадратов найдены оценки $\hat{a}^N$, а также рассчитаны ковариационная матрица оценок (или пропорциональная ей дисперсионная матрица плана C_N) и остаточная сумма квадратов S_N. Выражения для расчета этих величин отличаются от приведенных в разд. 3 только индексом N, определяющим число реализованных опытов:

$$\hat{a}^N = (F_N' F_N)^{-1} F_N' \tilde{Y}_N = C_N F_N' \tilde{Y}_N, \tag{11.2}$$

$$C_N = (F_N' F_N)^{-1}, \tag{11.3}$$

$$S_N = (\tilde{Y}_N - F_N \hat{a}^N)' (\tilde{Y}_N - F_N \hat{a}^N) = \tilde{Y}_N' \tilde{Y}_N - (\hat{a}^N)' F_N' \tilde{Y}_N. \tag{11.4}$$

В отличие от разд. 3 здесь с целью упрощения опущен индекс R у остаточной суммы квадратов. Выражение (11.4) легко получить из (3.89), если вместо $\tilde{Y}_N$ подставить $F_N \hat{a}^N$. Матрица F_N в выражениях (11.2)—(11.4) имеет вид [см. (3.78)]

$$F_N = \begin{bmatrix} f_0(x^1) & \dots & f_k(x^1) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ f_0(x^N) & \dots & f_k(x^N) \end{bmatrix}. \tag{11.5}$$

Добавим теперь к выборке объема N еще m экспериментов, определяемых матрицами

$$X_m = \begin{bmatrix} x_1^{N+1} & x_2^{N+1} & \dots & x_n^{N+1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_1^{N+m} & x_2^{N+m} & \dots & x_n^{N+m} \end{bmatrix}, \tag{11.6}$$

$$F_m = \begin{bmatrix} f_0(x^{N+1}) & f_1(x^{N+1}) & \dots & f_k(x^{N+1}) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ f_0(x^{N+m}) & f_1(x^{N+m}) & \dots & f_k(x^{N+m}) \end{bmatrix}, \tag{11.7}$$

$$\tilde{Y}_m = \begin{bmatrix} \tilde{y}^{N+1} \\ \tilde{y}^{N+2} \\ \vdots \\ \tilde{y}^{N+m} \end{bmatrix}. \qquad (11.8)$$

Предполагается, что дисперсия ошибки наблюдений в новых экспериментах та же, что и в предыдущих. Требуется построить алгоритмы вычисления:

а) оценок $\hat{a}^{N+m}$ через оценку $\hat{a}^N$ и дополнительную информацию, содержащуюся в матрицах F_m и $\tilde{Y}_m$;

б) дисперсионной матрицы C_{N+m} через матрицу C_N и F_m;

в) остаточной суммы квадратов S_{N+m} через S_N, $\hat{a}^N$, F_m и $\tilde{Y}_m$.

11.2.2. Рекуррентные алгоритмы вычисления оценок коэффициентов и дисперсионной матрицы плана

Приведем здесь вывод рекуррентных выражений для C_{N+m} и $\hat{a}^{N+m}$. Запишем аналогично (11.2) выражение для оценок коэффициентов $\hat{a}^{N+m}$, найденных по выборке объема $(N+m)$ с помощью метода наименьших квадратов:

$$\hat{a}^{N+m} = (F'_{N+m}F_{N+m})^{-1} F'_{N+m} \tilde{Y}_{N+m}. \qquad (11.9)$$

Здесь матрицы F_{N+m} и $\tilde{Y}_{N+m}$ образованы путем добавления к F_m и $\tilde{Y}_N$ m строк матриц F_m и $\tilde{Y}_m$ соответственно. При этом

$$\hat{a}^{N+m} = (F'_N F_N + F'_m F_m)^{-1} (F'_N \tilde{Y}_N + F'_m \tilde{Y}_m). \qquad (11.10)$$

Нетрудно убедиться, что

$$(F'_N F_N + F'_m F_m)^{-1} = C_N - C_N F'_m [F_m C_N F'_m + I_m]^{-1} F_m C_N, \qquad (11.11)$$

где I_m — единичная матрица размера m. Правильность формулы (11.11) легко проверить, умножив левую и правую части равенства (11.11) справа на матрицу $(F'_N F_N + F'_m F_m)$ [3, 4]. Матрица

$$(F'_{N+m}F_{N+m})^{-1} = (F'_N F_N + F'_m F_m)^{-1} = C_{N+m} \qquad (11.12)$$

есть дисперсионная матрица плана X_{N+m}. Отсюда следует искомая рекуррентная формула для дисперсионной матрицы

$$C_{N+m} = C_N - C_N F'_m [F_m C_N F'_m + I_m]^{-1} F_m C_N. \qquad (11.13)$$

Подставив выражение (11.11) в (11.10) и произведя несложные преобразования, найдем искомое рекуррентное выражение для

оценок коэффициентов:

$$\hat{a}^{N+m} = \hat{a}^N + C_N F'_m [F_m C_N F'_m + I_m]^{-1} (\tilde{Y}_m - F_m \hat{a}^N). \quad (11.14)$$

Применение выражений (11.13) и (11.14) при последовательном построении математической модели приводит при $m < k+1$ к более простым вычислительным операциям. Действительно, использование формулы (11.9) связано с необходимостью обращения матрицы $F'_{N+m} F_{N+m}$ размера $(k+1)(k+1)$, где $(k+1)$ — число оцениваемых коэффициентов. При проведении расчетов по формулам (11.13), (11.14) возникает необходимость обращения матрицы $(I_m + F_m C_N F'_m)$ размера $m \times m$. Если $m < k+1$, последний способ явно предпочтительнее.

Наиболее простой вид принимают вычислительные формулы в случае, когда $m = 1$, т. е. когда пересчет оценок коэффициентов осуществляется после каждого опыта:

$$C_{N+1} = C_N - \frac{C_N f(x^{N+1}) f'(x^{N+1}) C_N}{1 + f'(x^{N+1}) C_N f(x^{N+1})}, \quad (11.15)$$

$$\hat{a}^{N+1} = \hat{a}^N + \frac{C_N f(x^{N+1})}{1 + f'(x^{N+1}) C_N f(x^{N+1})} (\tilde{y}^{N+1} - f'(x^{N+1}) \hat{a}^N). \quad (11.16)$$

Здесь вообще нет необходимости производить обращение матриц, так как матрица $(I_m + F_m C_N F'_m)$ трансформировалась в скалярную величину.

Применение для вычисления C_{N+m} и $\hat{a}^{N+m}$ рекуррентных выражений (11.13), (11.14) или (11.15), (11.16) требует знания дисперсионной матрицы C_N и оценок $\hat{a}^N$. Если $N = 0$, т. е. требуется вычислить оценки для первой серии опытов, то величины C_0 и $\hat{a}_0$ необходимо каким-то образом задать. В этом случае можно поступить так: реализовать план, состоящий из $N \geqslant k+1$ экспериментов; для этого плана найти матрицу C_N и оценки $\hat{a}^N$, обращая матрицу $F'_N F_N$. После этого оказывается возможным применять приведенные выше рекуррентные формулы. Возможен, однако, и иной способ. Зададимся значениями

$$\begin{aligned} \hat{a}^0 &= 0, \\ C_0 &= I_{k+1} c^2, \end{aligned} \quad (11.17)$$

где c — большое число (согласно [3], $c \geqslant 10^{16}$), и будем применять рекуррентные формулы, начиная с первой серии опытов. Можно показать, что при этом, если величина c выбрана достаточно большой, после i-й серии из m опытов, где i таково, что $im \geqslant k+1$, оценки коэффициентов $\hat{a}^{mi}$ будут практически не отличаться от строго оптимальных оценок $\hat{a}^{mi}$, вычисленных по формуле (11.9) [4].

11.2.3. Рекуррентный алгоритм вычисления остаточной суммы квадратов

Запишем очевидное выражение для остаточной суммы квадратов

$$S_{N+m} = (\tilde{Y}_N - F_N\hat{a}^{N+m})'(\tilde{Y}_N - F_N\hat{a}^{N+m}) + (\tilde{Y}_m - F_m\hat{a}^{N+m})'(\tilde{Y}_m - F_m\hat{a}^{N+m}). \quad (11.18)$$

Первое слагаемое в (11.18) можно преобразовать следующим образом [6]:

$$(\tilde{Y}_N - F_N\hat{a}^{N+m})'(\tilde{Y}_N - F_N\hat{a}^{N+m}) = (\tilde{Y}_N - F_N\hat{a}^N)'(\tilde{Y}_N - F_N\hat{a}^N) + (\hat{a}^{N+m} - \hat{a}^N)'F_N'F_N(\hat{a}^{N+m} - \hat{a}^N) = S_N + (\hat{a}^{N+m} - \hat{a}^N)'F_N'F_N(\hat{a}^{N+m} - \hat{a}^N). \quad (11.19)$$

Выражение (11.19) позволяет вычислить увеличение остаточной суммы квадратов при замене оценок $\hat{a}^N$, полученных по выборке объема N методом наименьших квадратов, некоторым произвольным вектором $\hat{a}^{N+m}$. Такая замена часто имеет место при исключении из уравнения регрессии функций с незначимыми коэффициентами. Поэтому формула (11.19) представляет самостоятельный интерес. Имея в виду, что из (11.14) следует

$$(\hat{a}^{N+m} - \hat{a}^N) = C_N F_m'(F_m C_N F_m' + I_m)^{-1}(\tilde{Y}_m - F_m\hat{a}^N), \quad (11.20)$$

нетрудно найти следующие выражения для слагаемых формулы (11.18):

$$(\tilde{Y}_N - F_N\hat{a}^{N+m})'(\tilde{Y}_N - F_N\hat{a}^{N+m}) = S_N + (\tilde{Y}_m - F_m\hat{a}^N)' \times \times (I_m + F_m C_N F_m')^{-1} F_m C_N F_m' (I_m + F_m C_N F_m')^{-1}(Y_m - F_m\hat{a}^N), \quad (11.21)$$

$$(\tilde{Y}_m - F_m\hat{a}^{N+m})'(\tilde{Y}_m - F_m\hat{a}^{N+m}) = (\tilde{Y}_m - F_m\hat{a}^N)'(I_m + F_m C_N F_m')^{-1} \times \times (I_m + F_m C_N F_m')^{-1}(\tilde{Y}_m - F_m\hat{a}^N). \quad (11.22)$$

Подставляя (11.21) и (11.22) в (11.18) и производя несложные преобразования, окончательно получаем

$$S_{N+m} = S_N + (\tilde{Y}_m - F_m\hat{a}^N)'(I_m + F_m C_N F_m')^{-1}(\tilde{Y}_m - F_m\hat{a}^N). \quad (11.23)$$

В случае когда $m = 1$, т. е. когда добавлена только одна экспериментальная точка, имеем

$$S_{N+1} = S_N + \frac{1}{1 + f'(x^{N+1}) C_N f(x^{N+1})}(\tilde{y}^{N+1} - f'(x^{N+1})\hat{a}^N)^2. \quad (11.24)$$

Остаточная сумма квадратов может быть использована для вычисления остаточной дисперсии $(s_R^2)^{N+m}$ по формуле

$$(s_R^2)^{N+m} = \frac{S_{N+m}}{N + m - (k+1)}. \quad (11.25)$$

Остаточная дисперсия является оценкой дисперсии ошибок наблюдений с числом степеней свободы $\varphi = N + m - (k+1)$. В случае когда ошибка наблюдений e распределена по нормальному закону, оценка $(s_R^2)^{N+m}$ может быть использована при проверке различного рода гипотез о коэффициентах модели, а также при построении доверительных интервалов для оцениваемых величин (см. разд. 3).

11.2.4. Пример

Пример 11.1

Рассмотрим часть изложенного в работе [7] решения задачи определения оптимальных режимов процесса переэтерификации метилметакрилата (ММК) диметиламиноэтанолом. Здесь экспериментальным путем строилась модель, связывающая выход процесса (в процентах от теоретического выхода) со следующими переменными: z_1 — доза катализатора (условные единицы); z_2 — молярное отношение реагентов (аминоспирт/ММК); z_3 — время реакции (часы). Зависимость между указанными переменными искалась в виде

$$y(a, x) = a_0 + a_1x_1 + a_2x_2 + a_3x_3. \qquad (11.26)$$

Здесь x_1, x_2 и x_3 — нормированные входные переменные

$$x_1 = \frac{z_1 - 4}{1}, \quad x_2 = \frac{z_2 - 1{,}4}{0{,}1}, \quad x_3 = \frac{z_3 - 3}{0{,}5}.$$

Решение искалось таким путем: сначала был реализован насыщенный план, представляющий собой полуреплику полного факторного эксперимента с генератором $x_3 = -x_1x_2$.

Статистический анализ полученных результатов показал, что коэффициент при переменной x_3 незначимо отличается от нуля. Было высказано предположение, что причина получения такого результата кроется в чрезмерно большой дисперсии ошибок наблюдений. В связи с этим было решено продолжить эксперименты с тем, чтобы уточнить значения коэффициентов и проверить справедливость выбранного вида модели. При этом оказалось удобным все вычисления проводить по рекуррентным формулам, чтобы после каждого опыта располагать всеми результатами расчета. После добавления двух опытов, взятых из полуреплики полного факторного эксперимента с генератором $x_3 = x_1x_2$, были получены значимые величины оценок всех коэффициентов и опыты были прекращены.

Результаты расчетов представлены в табл. 11.1. Здесь $\tilde{y}$ — наблюдаемые значения зависимой переменной. Оценка s^2 дисперсии значений $\tilde{y}$, определенная по ряду параллельных опытов

Таблица 11.1

j	План x_1^j	x_2^j	x_3^j	$\tilde{y}^j$	Оценки коэффициентов a_0^j	a_1^j	a_2^j	a_3^j	Дисперсии оценок коэффициентов $(s_i^2)^j$	$t_{кр}(s_i)^j$	S_j	$(s_R^2)^j$	$F=\frac{(s_R^2)^j}{s^2}$
1	−1	−1	−1	45									
2	+1	−1	+1	84,6									
3	−1	+1	+1	59,1									
4	+1	+1	−1	82,4	67,77	15,72	2,97	4,07	2,73	3,53	0	0	—
5	−1	−1	+1	58,2	68,4	15,09	2,34	4,70	2,38	3,28	12,75	12,75	1,17
6	+1	−1	−1	67,9	67,33	14,02	3,41	5,77	2,04	3,04	49,35	24,68	2,26

$t_{кр}=2{,}132;\quad \varphi_2=4;\quad \alpha=0{,}1;$

$s^2=10{,}9;$

$F_{кр}=10{,}64;\quad \varphi_1=2;\quad \varphi_2=4;\quad \alpha=0{,}05.$

при числе степеней свободы $\varphi_2 = 4$, равна $s^2 = 10{,}9$ [1]. Дисперсионная матрица после первых четырех опытов имеет вид

$$C_4 = \frac{1}{4} I_4,$$

где I_4 — единичная матрица. Оценки коэффициентов $\hat{a}_i^4$; $i = 0, \ldots, 3$, найдены по формуле

$$\hat{a}_i^4 = \frac{1}{4} \sum_{l=1}^{4} \tilde{y}^l x_i^l.$$

Оценка дисперсии коэффициентов после четырех опытов

$$(s_i^2)^4 = \frac{s^2}{4} = 2{,}73.$$

Проверка значимости коэффициентов проводится с помощью критерия Стьюдента. Коэффициент a_i значимо отличается от нуля, если после j-го опыта выполняется условие

$$|\hat{a}_i^j| \geqslant t_{кр} (s_i)^j,$$

где $t_{кр}$ — критическое значение распределения Стьюдента при числе степеней свободы $\varphi_2 = 4$ и выбранном уровне значимости $\alpha = 0{,}1$ [2]. S_j — остаточная сумма квадратов после j-го эксперимента, а $(s_R^2)^j$ — остаточная дисперсия, определяемая по формуле

$$(s_R^2)^j = \frac{S_j}{j - (k+1)} = \frac{S_j}{j-4}.$$

В последней колонке табл. 11.1 приведено отношение остаточной дисперсии к оценке дисперсии ошибки, используемое для проверки адекватности модели по F-критерию. Найдем дисперсионную матрицу C_5 после пятого опыта:

$$C_5 = \frac{1}{32} \begin{bmatrix} 7 & 1 & 1 & -1 \\ 1 & 7 & -1 & 1 \\ 1 & -1 & 7 & 1 \\ -1 & 1 & 1 & 7 \end{bmatrix}.$$

---

[1] В работе [7] оценка дисперсии ошибок наблюдений производится по результатам дублирования опытов в каждой точке последовательно достраиваемого плана; мы отклонились от изложения этого примера, данного в работе [7], ибо там не приведены оценки дисперсии после добавленных в план экспериментов.

[2] В работе [7] уровень значимости $\alpha = 0{,}05$.

Вектор коэффициентов $\hat{a}^5$ рассчитываем по формуле (11.16), а остаточную сумму квадратов по формуле (11.24);

$$\hat{a}^5 = \begin{bmatrix} 67{,}77 \\ 15{,}72 \\ 2{,}97 \\ 4{,}07 \end{bmatrix} + \frac{1}{8}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}(58{,}2 - 53{,}15) = \begin{bmatrix} 68{,}4 \\ 15{,}09 \\ 2{,}34 \\ 4{,}70 \end{bmatrix},$$

$$S_5 = \frac{1}{2}(58{,}2 - 53{,}15)^2 = 12{,}75.$$

Дисперсии оценок коэффициентов равны

$$(s_l^2)^5 = \frac{7}{32s^2} = 2{,}38.$$

Из данных табл. 11.1 ясно, что после пятого эксперимента коэффициент при x_2 остается незначимым.

Оценки коэффициентов после шестого опыта определяются путем следующих расчетов:

$$\hat{a}^6 = \begin{bmatrix} 68{,}4 \\ 15{,}09 \\ 2{,}34 \\ 4{,}70 \end{bmatrix} + \frac{1}{64}\begin{bmatrix} 7 & 1 & 1 & -1 \\ 1 & 7 & -1 & 1 \\ 1 & -1 & 7 & 1 \\ -1 & 1 & 1 & 7 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}(67{,}9 - 76{,}45) = \begin{bmatrix} 67{,}33 \\ 14{,}02 \\ 3{,}41 \\ 5{,}77 \end{bmatrix}.$$

Остаточная сумма квадратов составляет

$$S_6 = 12{,}75 + \frac{(67{,}9 - 76{,}45)^2}{2} = 24{,}68.$$

Дисперсионная матрица после шестого опыта имеет вид

$$C_6 = \frac{1}{32}\begin{bmatrix} 6 & 0 & 2 & 0 \\ 0 & 6 & 0 & 2 \\ 2 & 0 & 6 & 0 \\ 0 & 2 & 0 & 6 \end{bmatrix}.$$

Отсюда дисперсии оценок коэффициентов:

$$(s_l^2)^6 = \frac{6}{32}s^2 = 2{,}04.$$

Из табл. 11.1 следует, что после шестого опыта все коэффициенты значимо отличаются от нуля. Кроме того, сопоставляя полученное значение $F = 2{,}26$ с $F_{кр} = 10{,}64$ при уровне значимости $\alpha = 0{,}05$ и числах степеней свободы $\varphi_1 = 2$ и $\varphi_2 = 4$, мы убеждаемся в адекватности линейной модели.

11.3. Последовательное планирование эксперимента

11.3.1. Постановка задачи последовательного планирования эксперимента

Пусть исследуется некоторый объект, причем по результатам N реализованных опытов методом наименьших квадратов найдены оценки коэффициентов модели объекта вида

$$y(a, x) = a'f(x).$$

Обозначим через X_N матрицу реализованного плана, а через F_N — матрицу значений функций независимых переменных для точек плана X_N. Требуется выбрать наилучшие условия проведения еще m опытов в заданной области планирования Ω_x, т. е. найти наилучшую в некотором смысле матрицу X_m. В качестве критерия планирования мы здесь будем использовать критерий D-оптимальности, рассмотренный в разд. 6.

Будем полагать, что дисперсия ошибок наблюдений при реализации планов X_N и X_m имеет одну и ту же величину. Дисперсионная матрица C_{N+m} после реализации плана X_{N+m} имеет вид

$$C_{N+m} = (F'_N F_N + F'_m F_m)^{-1}. \tag{11.27}$$

В соответствии с критерием D-оптимальности план X_m требуется выбрать так, чтобы определитель матрицы C_{N+m} имел минимальную величину, а определитель информационной матрицы M_{N+m} был бы максимален. Таким образом, план X_m^* находится из условий

$$|C_{N+m}(X_m^*)| = \min_{X_m \in \Omega_x} |C_{N+m}(X_m)| = \min_{X_m \in \Omega_x} |(F'_N F_N + F'_m F_m)^{-1}| \tag{11.28а}$$

или

$$|M_{N+m}(X_m^*)| = \max_{X_m \in \Omega_x} |M_{N+m}(X_m)| = \max_{X_m \in \Omega_x} |F'_N F_N + F'_m F_m|. \tag{11.28б}$$

Условия (11.28а) и (11.28б), как показано в разд. 6.2, обеспечивают минимальный объем эллипсоида рассеяния вектора $\hat{a}^{N+m}$ оценок коэффициентов модели после $N+m$ опытов. В случае нормального закона распределения ошибки эксперимента условия (11.28а) и (11.28б) приводят к минимальной величине объема доверительного эллипсоида для истинных значений коэффициентов.

Пример 11.2

Пусть модель имеет вид

$$y(a, x) = a_0 + a_1 x.$$

Положим, что $F'_N F_N = I_{k+1}$, где $1 + k = 2$, а область планирования Ω_x определяется следующим образом:

$$+1 \geqslant x \geqslant -1.$$

Требуется определить точки $x^{N+1}, \ldots, x^{N+m}$ постановки m экспериментов из условия (11.28а). В данном случае удобно воспользоваться следующим выражением для $|C_{N+m}|$:

$$|C_{N+m}| = |(F'_N F_N + F'_m F_m)^{-1}|.$$

Так как

$$F_m = \begin{bmatrix} 1 & x^{N+1} \\ 1 & x^{N+2} \\ \vdots & \\ 1 & x^{N+m} \end{bmatrix},$$

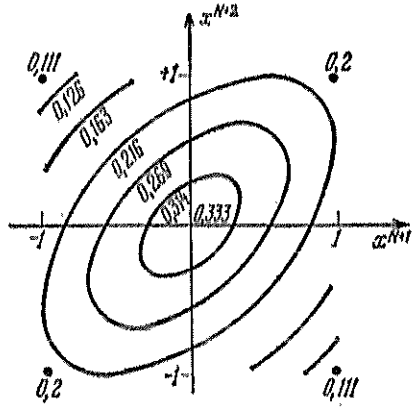
нетрудно найти

$$F'_m F_m = \begin{bmatrix} m & \sum_{l=1}^{m} x^{N+l} \\ \sum_{l=1}^{m} x^{N+l} & \sum_{l=1}^{m} (x^{N+l})^2 \end{bmatrix},$$

$$|C_{N+m}| = \left[(m+1)\left(1 + \sum_{l=1}^{m} (x^{N+l})^2\right) - \left(\sum_{l=1}^{m} x^{N+l}\right)^2 \right]^{-1}.$$

На фиг. 11.1 показаны линии равных значений определителя дисперсионной матрицы при $m = 2$. Здесь по оси абсцисс отложены значения x^{N+1}, по оси ординат x^{N+2}. Мы видим, что глобальный минимум имеет место при постановке эксперимента в точках $x^{N+1} = 1$, $x^{N+2} = -1$ или $x^{N+1} = -1$, $x^{N+2} = 1$. Выбор наилучшего плана X_m^* из условий (11.28) в общем случае требует решения достаточно сложной экстремальной задачи размерности $m \times n$, где n — число независимых переменных, а m — число планируемых опытов. Решение этой задачи, за исключением простейших случаев, может быть найдено лишь с помощью цифровых вычислительных машин. При поиске X_m^* трудности вычислительного порядка связаны прежде всего с размерностью задачи $m \times n$. Поэтому наибольшее распространение имеют одношаговые методы последовательного планирования, когда ищется наилучшая точка постановки только одного текущего опыта

(т. е при $m=1$). Кроме того, поиск X_m^* связан с многократным вычислением определителя матрицы C_{N+m} (или M_{N+m}) размера $(k+1)(k+1)$. Учитывая специфику последовательного планирования, здесь можно воспользоваться рекуррентными форму-



Фиг. 11.1. Линии равных значений определителя дисперсионной матрицы примера 11.2.

лами для определителя дисперсионной матрицы, более экономными с точки зрения объема вычислений.

11.3.2. Рекуррентная формула для определителя дисперсионной матрицы

Найдем рекуррентную формулу для вычисления определителя матрицы C_{N+m} с помощью определителя матрицы C_N и матрицы X_m.

В разд. 11.2 было получено следующее выражение для матрицы C_{N+m}:

$$C_{N+m}=C_N-C_NF'_m[F_mC_NF'_m+I_m]^{-1}F_mC_N.$$

Так как определитель произведения квадратных матриц равен произведению определителей сомножителей [8], то

$$|C_{N+m}|=|I_{k+1}-C_NF'_m[F_mC_NF'_m+I_m]^{-1}F_m||C_N|. \quad (11.29)$$

Формулу (11.29) целесообразно привести к иному виду, воспользовавшись следующим утверждением [8]: если B—матрица размера $(k\times m)$, а D—матрица размера $(m\times k)$, то

$$|I_k-BD|=|I_m-DB|. \quad (11.30)$$

Положив теперь в (11.30)

$$B=C_NF'_m[F_mC_NF'_m+I_m]^{-1},$$
$$D=F_m,$$

найдем

$$|I_{k+1} - C_N F'_m [F_m C_N F'_m + I_m]^{-1} F_m| = \\ = |I_m - F_m C_N F'_m [F_m C_N F'_m + I_m]^{-1}|. \qquad (11.31)$$

Подставляя (11.31) в (11.29), получим

$$|C_{N+m}| = |I_m - F_m C_N F'_m [F_m C_N F'_m + I_m]^{-1}| |C_N|. \qquad (11.32)$$

При пользовании формулой (11.32) операции вычисления определителя матрицы проводятся с матрицами размера $(m \times m)$ в отличие от формулы (11.29), где аналогичные операции осуществляются с матрицами размера $(k+1) \times (k+1)$. Поэтому формула (11.32) оказывается удобнее (11.29) при $m < k+1$; наоборот, при $m > k+1$ целесообразнее производить вычисление определителя матрицы C_{N+m} по формуле (11.29). Особенно простой формула (11.32) становится при $m = 1$, т. е. когда при вычислении определителя точки добавляются по одной. При этом имеем

$$|C_{N+1}| = \left(1 - \frac{f'(x^{N+1}) C_N f(x^{N+1})}{1 + f'(x^{N+1}) C_N f(x^{N+1})}\right) |C_N|. \qquad (11.33)$$

11.3.3 Вычислительная процедура последовательного планирования

Из (11.32) следует, что минимальное значение определителя матрицы C_{N+m} достигается при таком плане X_m^*, который обеспечивает

$$\min_{X_m \in \Omega_x} |I_m - F_m C_N F'_m [F_m C_N F'_m + I_m]^{-1}|. \qquad (11.34)$$

Из (11.33) при $m = 1$ вместо условия (11.34) имеем

$$\min_{x^{N+1} \in \Omega_x} f'(x^{N+1}) C_N f(x^{N+1}). \qquad (11.35)$$

Из (11.35) следует, что наилучшей точкой x^{N+1} постановки $(N+1)$-го эксперимента является точка, в которой достигается максимум квадратичной формы

$$\max_{x^{N+1} \in \Omega_x} f'(x^{N+1}) C_N f(x^{N+1}). \qquad (11.36)$$

Напомним, что дисперсия оценки зависимой переменной в точке x рассчитывается по формуле (3.85). Из сопоставления (11.36) и (3.85) следует, что наилучшая точка постановки $(N+1)$-го опыта есть точка области Ω_x, где имеет место максимальная дисперсия оценки зависимой переменной.

Таким образом, задача последовательного планирования при $m = 1$ фактически сводится к следующим вычислениям, прово-

димым последовательно для каждого нового опыта, т. е. при $i = 1, 2, \ldots$:

1) определение точки x_*^{N+i} из условия

$$f'(x_*^{N+i}) C_{N+i-1} f(x_*^{N+i}) = \max_{x^{N+i} \in \Omega_x} f'(x^{N+i}) C_{N+i-1} f(x^{N+i}); \quad (11.37)$$

2) расчет дисперсионной матрицы для плана после включения в него точки x_*^{N+i}

$$C_{N+i} = C_{N+i-1} - \frac{C_{N+i-1} f(x_*^{N+i}) f'(x_*^{N+i}) C_{N+i-1}}{1 + f'(x_*^{N+i}) C_{N+i-1} f(x_*^{N+i})}. \quad (11.38)$$

Естественно, что для использования этих формул требуется, чтобы был задан начальный план X_N с невырожденной информационной матрицей $|M_N| = |F_N' F_N| > 0$. Пример построения алгоритма последовательного планирования можно найти в работе [9].

Обратим внимание на то, что последовательность точек x_*^i, получаемая с помощью процедуры (11.37), (11.38), не зависит от результатов реализуемых опытов. Это означает, что, каковы бы ни были область планирования Ω_x, начальный план X_N и модель, такая последовательность может быть построена заранее, т. е. до начала эксперимента.

Приведем здесь два примера последовательностей экспериментов, полученных в [9] путем применения процедур (11.37), (11.38).

Пример 11.3

Для модели

$$y(a, x) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2 \quad (11.39)$$

и области планирования

$$-1 \leqslant x_i \leqslant +1, \quad i = 1, 2,$$

при начальном плане X_N, содержащем по одному разу точки непрерывного D-оптимального плана (см. разд. 6 и фиг. 6.3а), последовательность экспериментов приведена в табл. 11.2. В таблице указаны номера точек в соответствии с начальным планом и фиг. 6.3а.

Пример 11.4

Для модели вида

$$y(a, x) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_1^2 + a_5 x_2^2 + a_6 x_3^2 + a_7 x_1 x_2 + a_8 x_1 x_3 + a_9 x_2 x_3 \quad (11.40)$$

и области планирования

$$-1 \leqslant x_i \leqslant +1, \quad i = 1, 2, 3,$$

последовательность экспериментальных точек приведена в табл. 11.3. Начальный план X_N включает здесь 21 точку: это точки непрерывного D-оптимального плана Коно (см. табл. 6.4 и фиг. 6.5). Точки начального плана пронумерованы.

11.4. Правило остановки эксперимента

При реализации последовательного способа построения математической модели чрезвычайно важно уметь определять момент, когда следует прекратить дальнейшие эксперименты. Таким моментом может быть, например, момент, когда достигнута требуемая точность модели. Правила понятия решения о прекращении дальнейших опытов называют правилами остановки (или останова) эксперимента.

Таблица 11.2

Последовательность проведения экспериментов для модели (11.39)

Номер точки плана	Начальный план X_N	
	x_1	x_2
1	1	1
2	−1	1
3	−1	−1
4	1	−1
5	0	0
6	1	0
7	0	1
8	−1	0
9	0	−1

t	Номер точки плана	t	Номер точки плана	t	Номер точки плана
1	1	21	3	41	1
2	4	22	4	42	4
3	2	23	2	43	2
4	3	24	1	44	3
5	5	25	9	45	6
6	7	26	8	46	7
7	4	27	5	47	9
8	3	28	1	48	8
9	1	29	4	49	1
10	2	30	2	50	4
11	6	31	3		
12	9	32	7		
13	8	33	6		
14	1	34	8		
15	2	35	9		
16	4	36	1		
17	3	37	4		
18	5	38	2		
19	7	39	3		
20	6	40	5		

Представляется естественным прекращать эксперименты тогда, когда с заданной вероятностью $(1-\alpha)$ истинные значения оцениваемых коэффициентов попадают внутрь некоторой области, размеры и конфигурация которой задаются исследователем. Правило останова, базирующееся на проверке такого условия, рассматривается ниже [10].

11.4.1. Постановка задачи

Полагаем, что модель исследуемого объекта имеет вид (11.1). Экспериментальным путем ищутся оценки коэффициентов a модели на основе критерия наименьших квадратов, т. е. постановка и способ решения задачи не отличаются от изложенного в разд. 11.2 и 3.2. Будем предполагать, что ошибка наблюдений e (см. разд. 3.2) во всяком i-м опыте распределена по нормальному закону.

Полученное ниже правило прекращения эксперимента может быть использовано как в случае, когда осуществляется последовательное планирование, так и в случае так называемого „пассивного" эксперимента, когда модель строится на основе данных, получаемых от объекта в ходе его нормальной эксплуатации.

Положим, что требуемая точность оценок коэффициентов модели задана с помощью области в пространстве a, ограни-

Таблица 11.3

Последовательность проведения экспериментов для модели (11.40)

Номер точки плана	Начальный план X_N			
	x_1	x_2	x_3	
1	0	0	0	Центр куба
2	−1	−1	−1	Вершины куба
3	+1	+1	+1	
4	+1	−1	−1	
5	−1	+1	+1	
6	−1	+1	−1	
7	+1	−1	+1	
8	−1	−1	+1	
9	+1	+1	−1	
10	0	−1	−1	Середины ребер
11	−1	−1	0	
12	−1	0	−1	
13	−1	0	+1	
14	0	−1	+1	
15	−1	+1	0	
16	+1	−1	0	
17	+1	0	−1	
18	0	+1	−1	
19	+1	+1	0	
20	+1	0	+1	
21	0	+1	+1	

Продолжение табл. 11.3

l	Номер точки плана	l	Номер точки плана	l	Номер точки плана	l	Номер точки плана	l	Номер точки плана
1	1	13	9	25	16	37	4	49	6
2	2	14	4	26	17	38	5	50	4
3	3	15	3	27	18	39	3	51	8
4	4	16	5	28	9	40	10	52	20
5	5	17	4	29	19	41	13	53	2
6	6	18	6	30	20	42	11	54	3
7	7	19	7	31	21	43	14	55	7
8	8	20	8	32	1	44	16	56	19
9	9	21	13	33	8	45	9	57	18
10	10	22	14	34	6	46	5	58	21
11	11	23	15	35	7	47	17	59	15
12	12	24	2	36	2	48	1	60	13

ченной эллипсоидом

$$(a-\hat{a}^N)' B (a-\hat{a}^N) = \rho^2. \qquad (11.41)$$

Здесь $\hat{a}^N$—вектор оценок коэффициентов, найденный по результатам N опытов методом наименьших квадратов, а B—положительно определенная матрица. Область, ограниченную эллипсоидом (11.41), обозначим Ω_N и будем называть ее допустимой областью. Момент остановки должен иметь место при таком числе экспериментов N, когда с вероятностью, не меньшей заданной величины $1-\alpha$, истинные значения коэффициентов $\bar{a}$ попадают внутрь области Ω_N, т. е.

$$P(\bar{a} \in \Omega_N) \geqslant 1-\alpha. \qquad (11.42)$$

11.4.2. Метод решения задачи

Для решения задачи воспользуемся понятием доверительного эллипсоида для коэффициентов уравнения регрессии (см. разд. 6.2). Как известно, доверительный эллипсоид определяет область в пространстве a, накрывающую с заданной доверительной вероятностью $1-\alpha$ точку a действительных значений оцениваемых параметров. Уравнение доверительного эллипсоида записывается следующим образом:

$$(a-\hat{a}^N)' F_N' F_N (a-\hat{a}^N) = \delta^2 = (k+1) s^2 F_{\text{кр}}. \qquad (11.43)$$

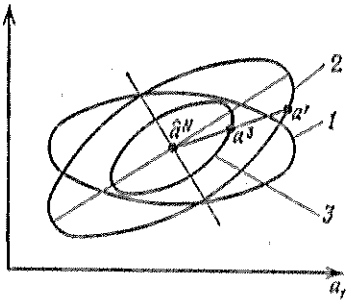
Здесь F_N—матрица значений функций независимых переменных в N опытах, определяемая согласно (11.5); s^2—оценка дисперсии ошибки наблюдений с числом степеней свободы φ_e; $(k+1)$—

размерность вектора коэффициентов; $F_{кр}$ — критическое значение распределения Фишера при уровне значимости α и числах степеней свободы $\varphi_1 = k+1$ и $\varphi_2 = \varphi_e$.

Заметим, что если известна истинная дисперсия σ^2, то

$$\delta^2 = \sigma^2 \chi^2_{кр}, \tag{11.44}$$

где $\chi^2_{кр}$ — критическое значение χ^2-распределения при уровне значимости α и числе степеней свободы $\varphi = k+1$.



Фиг. 11.2. Пояснения к правилу останова [1].
1 — граница допустимой области; *2* — доверительный эллипс для $\alpha = \alpha_3$; *3* — доверительный эллипс для $\alpha = \alpha_2$.

В качестве оценки s^2 может быть использована остаточная дисперсия $(s^2_R)^N$ с числом степеней свободы $\varphi = N - (k+1)$.

Очевидно, что условие останова (11.42) будет выполнено, если доверительный эллипсоид окажется внутри поверхности (11.41), ограничивающей допустимую область. Этот вывод иллюстрирует фиг. 11.2, где рассмотрен случай оценки двух коэффициентов. Пусть эллипс *1* соответствует границе допустимой области, а *2* и *3* — доверительные эллипсы при некотором количестве экспериментов N и доверительных вероятностях $1-\alpha_2$ и $1-\alpha_3$ соответственно (очевидно, что $\alpha_2 < \alpha_3$). Поскольку эллипс *2* не охватывается границей допустимой области, то нет оснований для утверждения, что

$$P(\bar{a} \in \Omega_N) \geqslant 1-\alpha_2 \tag{11.45}$$

и, следовательно, в случае $\alpha = \alpha_2$ эксперимент необходимо продолжить. Эллипс *3* расположен внутри эллипса *1*, и, следовательно, можно утверждать, что

$$P(\bar{a} \in \Omega_N) \geqslant 1-\alpha_3. \tag{11.46}$$

Это значит, что при $\alpha = \alpha_3$ эксперимент следует прекратить.

[1] На фиг. 11.2 точка a^1 ошибочно показана на эллипсе *2*. — *Прим. ред.*

11.4.3. Правило останова

Покажем, что с вычислительной точки зрения задача проверки условий останова (11.42) сводится к отысканию минимального собственного числа матрицы B^{-1}. Предварительно убедимся, что в случае, когда доверительный эллипсоид (11.43) охватывается границей допустимой области, значения квадратичной формы

$$(a-\hat{a}^N)'F'_NF_N(a-\hat{a}^N)=r_N(a) \qquad (11.47)$$

в точках a, принадлежащих границе допустимой области, описываемой уравнением (11.41), не меньше δ^2, т. е.

$$r_N(a)\geqslant\delta^2 \quad \text{при} \quad (a-\hat{a}^N)'B(a-\hat{a}^N)=\rho^2.$$

Действительно, соединим на фиг. 11.2 некоторую произвольную точку a^1 границы допустимой области с точкой $\hat{a}^N$ вектором $(a^1-\hat{a}^N)$, который пересечет эллипс *3* в точке a^3. Так как эллипс *1* охватывает доверительный эллипс *3*, то $(a^1-\hat{a}^N)=\mu(a^3-\hat{a}^N)$, где $\mu\geqslant 1$. Отсюда

$$r_N(a^1)=(a^1-\hat{a}^N)'F'_NF_N(a^1-\hat{a}^N)=\mu^2(a^3-\hat{a}^N)'F'_NF_N(a^3-\hat{a}^N)=$$
$$=\mu^2 r_N(a^3)\geqslant r_N(a^3)=\delta^2, \qquad (11.48)$$

что и требовалось доказать. Тогда условие останова может быть записано следующим образом:

$$\min_a r_N(a)\geqslant\delta^2 \quad \text{при} \quad (a-\hat{a}^N)'B(a-\hat{a}^N)=\rho^2. \qquad (11.49)$$

Таким образом, для проверки условий останова достаточно найти минимум квадратичной формы (11.47) на поверхности (11.41). Применяя метод неопределенных множителей Лагранжа, найдем решение из условия

$$\min_{\Delta a}\{\Delta a'F'_NF_N\Delta a+\lambda(\rho^2-\Delta a'B\Delta a)\}, \qquad (11.50)$$

где $\Delta a=a-\hat{a}^N$. Дифференцируя по Δa и приравнивая производную нулю, получим

$$F'_NF_N\Delta a-\lambda B\Delta a=0. \qquad (11.51)$$

Система однородных линейных уравнений (11.51) имеет нетривиальное решение относительно Δa при

$$|B^{-1}F'_NF_N-\lambda I_{k+1}|=0. \qquad (11.52)$$

Корни уравнения (11.52) $\lambda_1, \lambda_2, \ldots, \lambda_{k+1}$ являются собственными числами матрицы $B^{-1}F'_NF_N$ [8]. Для каждого собственного

числа $\lambda_1, \lambda_2, \ldots, \lambda_{k+1}$ может быть найден собственный вектор $\Delta a^1, \Delta a^2, \ldots, \Delta a^{k+1}$, причем

$$(\Delta a^l)' F_N' F_N (\Delta a^l) = \lambda_l (\Delta a^l)' B (\Delta a^l) = \lambda_l \rho^2. \qquad (11.53)$$

Из (11.53) непосредственно следует, что минимальное значение квадратичной формы (11.47) на границе допустимой области равно

$$\min_a r_N(a) = \lambda_{N\,\min} \rho^2, \qquad (11.54)$$

где $\lambda_{N\,\min}$ — минимальное собственное число матрицы $B^{-1} F_N' F_N$. При этом имеем следующее правило останова: эксперимент следует прекратить после N опытов, если

$$\lambda_{N\,\min} \rho^2 \geqslant \delta^2. \qquad (11.55)$$

Заметим, что если до начала эксперимента известна оценка s^2 дисперсии ошибки наблюдений и, кроме того, заранее выбрана последовательность опытов (см., например, примеры 11.3 и 11.4), то с помощью условия (11.55) можно до начала эксперимента определить требуемое число опытов.

Пример 11.5

Пусть модель исследуемого объекта имеет вид

$$y(a, x) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2. \qquad (11.56)$$

Для нахождения оценок коэффициентов предполагается последовательно реализовывать опыты в соответствии с табл. 11.4 до тех пор, пока с вероятностью $1 - \alpha = 0{,}975$ истинные значения коэффициентов не окажутся внутри окружности радиуса $\rho = 2{,}5$. При этом уравнение границы допустимой области (11.41) имеет вид

$$(a - \hat{a}^N)' (a - \hat{a}^N) = \rho^2,$$

т. е. $B = I_3$. Требуется определить необходимое число опытов N, если известна оценка дисперсии ошибок наблюдения $s^2 = 0{,}27$ с числом степеней свободы $\varphi_e = 5$.

Из (11.55) и (11.43) следует, что в момент останова необходимо, чтобы было

$$\lambda_{N\,\min} \geqslant \frac{k+1}{\rho^2} s^2 F_{\text{кр}} = \frac{6}{6{,}25} \times 0{,}27 \times 6{,}977 = 1{,}81. \qquad (11.57)$$

Здесь $F_{\text{кр}} = 6{,}977$ взято из таблиц F-распределения при уровне значимости $\alpha = 0{,}025$ и $\varphi_1 = k + 1 = 6$; $\varphi_2 = \varphi_e = 5$. В табл. 11.4 нумерация точек соответствует фиг. 6.3, *а*. Первые 9 точек табл. 11.4 — это точки начального плана, включающего по одному разу точки непрерывного D-оптимального плана. В таблице, кроме того, приведены значения $\lambda_{N\,\min}$ минимальных собствен-

Таблица 11.4

N	Номер точки плана	$\lambda_{N\min}$	N	Номер точки плана	$\lambda_{N\min}$	N	Номер точки плана	$\lambda_{N\min}$
1	1	—	10	4	1,031	19	6	1,827
2	2	—	11	1	1,067	20	8	1,985
3	3	—	12	3	1,1			
4	4	—	13	2	1,238			
5	5	—	14	5	1,717			
6	6	—	15	9	1,744			
7	7	—	16	1	1,762			
8	8	—	17	2	1,781			
9	9	—	18	4	1,803			

ных чисел матрицы $F_N'F_N$ после каждого эксперимента (начиная с $N=10$). Нетрудно видеть, что условие (11.57) выполняется, начиная с 19-го опыта. Таким образом, для данной задачи достаточно осуществить $N=19$ опытов.

11.5. Последовательное планирование эксперимента в задачах построения нелинейных по параметрам моделей

В этом разделе будет рассмотрена задача последовательного планирования эксперимента для случая, когда искомые коэффициенты a входят в модель $y(a, x)$ нелинейно [11, 12]. Задачи вычисления по экспериментальным данным оценок нелинейных коэффициентов часто возникают тогда, когда структура модели выбирается из физико-химических соображений. Примером модели такого типа является, например, следующее уравнение, описывающее кинетику последовательной реакции первого порядка:

$$y(t, a_1, a_2) = \frac{a_1}{a_1 - a_2}\left(e^{-a_2 t} - e^{-a_1 t}\right). \qquad (11.58)$$

Здесь a_1 и a_2 — неизвестные константы скорости реакции, t — время, которое в данном случае является независимой переменной, определяющей условия эксперимента.

Положим, что решается задача определения значений коэффициентов зависимости $y(a, x)$ известного вида, в которую коэффициенты a входят нелинейно. Пусть реализовано N опытов в точках, задаваемых матрицей X_N размера $(N \times n)$. Здесь n — число варьируемых переменных, а N — число реализованных опытов. Результат измерения зависимой переменной в точке x^j обозначим через $\tilde{y}^j$. Предположим, что эти N экспериментов ис-

пользованы для нахождения оценок $\hat{a}^N$ в соответствии с критерием

$$Q(\hat{a}^N) = \min_a Q(a) = \min_a \sum_{j=1}^{N} [\tilde{y}^j - y(x^j, a)]^2. \qquad (11.59)$$

Задача нахождения оценок $\hat{a}^N$ из условия (11.59) может быть решена с помощью методов нелинейной оптимизации, описанных в разд. 10. Пусть теперь требуется определить наилучшие (с точки зрения некоторого показателя точности решения) условия проведения последующих m опытов. Приближенное решение этой задачи можно получить, применяя теорию планирования эксперимента, развитую для моделей, линейно зависящих от искомых коэффициентов. Для этого зависимость $y(x, a)$ следует разложить в ряд Тейлора по коэффициентам a в точке $a = \hat{a}^N$ и ограничиться в этом разложении линейными членами

$$y(x, a) \approx y(x, \hat{a}^N) + \frac{\partial y(x, a)}{\partial a_1}\Big|_{a=\hat{a}^N} (a_1 - \hat{a}_1^N) + \ldots + \frac{\partial y(x, a)}{\partial a_k}\Big|_{a=\hat{a}^N} (a_1 - \hat{a}_k^N). \qquad (11.60)$$

Подставляя (11.60) в выражение для критерия $Q(a)$

$$Q(a) = \sum_{j=1}^{N+m} (\tilde{y}^j - y(x^j, a))^2 \qquad (11.61)$$

и вводя следующие обозначения:

$$\tilde{y}^j - y(x^j, \hat{a}^N) = \tilde{z}(x^j, \hat{a}^N), \qquad (11.62)$$

$$\frac{\partial y(x^j, a)}{\partial a_l}\Big|_{a=\hat{a}^N} = f_l(x^j, \hat{a}^N), \qquad (11.63)$$

$$(a_l - \hat{a}_l^N) = b_l, \qquad (11.64)$$

получим

$$Q(\hat{a}^N, a) = \sum_{j=1}^{N+m} (\tilde{z}(x^j, \hat{a}^N) - \sum_{l=1}^{k} b_l f_l(x^j, \hat{a}^N))^2 = \sum_{j=1}^{N+m} (\tilde{z}(x^j, \hat{a}^N) - b^T f(x^j, \hat{a}^N))^2. \qquad (11.65)$$

Здесь b и $f(x^j, \hat{a}^N)$ — векторы-столбцы размерности k.

Для уточнения значений коэффициентов модели теперь необходимо рассчитать оценки $\hat{b}_l$ из условия минимума критерия (11.65). Отметим, что величины $Q(a)$ и $Q(\hat{a}^N, a)$ не идентичны. В выражение для $Q(a)$ входит нелинейная модель, в то время как в $Q(\hat{a}^N, a)$ используется модель, линеаризованная по коэф-

фициентам в точке $\hat{a}^N$. Оценки коэффициентов, полученные при минимизации $Q(\hat{a}^N, a)$, будут давать лишь приближенное решение задачи, отличающееся в общем случае от решения, получаемого путем минимизации $Q(a)$.

Обозначим через $\tilde{Z}_{N+m}$ вектор-столбец размерности $N+m$ с элементами $\tilde{z}(x^j, \hat{a}^N) = \tilde{y}^j - y(x^j, \hat{a}^N)$, а через F_{N+m} — матрицу размера $(N+m)\times k$ с элементами $f_i(x^j, \hat{a}^N)$, $i=1, 2, \ldots, k$, $j=1, 2, \ldots, (N+m)$. Тогда оценки $\hat{b}$ определятся выражением

$$\hat{b} = (F'_{N+m}F_{N+m})^{-1} F'_{N+m}\tilde{Z}_{N+m}. \qquad (11.66)$$

Для выбора наилучшего плана X_m^* постановки m новых опытов воспользуемся критерием D-оптимальности, потребовав, чтобы определитель информационной матрицы M_{N+m} плана X_{N+m} был максимален:

$$|M_{N+m}(X_m^*)| = \max_{X_m \in \Omega_x} |F'_{N+m}F_{N+m}|. \qquad (11.67)$$

Здесь Ω_x — область планирования эксперимента.

Естественно, что вместо условия (11.67) можно искать план X_m^* из требования минимальной величины определителя матрицы $C_{N+m} = (F'_{N+m}F_{N+m})^{-1}$. Эффективность описанной здесь методики планирования эксперимента будет зависеть от близости оценок $\hat{a}^N$ к истинным значениям коэффициентов. Поэтому планирование сразу большого числа m новых опытов может оказаться нецелесообразным. Лучшей стратегией здесь является следующая. Положим, что после проведения i экспериментов рассчитаны оценки $\hat{a}^i$. Тогда для выбора условий последующих опытов необходимо:

1) рассчитать матрицу F_i значений частных производных

$$\left.\frac{\partial y(x^j, a)}{\partial a_\mu}\right|_{a=\hat{a}^i} \quad \text{при} \quad j=1, 2, \ldots, i \quad \text{и} \quad \mu=1, \ldots, k;$$

2) выбрать условия x_*^{i+1} проведения $(i+1)$-го опыта, осуществляя поиск максимума функции

$$|M_{i+1}(x^{i+1})| = |F_i'F_i + f(x^{i+1}, \hat{a}^i) f'(x^{i+1}, \hat{a}^i)|$$

по x^{i+1} в заданной области Ω_x;

3) провести эксперимент в точке x_*^{i+1} и, используя его результаты, рассчитать с помощью методов нелинейной оптимизации оценки $\hat{a}^{i+1}$ из условия

$$Q(\hat{a}^{i+1}) = \min_a Q(a) = \min \sum_{j=1}^{i+1} (\tilde{y}^j - y(x^j, a))^2; \qquad (11.68)$$

Таблица 11.6

Последовательность экспериментальных точек (пример 11.6)

Номер опыта l	x_1^l	x_2^l	
1	0,01	0,01	Начальный план
2	0,03	0,01	
3	0,01	0,03	
4	0,03	0,03	
5	0,05	0,05	
6	0,05	0,02	
7	0,02	0,05	
8	0,05	0,05	
9	0,05	0,02	
10	0,02	0,05	
11	0,05	0,05	
12	0,05	0,02	

4) перейти к п. 1 — расчету матрицы F_{l+1} при новых значениях коэффициентов $a = \hat{a}^{l+1}$ с целью выбора следующей $(l+2)$-й точки.

Пример 11.6 [12]

Исследовалась реакция каталитического восстановления окиси азота

$$NO + H_2 \rightleftarrows H_2O + \frac{1}{2} N_2.$$

Скорость этой реакции можно описать с помощью модели вида

$$y(x, a) = \frac{a_1 a_3 a_0 x_2}{(1 + a_1 x_1 + a_2 x_2)^2}. \qquad (11.69)$$

Здесь y — скорость реакции (г·моль/мин на 1 г катализатора); a_1 и a_2 — равновесные коэффициенты абсорбции окиси азота (NO) и водорода (H_2) соответственно (см^2/кгс); a_3 — константа скорости прямой реакции (г·моль/мин на 1 г катализатора); x_1 и x_2 — парциальные давления окиси азота и водорода соответственно (кгс/см^2).

С целью получения оценок коэффициентов a_1, a_2 и a_3 был использован метод последовательного планирования эксперимента. Область варьирования независимых переменных в данном случае определялась неравенствами

$$0{,}01 \leqslant x_1 \leqslant 0{,}05,$$
$$0{,}01 \leqslant x_2 \leqslant 0{,}05,$$

Для получения начальных оценок коэффициентов был использован полный факторный план с основным уровнем обоих факторов, равным 0,02, и интервалом варьирования, равным 0,01 (см. первые четыре опыта табл. 11.5) [1]. По результатам четырех опытов с помощью методов нелинейного оценивания (см. разд. 10) были рассчитаны оценки $\hat{a}_1^4$, $\hat{a}_2^4$, $\hat{a}_3^4$. Дальнейшие опыты планировались последовательно в соответствии с изложенной выше методикой.

Для выбора условий проведения пятого опыта необходимо рассчитать матрицу F_4 частных производных, элементами которой являются величины $f_l(x^j, \hat{a}^4)$ (здесь l — номер столбца матрицы F_4, j — номер строки):

$$f_l(x^j, \hat{a}^4) = \frac{\partial}{\partial a_l}\left(\frac{a_1 a_3 a_3 x_1^j x_2^j}{(1 - a_1 x_1^j + a_2 x_2^j)^2}\right)\Bigg|_{a=\hat{a}^4}. \qquad (11.70)$$

Очевидно, что

$$f_1(x^j, \hat{a}^4) = \frac{\hat{a}_2^4 \hat{a}_3^4 x_1^j x_2^j (1 + \hat{a}_1^4 x_1^j + \hat{a}_2^4 x_2^j)}{(1 - \hat{a}_1^4 x_1^j + \hat{a}_2^4 x_2^j)^3},$$

$$f_2(x^j, \hat{a}^4) = \frac{\hat{a}_1^4 \hat{a}_3^4 x_1^j x_2^j (1 - \hat{a}_1^4 x_1^j - \hat{a}_2^4 x_2^j)}{(1 - \hat{a}_1^4 x_1^j + \hat{a}_2^4 x_2^j)^3},$$

$$f_3(x^j, \hat{a}^4) = \frac{\hat{a}_1^4 \hat{a}_2^4 x_1^j x_2^j}{(1 - \hat{a}_1^4 x_1^j + \hat{a}_2^4 x_2^j)^2}.$$

Информационная матрица плана после пятого опыта имеет вид

$$M_5(x^5) = \begin{bmatrix} \sum_{j=1}^{5} f_1^2(x^j, \hat{a}^4) & \sum_{j=1}^{5} f_1(x^j, \hat{a}^4) f_2(x^j, \hat{a}^4) & \sum_{j=1}^{5} f_1(x^j, \hat{a}^4) f_3(x^j, \hat{a}^4) \\ \sum_{j=1}^{5} f_2(x^j, \hat{a}^4) f_1(x^j, \hat{a}^4) & \sum_{j=1}^{5} f_2^2(x^j, \hat{a}^4) & \sum_{j=1}^{5} f_2(x^j, \hat{a}^4) f_3(x^j, \hat{a}^4) \\ \sum_{j=1}^{5} f_3(x^j, \hat{a}^4) f_1(x^j, \hat{a}^4) & \sum_{j=1}^{5} f_3(x^j, \hat{a}^4) f_2(x^j, \hat{a}^4) & \sum_{j=1}^{5} f_3^2(x^j, \hat{a}^4) \end{bmatrix}$$

[1] В данном примере наблюдения имитировались путем расчета по модели (11.69) при значениях коэффициентов $a_1 = 14{,}64$ см³/кгс, $a_2 = 19{,}00$ см³/кгс, $a_3 = 4{,}94 \cdot 10^{-4}$ г·моль/мин на 1 г катализатора. К результату расчета добавлялась случайная нормально распределенная ошибка с нулевым средним и дисперсией, равной $9{,}508 \cdot 10^{-12}$.

Таблица 11.6

Значения $|M_6(x^6)| \times 10^{27}$ при поиске шестой точки (пример 11.6)

x_1 \ x_2	0,01	0,02	0,03	0,04	0,05
0,05	18,4	20,4	13,8	8,4	7,1
0,04	11,1	10,2	6,4	5,3	7,4
0,03	6,8	5,9	5,1	6,7	10,0
0,02	5,1	5,4	6,4	8,5	11,1
0,01	4,4	5,1	6,0	6,9	7,8

Так как точки проведения первых четырех наблюдений зафиксированы (см. табл. 11.5), то величина определителя матрицы $M(x^5)$ зависит только от координат пятого эксперимента x^5. Поиск значений x_1^5 и x_2^5, в которых достигается максимум определителя матрицы $M(x^5)$, осуществлялся методом сетки. В результате была найдена точка с координатами $x_1^5 = 0{,}05$; $x_2^5 = 0{,}05$. Результат наблюдения в этой точке $\tilde{y}^5$ был использован для определения оценок $\hat{a}_1^5$, $\hat{a}_2^5$, $\hat{a}_3^5$ путем минимизации критерия

$$Q(a) = \sum_{j=1}^{5} (\tilde{y}^j - y(a, x^j))^2.$$

Далее аналогично вышеизложенному была вычислена матрица F_5 частных производных при $a = \hat{a}^5$.

Точка постановки шестого опыта определялась из условия достижения максимума определителя матрицы $M_6(x^6)$. В табл. 11.6 приведены значения определителя информационной матрицы, полученные при различных значениях координат x_1 и x_2 шестой точки. Очевидно, что максимум достигается здесь в точке с координатами $x_1^6 = 0{,}05$; $x_2^6 = 0{,}02$. Эта точка и была выбрана для постановки шестого опыта. Точки постановки последующих опытов, полученные аналогичным образом, приведены в табл. 11.5.

После реализации $N = 12$ опытов были получены следующие оценки коэффициентов модели:

$$\hat{a}_1 = 10{,}8 \pm 5{,}5,$$
$$\hat{a}_2 = 20{,}7 \pm 6{,}0,$$
$$\hat{a}_3 = 5{,}75 \pm 2{,}1.$$

Здесь, помимо величин оценок, указаны границы доверительных интервалов для коэффициентов, соответствующие доверительной вероятности $P = 0{,}95$. Заметим, что дисперсии оценок коэффи-

цнентов равны произведению соответствующего диагонального элемента матрицы $C_N = M_N^{-1} = (F_N' F_N)^{-1}$, где N — число реализованных опытов, на оценку дисперсии ошибки наблюдений, в качестве которой в данном примере использовалась остаточная дисперсия. Элементы матрицы F_N определяются согласно (11.63), где $\hat{a}^N$ — полученные в результате исследования оценки коэффициентов.

ЛИТЕРАТУРА

1. Федоров В. В., Последовательные методы планирования экспериментов при изучении механизма явлений, см. сб. „Новые идеи в планировании эксперимента", изд-во „Наука", М., 1969.
2. Ермуратский П. В., Круг Г. К., Лецкий Э. К., Методологические вопросы исследования и управления сложными технологическими процессами, см. сб. „Идентификация и аппаратура для статистических исследований", изд-во „Наука", М., 1970.
3. Lee R. C. K., Optimal Estimation, Identification and Control, Research monogr. № 28, M.I.T. Press, Cambridge, Mass.; русский перевод: Ли Р. Оптимальные оценки, определение характеристик и управление, изд-во „Наука", М., 1966.
4. Федоров В. В., Теория оптимального эксперимента, изд-во „Наука", М., 1971.
5. Plackett R. W., Some Theorems in Last Squares, *Biometrika*, **6**, № 1, 149 (1950).
6. Круг Г. К., Коршунов М. А., Лазарянц В. Э., Лисенков А. Н., Оптимизация одного химического процесса методом последовательного планирования эксперимента, Доклады научно-технической конференции по итогам научно-исследовательских работ за 1966—1967 гг., подсекция Автоматики и телемеханики, ч. I, изд-во МЭИ, М., 1967.
7. Гантмахер Ф. Р., Теория матриц, изд-во „Наука", М., 1967.
8. Вучков И. Н., Ермуратский П. В., Круг Г. К., Лецкий Э. К., Никифорова Е. С., Об одном алгоритме получения D-оптимальных планов, см. Труды МЭИ, вып. 68, изд-во МЭИ, М., 1969.
9. Лецкий Э. К., Вучков И. Н., Об одном способе последовательной идентификации, *Техническая кибернетика*, № 2, 1970.
10. Box G. E., Lucas H. L., Design of Experiments in Non-Linear Situations, *Biometrica*, **46**, 77—90 (1959).
11. Box G. E., Hunter W. G., Sequential Design of Experiments for Nonlinear Models, IBM Scientific Computing Symp. in Statistics, Okt. 1965, pp. 113—137.
12. Kittrell J. R., Hunter W. G., Natson C. C., Obtaining Precise Parameter Estimates for Nonlinear Catalytic Rate Models, *AIChE. J.*, **12**, № 1 (1966).
13. Himmelblau D., Process Analysis by Statistical Methods, Willey, New York, 1970; русский перевод: Химмельблау Д., Анализ процессов статистическими методами, изд-во „Мир", М., 1973.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Лецкий Э. К., О некоторых задачах управления экспериментом с использованием автоматизированных систем обработки информации, Информ. бюлл. Совета по автоматизации научных исследований, № 3, М., 1971.

2. Atkinson A. C., Hunter W. G., The Design of Experiments for Parameter Estimation, *Technometries*, **10**, 271 (1968).
3. Draper N. R., Hunter W. G., The Use of Prior Distribution in the Design of Experiments for Parameter Estimation in Nonlinear Situations, *Biometrika*, **54**, 147 (1967).
4. Gray K. B., Sequential Selection of Experiments, *Ann. Math. Stat.*, **39**, № 6 (1968).
5. Божанов Е., Вучков И., Статистически методи за моделиране и оптимизиране на многофакторни обекти, „Техника", София, 1973.
6. Борчанинов М. Г., Никифорова Е. С., Фомин Г. А., Алгоритмы управления экспериментом в системах автоматизации научных исследований, см. сб. „Планирование и автоматизация эксперимента в научных исследованиях", изд-во „Советское радио", М., 1974.
7. Игонина Н. В., Борисова А. П., *Заводск. лаб.*, № 5 (1974).

12.

МЕТОДЫ ЭКСПЕРИМЕНТАЛЬНОЙ ОПТИМИЗАЦИИ

12.1. Постановка задачи

Экономические и технические причины обычно заставляют стремиться к тому, чтобы процессы, установки, аппараты и прочие устройства работали при оптимальных условиях, а производимая продукция обладала оптимальными свойствами. Это означает, например, что для химического реактора следует устанавливать такие значения температуры, давления, концентрации компонентов химической реакции, при которых реакция протекает с максимальным выходом целевого продукта. Подобно этому можно потребовать, чтобы установка действовала с максимальной прибылью или минимальными эксплуатационными расходами. Часто требуется, чтобы продукция обладала определенными оптимальными свойствами, например имела бы как можно большую прочность или чистоту; к специальным строительным материалам может быть предъявлено требование высокой звукопоглощаемости, к катализаторам — наибольшей активности и избирательности и т. д. Другими словами, определенные качественные показатели продукции должны быть наилучшими. Целевые величины, такие, как пропускная способность, производительность, доход, часто связаны с величинами, характеризующими качественные показатели. При этом одна из них (целевая величина или характеристика качества) выступает в виде ограничения. Например, можно потребовать максимального выхода продукта при заданном качестве.

Необходимость решения рассматриваемых здесь задач возникает при различных формах рационализации производства, таких, как интенсификация производственных процессов, улучшение качества продукции, устранение узких мест, увеличение выпуска продукции, а также при выборе наилучших режимов для заново проектируемых производственных единиц, лабораторных и полупромышленных установок.

Для решения задачи нахождения оптимума выбранной целевой величины в соответствии с конкретными условиями и требованиями применимы два принципиально различных подхода:

1. Определение оптимальных условий с помощью модели объекта.

Здесь предполагается, что имеется статистическая (экспериментальная) модель, т. е. полученное с помощью эксперимента математическое описание объекта.

Теоретические модели, т. е. модели, которые адекватно отражают физико-химические взаимосвязи процессов, в этой главе рассматриваться не будут. Для оптимизации при наличии таких моделей имеются специальные методы, описанные, например, в работе [1].

2. Нахождение оптимальных условий непосредственно на объекте без использования модели.

Этот метод предполагает существование объекта и возможность проводить на объекте эксперимент. Такой подход поэтому не может быть использован при создании новых процессов до их технической реализации.

Оба метода используют (первый явно, второй неявно) понятие функции отклика:

$$y = f(x_1, x_2, \ldots, x_n), \tag{12.1}$$

где y — целевая (выходная) величина, x_i — влияющие переменные (факторы). Выбором значений факторов необходимо обеспечить экстремальное значение целевой величины

$$y = \text{extr}. \tag{12.2}$$

Между экспериментами на лабораторных и промышленных установках имеется ряд существенных различий, к которым относятся:

— малые возможности варьирования режимов в крупных промышленных установках из-за опасений нарушения качественных показателей продукции;

— относительно высокий уровень „шума“ из-за большого числа неконтролируемых факторов в промышленных условиях;

— применение высокоточной измерительной аппаратуры в лабораторных условиях в отличие от неточных технических измерителей в промышленных установках;

— дрейф характеристик объекта в промышленных установках.

Эти различия привели к разработке специальных методов экспериментального поиска оптимума в промышленных условиях, таких, как метод эволюционного планирования (ЭВОП) и симплексный метод. Оба эти метода работают без использования модели.

Для получения максимальной информации о положении оптимума с помощью небольшого числа опытов применяются специальные методы и стратегии планирования эксперимента. К ним относится, например, метод крутого восхождения. В предыдущих разделах рассматривались экспериментальные планы, используемые при реализации этого метода. В этом разделе,

кроме метода крутого восхождения, будут описаны также последовательный симплексный метод и метод случайного поиска.

Для исследования функции в окрестности точки оптимума развиты специальные методы, которые кратко излагаются в данном разделе.

Прежде чем перейти к рассмотрению многомерных задач, составляющих основное содержание данного раздела, опишем предварительно некоторые методы, разработанные для одномерного случая.

12.2. Экспериментальная оптимизация в случае одной переменной

Существует ряд способов выбора координат точек проведения эксперимента x^i в одномерном случае. Простейший способ — эквидистантные планы. В этих планах точки проведения опытов отделены одна от другой равными расстояниями. Для случая, когда экстремум целевой функции должен быть определен по возможности точнее с помощью относительно малого числа опытов, разработаны специальные планы. При применении этих планов к виду исследуемой функции $y = f(x)$ предъявляется требование, чтобы она имела только один экстремум (унимодальность), а величина y определялась без ошибки. При этом интервал поиска ограничен: $a \leqslant x \leqslant b$. Сокращение начального интервала осуществляется путем сравнения значений целевой функции в двух точках, выбираемых соответствующим образом. В качестве меры эффективности для плана эксперимента используется отношение величины начального интервала I_0 к остаточному интервалу Δx_N после N опытов, внутри которого находится экстремум целевой функции

$$E = \frac{b-a}{\Delta x_N}.$$

Эквидистантные планы

N точек постановки экспериментов на интервале $[a, b]$ располагаются на равных расстояниях одна от другой. Опыты проводятся при значениях

$$x^i = a + \frac{b-a}{N-1}, \qquad i = 0\,(1)\,N. \tag{12.3}$$

Для меры эффективности имеем $E = N - 1$.

Метод последовательной дихотомии

Каждые два опыта располагаются в центре интервала варьирования как можно ближе друг к другу:

$$x_1 = \frac{a+b+e}{2}, \qquad x_2 = \frac{a+b-e}{2}, \qquad x_1 > x_2, \tag{12.4}$$

где ε — малая величина. Путем сравнения значений функций $f(x_1)$ и $f(x_2)$ решается, в какой части области варьирования процесс деления следует продолжить. Мера эффективности равна $E = 2^{N/2}$.

Планы, использующие числа Фибоначчи

Эффективное планирование достигается применением чисел Фибоначчи, которые определяются последовательностью

$$F_0 = F_1 = 1, \quad F_k = F_{k-1} + F_{k-2}, \qquad k > 1. \tag{12.5}$$

Планирование эксперимента осуществляется следующим образом: координаты первых двух экспериментов устанавливаются равными

$$x_1 = a + (b-a)\,q, \quad x_2 = b - (b-a)\,q \quad \text{при} \quad q = \frac{F_{N-2}}{F_N}, \tag{12.6}$$

где N выбирается достаточно большим. Использование результатов опытов в этих двух точках позволяет уменьшить длину I_0 интервала поиска в

$$1 - q = \frac{F_{N-1}}{F_N} \tag{12.7}$$

раз. Эта процедура применяется затем к остаточному интервалу при

$$q = \frac{F_{N-2}}{F_{N-1}},$$

причем здесь нужно проводить только один новый опыт, так как на этом шаге может быть использован опыт, поставленный на предыдущем шаге и расположенный внутри остаточного интервала. Фактор уменьшения длины интервала при этом принимает значение

$$1 - q = \frac{F_{N-3}}{F_{N-1}}.$$

Мера эффективности метода составляет

$$E = \left(\frac{F_0}{F_N}\right)^{-1} = \left(\frac{1}{F_N}\right)^{-1}. \tag{12.8}$$

Так как при $N = 11$ $F_N = 144$, то с помощью 11 экспериментов можно зафиксировать окончательную область, не превышающую 1% размера начальной области поиска. Недостаток этого метода состоит в том, что до начала исследования экспериментатор должен задать число опытов, исходя из желаемой точности результата.

Золотое сечение

Метод золотого сечения является частным случаем метода Фибоначчи, в котором вместо величины

$$q = \frac{F_{k-2}}{F_k}$$

на каждом шаге используется ее предельное значение при $k \to \infty$:

$$q = \lim_{k \to \infty} \frac{F_{k-2}}{F_k} = \left(\frac{2}{1+\sqrt{5}}\right)^2 = \frac{3-\sqrt{5}}{2} = 0{,}382. \qquad (12.9)$$

Преимущество метода состоит в том, что число опытов не обязательно должно определяться заранее. Из-за этого эффективность оказывается несколько меньшей. После проведения N опытов остаточный интервал составляет

$$\left(\frac{2}{1+\sqrt{5}}\right)^{N-1} = (0{,}618)^{N-1} \qquad (12.10)$$

часть исходного.

Решетчатый метод поиска

Этот метод является частным случаем метода Фибоначчи, когда варьируемая переменная изменяется не непрерывно, а может принимать лишь конечное множество дискретных значений. С помощью N опытов можно из множества, состоящего из $k_N = F_{N+1} - 1$ точек, выделить точку с наибольшим значением функции.

Пример 12.1 [3]

С помощью плана, использующего числа Фибоначчи, требуется выбрать время пребывания вещества в смесителе, оптимальное с точки зрения выхода целевого продукта. Исходная область $[a, b]$ составляет $[0, 1200]$, остаточный интервал Δx_N, в котором находится экстремум, должен составлять 1/10 исходного интервала, т. е. $F_N = 1200/120 = 10$. Ближайшее подходящее число Фибоначчи есть $F_6 = 13$. Следовательно, необходимо провести шесть опытов. Для получения координат первых двух наблюдений $x^1 = a + (b-a)$ и $x^2 = b(b-a)q$ рассчитываем

$$q_2 = \frac{F_{N-2}}{F_N} = \frac{5}{13}$$

и получаем $x^1 = 462$ и $x^2 = 1200 - 462 = 738$. После реализации этих опытов сравниваем $\tilde{y}^1$ и $\tilde{y}^2$. Сопоставление результатов дает $\tilde{y}^1 > \tilde{y}^2$, при этом интервал укорачивается до $[0, 738]$. Для получения координат точки постановки третьего опыта x^3 рассчитываем

$$q_3 = \frac{F_{N-3}}{F_{N-1}} = \frac{3}{8}$$

и получаем $x^3 = (a - x^2)\, q_3 = 276$. В результате реализации третьего опыта получаем $\tilde{y}^3 < \tilde{y}^1$ и интервал сокращается до размера [276, 738]. Дальнейшие результаты и общая схема эксперимента представлены в табл. 12.1 [1].

Таблица 12.1

Номер эксперимента i	Границы интервала I_i	$I_i q_i = I_i \frac{F_{N-i}}{F_{N-i+2}}$	x^i	Соотношение результатов эксперимента
0	0—1200	—	—	—
1	0—1200	462	462	—
2	0—1200	462	738	$\tilde{y}_1 > \tilde{y}_2$
3	0— 738	276	276	$\tilde{y}_1 > \tilde{y}_3$
4	276— 738	174	450	$\tilde{y}_4 \leqslant \tilde{y}_1$
5	450— 738	90	540	$\tilde{y}_5 = \tilde{y}_4$
6	450— 540	Половина интервала	495	$\tilde{y}_6 > \tilde{y}_5$ $\tilde{y}_6 > \tilde{y}_4$

12.3. Метод Бокса — Уилсона (метод крутого восхождения)

Стратегия метода Бокса—Уилсона при исследовании поверхности отклика с целью нахождения экстремума состоит в следующем. На основе малой серии опытов находится локальное описание поверхности отклика в некоторой исходной области с помощью модели линейного вида. В центре области рассчитывается линейное приближение градиента, и в направлении градиента, т. е. в направлении наискорейшего подъема (крутого восхождения), проводятся опыты до достижения стационарной области, в которой расположен экстремум. Если найденное линейное приближение градиента для центра области существенно отличается от значения градиента в некоторой точке по направлению движения, то можно найти новое описание поверхности отклика с помощью полиномов первого порядка в окрестности этой точки и рассчитать новое значение градиента. В стационарной области линейная модель оказывается уже не адекватной, и для описания поверхности отклика в этой области, как правило, используются полиномы второй и даже третьей степени. В стационарной области требуется провести большее число экспериментов, так как здесь необходимо получить возможно более точное описание. Стратегия метода Бокса—Уилсона для

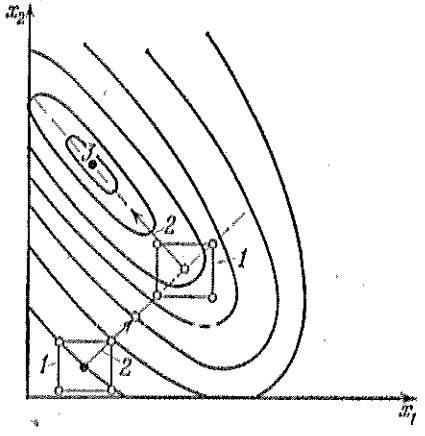
[1] В табл. 12.1 через I_i обозначена величина исходного интервала (интервала неопределенности) перед i-м опытом.—*Прим. ред.*

двумерного случая представлена на фиг. 12.1. При применении метода необходимо обращать внимание на то, что градиент не инвариантен к изменению масштаба переменных.

Вектор-градиент непрерывной функции $y = f(x_1, x_2, \ldots, x_n)$ имеет вид

$$\operatorname{grad} y = \left(\frac{\partial f}{\partial x_1} i + \frac{\partial f}{\partial x_2} j + \ldots \right). \tag{12.11}$$

Этот вектор перпендикулярен поверхности равного уровня $y = \text{const}$ и указывает направление наискорейшего подъема (кру-



Фиг. 12.1. Стратегия метода Бокса—Уилсона.
1 — план эксперимента для линейной модели; *2* — линейное приближение градиента; *3* — стационарная область.

того восхождения) (фиг. 12.1). В (12.11) i, j — единичные векторы по осям переменных x. Если модель линейная, т. е.

$$y = a_0 x_0 + a_1 x_1 + \ldots a_n x_n, \tag{12.12}$$

то коэффициенты a_i являются координатами вектор-градиента. Если изменять факторы пропорционально найденным оценкам коэффициентов $\hat{a}_i$, то реализуется движение в направлении наискорейшего приближения к оптимуму. Компоненты градиента и нормированного градиента

$$\frac{\operatorname{grad} y}{|\operatorname{grad} y|} = \frac{\frac{\partial f}{\partial x_1} i + \frac{\partial f}{\partial x_2} j + \ldots}{\left[\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 \right]^{1/2}} \tag{12.13}$$

экспериментально полученной линейной модели зависят от выбора основного уровня и интервалов варьирования переменных.

Необходимо стремиться так выбирать интервалы варьирования, чтобы величины коэффициентов (в случае их значимости) имели бы одинаковый порядок. В этом случае применение метода крутого восхождения является наиболее эффективным.

Метод Бокса—Уилсона состоит из следующих шагов:

1. Построение линейной модели. Предполагается, что эксперимент начинается с точки, достаточно удаленной от оптимума (или так называемой стационарной области); поэтому для описания поверхности отклика в окрестности начальной точки можно использовать линейную модель. С помощью факторного планирования можно найти точки проведения наблюдений и рассчитать оценки коэффициентов. Статистический анализ результатов завершает этот этап.

2. Пошаговое улучшение величины целевой функции. Чтобы осуществить движение в направлении оптимума, необходимо рассчитать направление градиента и определить величину шага до следующего опыта. Координаты l-й точки наблюдения по направлению градиента находятся по формуле

$$x_i^l = x_i^{l-1} + \lambda \hat{a}_i, \quad l = 1\,(1)\,L, \tag{12.14}$$

где $\hat{a}_i$ — оценка i-го коэффициента [1]. Так как градиент всегда проходит через центр плана $x_i^0 = 0$, то

$$x_i^l = l\lambda \hat{a}_i, \tag{12.15}$$

где λ — выбранная величина нормированного шага. При выборе величины шага имеют значение соображения, близкие к тем, которые используются при выборе интервала варьирования. Величину шага для k-го фактора λ_k выбирают из физических соображений (целесообразно, чтобы этим фактором была переменная, в наибольшей степени влияющая на целевую функцию), выражая ее в натуральных единицах (°С, секунды, концентрация и т. д.); затем рассчитывают

$$\lambda = \frac{\lambda_k}{\Delta_k |\hat{a}_k|}, \tag{12.16}$$

где Δ_k — интервал варьирования $[\Delta_k = {}^1/_2 (x_{k\,\text{макс}} - x_{k\,\text{мин}})]$. Для определения величины шага λ_i по остальным переменным в натуральных единицах используется следующее соотношение:

$$\lambda_i = \lambda |\hat{a}_i| \Delta_i. \tag{12.17}$$

При этом получается, что величины шагов λ_i всегда соотносятся

[1] Здесь и ниже λ — нормированная величина шага движения в направлении градиента. Кроме того, через x_i обозначены нормированные значения факторов, а через x_i^* — значения факторов в натуральных единицах. — *Прим. ред.*

друг с другом так же, как модули величин $\hat{a}_l\Delta_l$, т. е.

$$\lambda_1 : \lambda_2 : \ldots : \lambda_n = |\hat{a}_1\Delta_1| : |\hat{a}_2\Delta_2| : \ldots : |\hat{a}_n\Delta_n|. \qquad (12.18)$$

Пример 12.2 [5]

В табл. 12.2 для двумерного случая указаны условия проведения эксперимента, план эксперимента (типа 2^2) и результаты наблюдений, полученные путем вычисления среднего по двум опытам в каждой точке плана. Требуется рассчитать величину шага и координаты дальнейших опытов в направлении крутого восхождения, приняв величину шага по переменной x_2^* равной 0,5. Результаты расчетов приводятся в нижней части табл. 12.2[1]. Здесь через x_l^* обозначены факторы в натуральных единицах.

Таблица 12.2

План и результаты эксперимента (пример 12.2)

Переменные	x_1^*	x_2^*	$\bar{y}$
Основной уровень	1,5	7,0	
Интервал варьирования	0,5	1,0	
Верхний уровень	2,0	8,0	
Нижний уровень	1,0	6,0	
Нормированные значения факторов (переменных)	x_1	x_2	
Опыты: 1	−1	−1	95,0
2	+1	−1	90,0
3	−1	+1	85,0
4	+1	+1	82,0
$\hat{a}_l$ ($\hat{a}_0 = 88{,}0$)	−2,0	−4,5	
$\hat{a}_l \times$ Интервал варьирования	−1,0	−4,5	
Шаг при изменении x_2^* на 0,5	−0,11	−0,5	
Округленная величина шага	−0,1	−0,5	
Опыты в направлении крутого восхождения	x_1^*	x_2^*	$\bar{y}$
5	1,4	6,5	
6	1,3	6,0	
7	1,2	5,5	95
8	1,1	5,0	98,6
9	1,0	4,5	

[1] В табл. 12.2 шаг указан со знаком, определяющим направление изменения соответствующего фактора.—*Прим. ред.*

Важную роль в методе крутого восхождения играют „мысленные" эксперименты. Мысленным экспериментом называют расчет выходной величины по модели в точках, расположенных в направлении градиента. На основе этих расчетов исследуется поведение поверхности отклика в направлении градиента.

Если модель адекватна, то реализацию мысленных экспериментов начинают обычно с точки, лежащей вне области, ограниченной точками начального плана. В примере 12.2 это опыт под номером 7. Величина выходной переменной, рассчитанная в этой точке с помощью модели, приведена в табл. 12.2. При реализации опытов в направлении градиента могут быть использованы две различные стратегии: в первом случае все выбранные для реализации опыты проводятся одновременно, во втором — последовательно по некоторой программе. Одновременную реализацию всех опытов проводят обычно тогда, когда исследуемый процесс подвержен временно́му дрейфу, а реализация каждого опыта не требует больших временны́х или материальных затрат. В случае когда постановка опытов связана с большими временны́ми или материальными затратами, предпочтительнее последовательная стратегия.

Пример 12.3 [5]

С помощью метода крутого восхождения требуется увеличить выход целевого продукта некоторого химического процесса. Этот выход зависит от следующих переменных: x_1^* — отношение количества растворителя к количеству основного вещества, x_2^* — температура, x_3^* — длительность реакции. Уровни варьирования факторов, матрица плана, результаты эксперимента и рассчитанные значения коэффициентов приведены в табл. 12.3. Нижняя часть таблицы содержит величины шагов, координаты мысленных опытов в направлении градиента и результаты реализованных опытов.

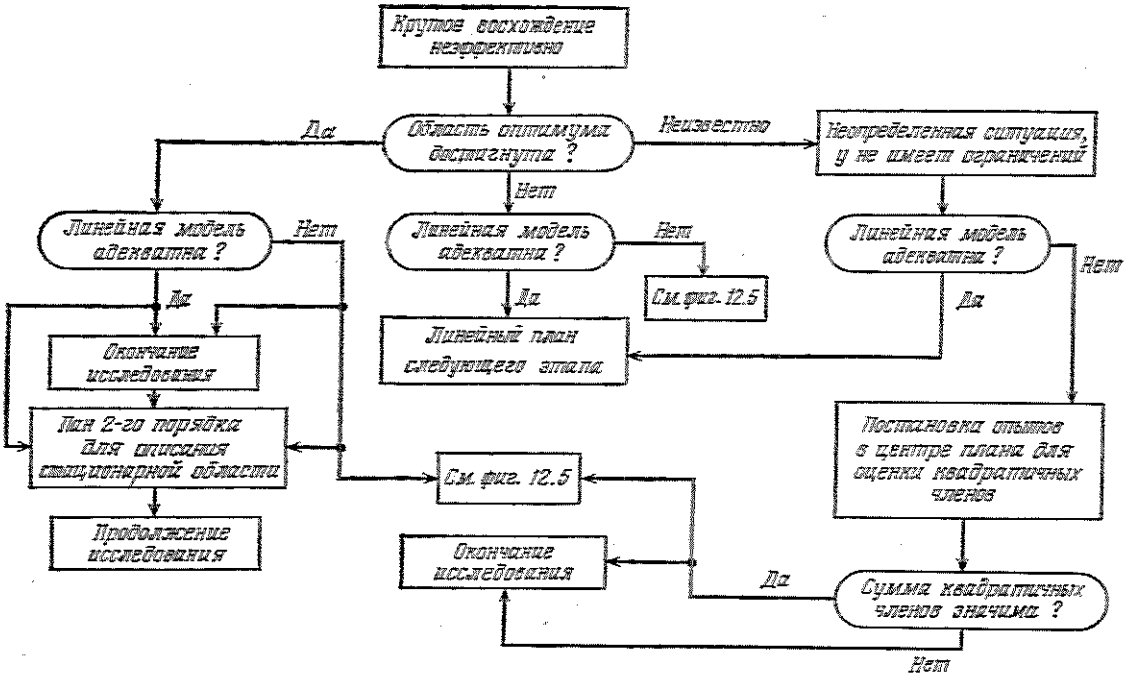
Опыт 9 расположен внутри области, ограниченной точками начального плана; опыты 10 и 12 проводятся мысленно, а опыты 11, 13 и 14 реализуются на исследуемой установке. Результаты экспериментов показывают, что стратегия крутого восхождения оказалась в данном случае весьма эффективной: если в начальной области наибольший выход продукта составлял 46,8% (опыт 1), то в точке с номером 13 была достигнута величина выхода, равная 72,5%. При реализации опытов был использован метод „ножниц": сначала были проведены два опыта в точках, расположенных на концах последовательности мысленных экспериментов (опыты 11 и 14), а затем осуществлен эксперимент в точке с номером 13 с целью проверки характера изменения целевой функции между первыми двумя точками.

Применение метода крутого восхождения считается эффективным, если по крайней мере один из реализованных опытов приведет к лучшему результату по сравнению с точками начального плана. На фиг. 12.2—12.5 представлены возможные решения, принимаемые после окончания движения в направлении градиента в случаях, когда метод крутого восхождения оказывается эффективным или неэффективным, а линейная модель адекватно или неадекватно описывает поверхность отклика [5].

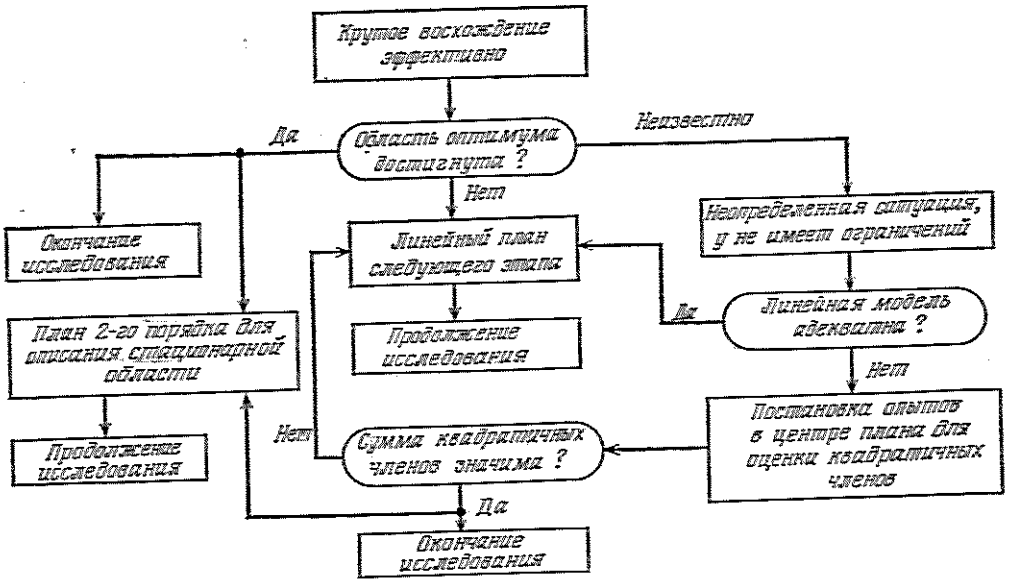
Таблица 12.3

План и результаты эксперимента (пример 12.3)

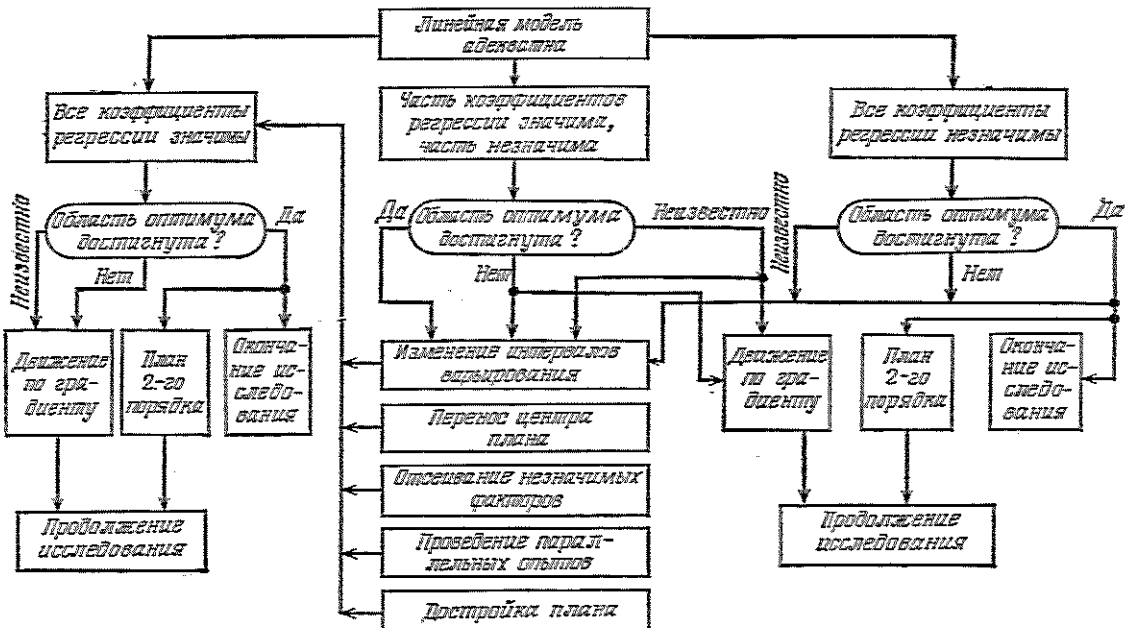
Переменные	x_1^*	x_2^*	x_3^*	$\tilde{y}$
Основной уровень	0,7	135	30	
Интервал варьирования	0,2	5	15	
Верхний уровень	0,9	140	45	
Нижний уровень	0,5	130	15	
Нормированные значения факторов	x_1	x_2	x_3	
Опыты: 1	+1	+1	+1	46,80
2	+1	−1	+1	20,47
3	−1	+1	+1	16,80
4	−1	−1	−1	5,08
5	+1	+1	−1	24,15
6	+1	−1	−1	8,80
7	−1	+1	−1	16,63
8	−1	+1	+1	46,45
$\hat{a}_i$ ($\hat{a}_0 = 23{,}28$)	1,78	10,23	9,36	
	$\hat{a}_{12} = 0{,}17$;	$\hat{a}_{23} = 3{,}77$;		
	$\hat{a}_{13} = -0{,}79$;	$\hat{a}_{123} = 1{,}00$		
$\hat{a}_i \times$Интервал варьирования	0,356	51,40	140,40	
Шаг при изменении x_2^* на 5	0,0346	5	13,60	
Округленная величина шага	0,03	5	14,00	
Опыты в направлении крутого восхождения	x_1^*	x_2^*	x_3^*	
9	0,73	140	44	—
10	0,76	145	58	—
11	0,79	150	72	66,70
12	0,82	155	86	—
13	0,85	160	100	72,50
14	0,88	165	114	68,40



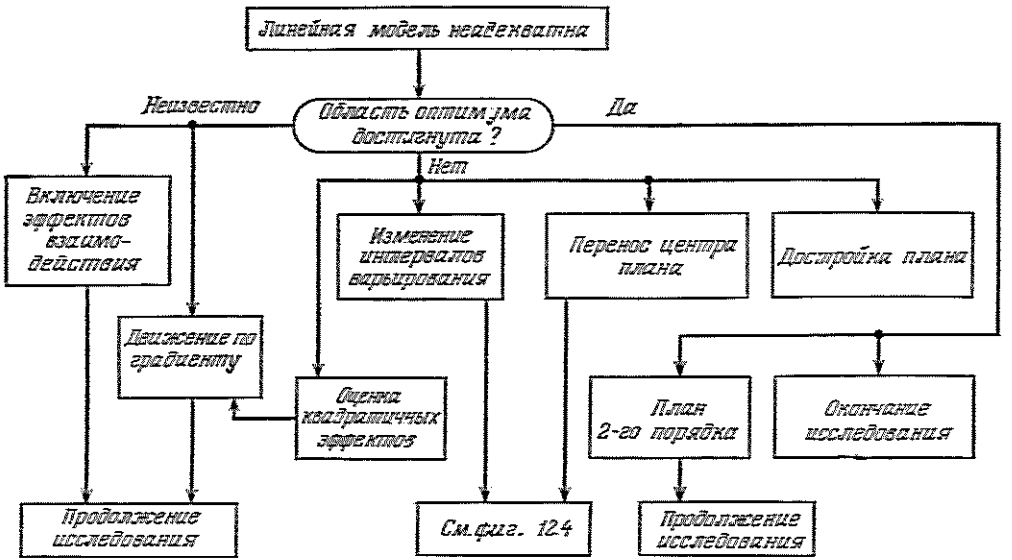
Фиг. 12.2. Принятие решений после крутого восхождения в случае, когда крутое восхождение неэффективно.



Фиг. 12.3. Принятие решений после крутого восхождения в случае, когда крутое восхождение эффективно.



Фиг. 12.4. Принятие решений в случае, когда линейная модель адекватна.



Фиг. 12.5. Принятие решений в случае, когда линейная модель неадекватна.

Обсуждение и обоснование вариантов решений читателю предлагается провести самостоятельно.

Если стационарная область достигнута, то для описания поверхности отклика в этой области необходимо использовать полином по крайней мере второй степени. С планами эксперимента для таких полиномов (например, ортогональные и ротатабельные центральные композиционные планы второго порядка) мы познакомились в разд. 5.

В случае когда стационарную область не удается адекватно описать полиномом второго порядка, может быть использован полином третьего порядка. Соответствующие этому случаю ротатабельные планы третьего порядка описаны, например, в [6].

Пример 12.4 [7]

Выход некоторого продукта зависит от следующих факторов:
x_1^* — температура;
x_2^* — входная концентрация компонента реакции;
x_3^* — расход.

С помощью метода крутого восхождения достигнута стационарная область. Для описания этой области был использован полином второй степени и применен ротатабельный центральный композиционный план второго порядка (разд. 5.6). Значения независимых переменных, план эксперимента и результаты представлены в табл. 12.4. Обработка результатов эксперимента привела к следующей модели:

$$\hat{y}_1 = 83{,}70 - 0{,}49x_1 + 7{,}63x_2 + 1{,}38x_3 - 0{,}21x_1x_2 - 4{,}71x_1x_3 - 0{,}94x_2x_3 - 4{,}41x_1^2 - 7{,}60x_2^2 - 0{,}5x_3^2.$$

Анализ модели

Чтобы получить координаты экстремума целевой функции, необходимо приравнять нулю производные по факторам полинома второй степени и решить полученную систему линейных уравнений. Как правило, исследователя интересуют не только координаты оптимума, но и вид поверхности отклика. Исследование формы поверхности отклика может быть проведено с помощью следующих двух методов:

1. Путем приведения уравнения поверхности отклика к каноническому виду.

2. С помощью метода гребневого анализа (Kammlinienanalyse), предложенного Хёрлем [8].

Таблица 12.4

План и результаты эксперимента (пример 12.4)

Переменные	x_1^*	x_2^*	x_3^*	$\tilde{y}$
Основной уровень	255	20	0,5	82,1
Интервал варьирования	15	10	0,1	
Верхний уровень	270	30	0,6	
Нижний уровень	240	10	0,4	
Нормированные значения факторов	x_1	x_2	x_3	
Опыты: 1	—	—	—	53,0
2	+	—	—	65,3
3	—	+	+	76,1
4	+	+	—	77,0
5	—	—	+	72,7
6	+	—	+	56,1
7	—	+	+	81,0
8	+	+	+	74,1
$\hat{a}_i$ ($\hat{a}_0 = 69{,}41$)	—1,26	7,62	1,56	
$a_i \times$Интервал варьирования	—	76,2	0,156	
Величина шага	Фактор фиксируется на основном уровне	3	0,03 [1]	
Опыты в направлении градиента	x_1^*	x_2^*	x_3^*	$\tilde{y}$
I	255	23	0,53	86,7
II	255	26	0,56	84,2
Добавляемые опыты композиционного ротатабельного плана	x_1	x_2	x_3	$\tilde{y}$
9	—1,68	0	0	70
10	+1,68	0	0	72,1
11	0	—1,68	0	49,1
12	0	+1,68	0	74,8
13	0	0	—1,68	79,7
14	0	0	+1,68	83,5
15	0	0	0	82,0
16	0	0	0	82,9
17	0	0	0	83,6
18	0	0	0	82,8
19	0	0	0	83,1

[1] По техническим причинам изменения на 0,006 оказались невозможными, поэтому шаг установлен равным 0,03.

1. Каноническое преобразование поверхности отклика

В общем случае поверхность второго порядка может быть описана выражением

$$y(x) = a_0 + 2a'x + x'Ax, \quad A' = A, \tag{12.19}$$

где $x' = (x_1, \ldots, x_n)$ — вектор-строка переменных, a_0 — некоторая константа, $a' = (a_1, \ldots, a_n)$ — постоянный вектор-строка, а $A = (a_{ij})$ — некоторая симметричная матрица порядка n.

Координаты экстремума (стационарной точки) x^* можно получить из (12.19), приравнивая нулю первую производную y по x:

$$\frac{\partial y}{\partial x} = 2a + 2Ax^* = 0$$

и решая полученную систему линейных уравнений

$$Ax^* = -a, \quad x^* = -A^{-1}a.$$

Перенося начало координат в точку x^* и поворачивая систему координат с целью приведения ее к главным осям, получаем запись уравнения поверхности (12.19) в канонической форме, позволяющей легко интерпретировать вид поверхности. Обозначим новые координаты через ξ. Тогда

$$x = x^* + B\xi, \quad B'B = I,$$

где матрица B определяет угол между новыми и старыми координатными осями. Подставляя выражение для x в новых переменных в (12.19) и принимая во внимание выражение для x^*, получаем

$$y(\xi) - y^* = \xi' L \xi,$$

где $y^* = a_0 + 2a'x'$ и $L = B'AB$. Через y^* здесь обозначено значение целевой величины в экстремальной точке, а матрица L является диагональной, содержащей собственные значения матрицы A [1]. Для анализа вида поверхности собственные векторы

[1] Для того чтобы уравнение поверхности в координатах ξ содержало только квадраты переменных ξ, матрица L должна быть диагональной. Если A — действительная симметричная матрица, то всегда существует действительная матрица B, удовлетворяющая условию $B'B = I$, где I — единичная матрица, такая, что матрица $L = B'AB$ является диагональной, причем диагональными элементами являются собственные числа матрицы A. Подобное преобразование называют ортогональным преобразованием матрицы A (или приведением матрицы к диагональному виду). Подробнее об этом см. Гантмахер Ф. Р., Теория матриц, изд-во „Наука", М., 1967, Форсайт Дж., Молер К., Численное решение систем линейных алгебраических уравнений, изд-во „Мир", М., 1969. — *Прим. ред.*

(матрица B) интереса не представляют, поэтому достаточно определить собственные значения матрицы A, которые в силу симметричности матрицы A являются действительными числами. Собственные числа λ могут быть найдены как корни уравнения

$$|A-\lambda I|=0. \tag{12.20}$$

Таким образом, процедура канонического преобразования состоит из следующих этапов:

1) определение координат стационарной точки x^* путем решения системы линейных уравнений

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \\ \cdot \\ \cdot \\ \cdot \\ x_n^* \end{bmatrix} = - \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix}; \tag{12.21}$$

2) вычисление значения целевой величины в стационарной точке

$$y^*=a_0+2a'x^*+x^{*\prime}Ax^*; \tag{12.22}$$

3) определение n собственных значений $\lambda_1, \lambda_2, \dots, \lambda_n$ матрицы A путем отыскания корней полинома (12.20)

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix} = (-\lambda)^n+P_1\lambda^{n-1}+\dots+P_n=0; \tag{12.23}$$

4) запись уравнения поверхности в виде

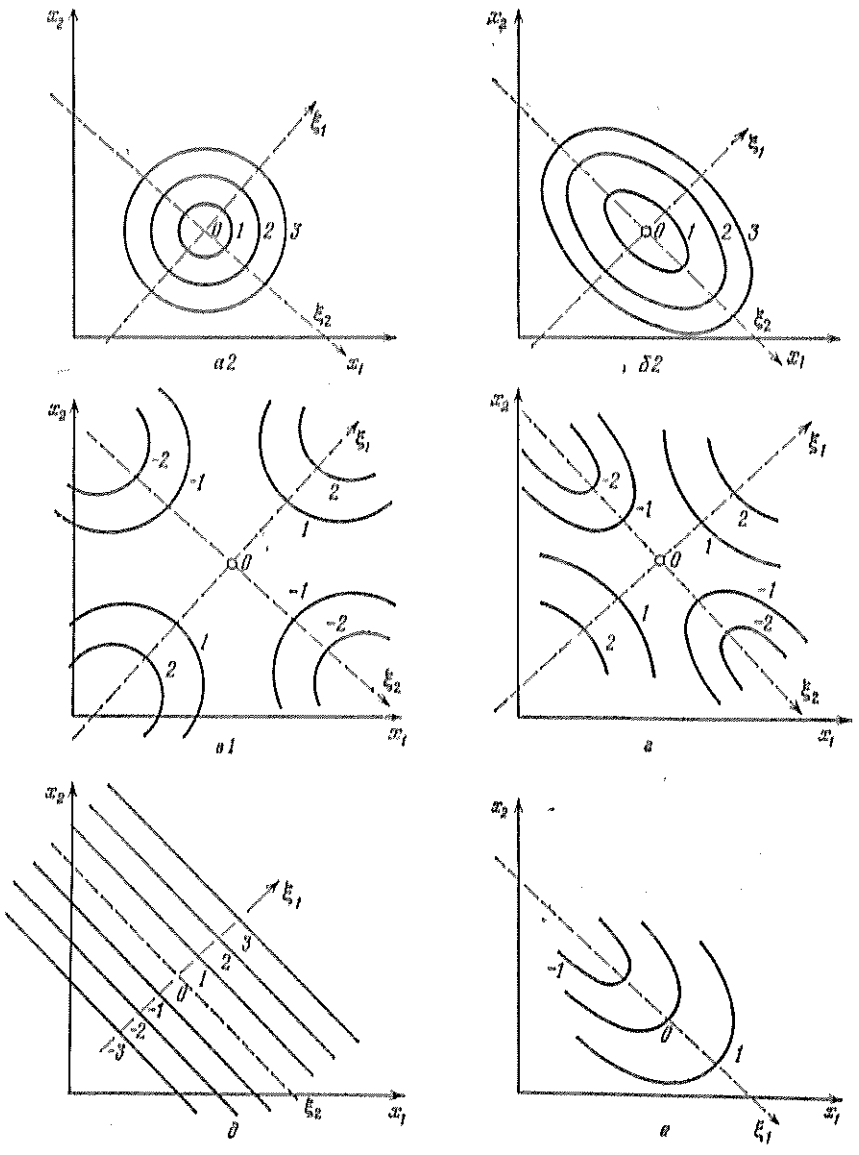
$$y-y^*=\lambda_1\xi_{b1}^2+\lambda_2\xi_{b2}^2+\dots+\lambda_n\xi_{bn}^2. \tag{12.24}$$

Из выражения (12.24) легко определить вид поверхности.

Различные сочетания знаков собственных чисел и соотношения между ними соответствуют поверхностям различного типа. В особом случае, когда ранг матрицы A меньше n, стационарной точки x^* не существует. В лучшем случае при этом может существовать стационарная гиперплоскость. Более точно, стационарной точки не существует, если ранг матрицы A $\mathrm{Rg}(A)\neq \mathrm{Rg}(A, a)$, где $\mathrm{Rg}(A, a)$—ранг расширенной матрицы. Все точки некоторой гиперплоскости размерности $(n-r)$ являются стационарными, если $\mathrm{Rg}(A)=\mathrm{Rg}(A, a)=r<n$.

Для случая двух факторов уравнение в канонической форме имеет вид

$$y-y^*=\lambda_1\xi_{b1}^2+\lambda_2\xi_{b2}^2. \tag{12.25}$$



Фиг. 12.6. Двумерные сечения поверхностей второго порядка (пояснение к табл. 12.5).

Различные виды поверхностей для двухфакторных моделей представлены в табл. 12.5, а двумерные сечения этих поверхностей показаны на фиг. 12.6.

Анализ уравнения, полученного в примере 12.4, позволяет найти следующие координаты оптимума:

$$x_1^* = 0{,}41,$$
$$x_2^* = 0{,}54,$$
$$x_3^* = -0{,}96.$$

Значение целевой величины в этой точке

$$y^* = 85{,}15\%.$$

Таблица 12.5

Виды поверхностей

Номер	Соотношение собственных чисел		Форма кривых равного уровня	Стационарная точка	Номер рисунка
	по величине	по знаку			
1	$\lvert\lambda_1\rvert = \lvert\lambda_2\rvert$	$- -$	Окружности	Максимум	*а1*
2	$\lvert\lambda_1\rvert = \lvert\lambda_2\rvert$	$+ +$	»	Минимум	*а2*
3	$\lvert\lambda_1\rvert \approx \lvert\lambda_2\rvert$	$- -$	Эллипсы	Максимум	*б1*
4	$\lvert\lambda_1\rvert \approx \lvert\lambda_2\rvert$	$+ +$	»	Минимум	*б2*
5	$\lvert\lambda_1\rvert = \lvert\lambda_2\rvert$	$+ -$	Гиперболы	Седловая точка	*в1*
6	$\lvert\lambda_1\rvert = \lvert\lambda_2\rvert$	$- +$	»	» »	*в2*
7	$\lvert\lambda_1\rvert \approx \lvert\lambda_2\rvert$	$+ -$	»	» »	*в*
8	$\lambda_2 = 0$	$-$	Прямые	Нет	*д*
9	$\lambda_2 = 0$	$-$	Параболы	На бесконечности	*е*

Исследование поверхности с помощью канонического преобразования приводит к уравнению

$$y - 85{,}15 = -3{,}94\xi_{б1}^2 - 9{,}28\xi_{б2}^2 + 0{,}71\xi_{б3}^2,$$

где ξ_i, $i = 1, 2, 3$, — новые координатные оси. Детальное исследование показывает, что мы имеем дело с двухполостным гиперболоидом, в стационарной точке которого достигается максимум целевой величины.

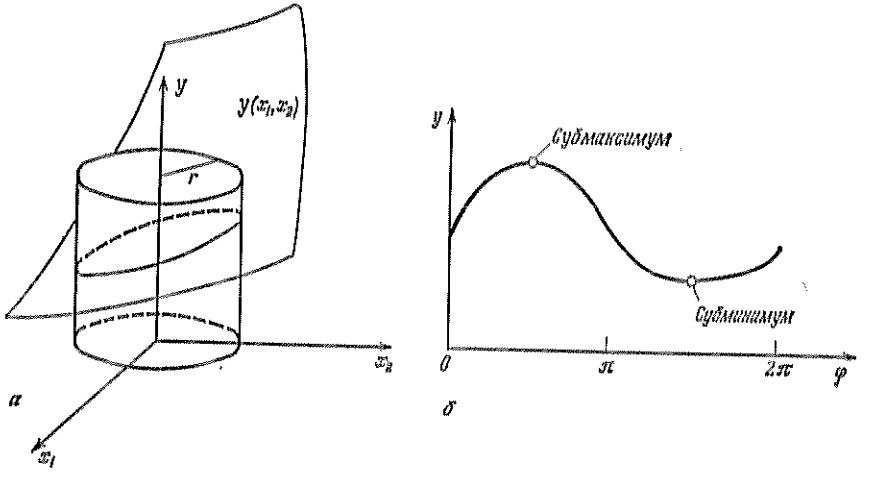
2. Гребневый анализ по Хёрлю [1]

Второй метод исследования поверхности отклика не требует представления этой поверхности в виде полинома второй степени, хотя и был разработан Хёрлем специально для этого случая. Целевая функция представляется заданной дифференцируемой функцией n переменных $x_1, \ldots, x_n$. Основная идея метода заключается в последовательном нахождении точек экстремума целевой функции на поверхностях шара с центром в точке $x_i = 0$, $i = 1\,(1)\,n$, при различных значениях радиуса шара. При

[1] Используется и другое название этого метода: «ридж-анализ». — *Прим. ред.*

этом в $(n+1)$-мерном пространстве (x, y) выделяется некоторый цилиндр, на поверхности которого находятся субоптимальные (suboptima) значения целевой функции [1].

На фиг. 12.7 для $n=2$ показан характер изменения величины целевой функции на пересечении поверхности цилиндра с функцией отклика. В общем случае существует множество субопти-



Фиг. 12.7. Целевая функция.

а — целевая функция $y(x_1, x_2)$ и цилиндр; *б* — значения целевой функции в точках пересечения с поверхностью цилиндра радиуса r.

мальных значений. Изменяя радиус r, можно найти зависимость значений функции в экстремальной точке от величины радиуса r. Так как при всяком r возможно наличие нескольких субоптимальных точек, то для различных r могут существовать кривые, связанные точками слияния и разъединения. Полученные таким образом кривые образуют множество линий гребней. Поэтому и сам метод получил название метода гребневого анализа.

Очевидно, что искомое экстремальное значение функции $y(x_1, \ldots, x_n)$ может быть найдено как экстремальное значение на линиях гребней. Преимущество метода гребневого анализа как раз в том и состоит, что определение экстремального значения может осуществляться через двумерное представление функции $y(r)$.

Рассмотрим теперь применение этого метода к квадратичному полиному (12.19).

[1] Под субоптимальным (субмаксимальным, субминимальным) значением функции здесь понимаются экстремальные значения функции на сфере заданного радиуса. Иными словами, речь идет об условно оптимальных значениях функции. — *Прим. ред.*

Пусть необходимо отыскать оптимум функции $y(x)$, представляющей собой адекватную модель некоторого процесса, полученную с помощью метода множественной регрессии. Найдем экстремум функции $y(x)$ при дополнительном условии

$$x'x = r^2 \tag{12.26}$$

и при фиксированном радиусе r. Вводя множитель Лагранжа λ, определяем необходимое условие стационарной точки x^*, приравнивая нулю производную функции $y(x, \lambda)$ по x. Функция $y(x, \lambda)$ имеет вид

$$y(x, \lambda) = a_0 + 2a'x + x'Ax + \lambda(r^2 - x'x).$$

Необходимым условием стационарной точки при выполнении (12.26) является выражение

$$\frac{1}{2}\frac{\partial y}{\partial x'} = a + Ax^* - \lambda x^* = 0$$

или

$$(A - \lambda I)x^* = -a.$$

Единственное решение x^* системы этих линейных уравнений может быть найдено, если матрица $(A - \lambda I)$ имеет ранг n. Тогда, очевидно, что решение может быть представлено в следующей форме:

$$\left.\begin{aligned} x^*(\lambda) &= (\lambda I - A)^{-1}a, \\ r(\lambda) &= \sqrt{x^{*\prime}x^*}, \\ y^*(\lambda) &= a_0 + 2a'x^* + x^{*\prime}Ax^* \end{aligned}\right\} -\infty < \lambda < \infty. \tag{12.27}$$

Решения x^*, r и y^* системы уравнений (12.27) являются однозначными функциями λ. Два последних уравнения представляют собой параметрическую форму записи уравнения гребневых линий.

Рассмотрим детально решение $x^*(\lambda)$. Матрица A имеет не более n действительных собственных чисел λ_i, которые могут быть найдены из условия

$$|A - \lambda I| = 0. \tag{12.28}$$

При $\lambda \to \lambda_i$ одна из компонент вектора x стремится к нулю, а величина r — к бесконечности. Каждому монотонному отрезку функции $r(\lambda)$ соответствует линия гребня.

Рассмотрим числовой пример. Пусть требуется найти экстремум функции двух переменных

$$y = x_1^2 + 2x_2^2 + 2x_1 - 2x_2 + 2.$$

Для векторной записи функции y имеем

$$a_0 = 2, \quad a = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Из уравнения

$$0 = \begin{vmatrix} 1-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} = (\lambda-1)(\lambda-2)$$

находим собственные числа λ:

$$\lambda_1 = 1 \quad \text{и} \quad \lambda_2 = 2.$$

Решение $x^*(\lambda)$ определяется уравнением (12.27)

$$x^* = \begin{pmatrix} \lambda-1 & 0 \\ 0 & \lambda-2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\lambda-1} & 0 \\ 0 & \frac{1}{\lambda-2} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\lambda-1} \\ -\frac{1}{\lambda-2} \end{pmatrix},$$

$$x_1^* = \frac{1}{\lambda-1}, \quad x_2^* = -\frac{1}{\lambda-2}.$$

Для r получаем

$$r^2 = x^{*\prime} x^* = \frac{1}{(\lambda-1)^2} + \frac{1}{(\lambda-2)^2}.$$

Вид функции $r(\lambda)$ показан на фиг. 12.8. Функция $r(\lambda)$ имеет полюсы при $\lambda_1 = 1$ и $\lambda_2 = 2$; она монотонно убывает при $\lambda \to \infty$; $\lambda > \lambda_2$ и при $\lambda \to -\infty$; $\lambda < \lambda_1$. При изменении λ в интервале $1 < \lambda < 2$ функция $r(\lambda)$ сначала убывает до минимального значения $r_0 = 2\sqrt{2}$ в точке $\lambda_0 = 3/2$ и затем монотонно возрастает. Виду функции $r(\lambda)$ соответствуют четыре гребневые линии, которые представлены на фиг. 12.9 в виде зависимости $y^*(\lambda)$ от $r(\lambda)$:

$$y^*(\lambda) = 2 + 2(1, -1) x^* + x^{*\prime} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} x^*,$$

$$y^*(\lambda) = 2 + 2(1, -1) \begin{pmatrix} \frac{1}{\lambda-1} \\ -\frac{1}{\lambda-2} \end{pmatrix} + \left(\frac{1}{\lambda-1}, -\frac{1}{\lambda-2}\right) \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda-1} \\ -\frac{1}{\lambda-2} \end{pmatrix},$$

$$\left.\begin{aligned} y^*(\lambda) &= 2 + \frac{2}{\lambda-1} + \frac{2}{\lambda-2} + \frac{1}{(\lambda-1)^2} + \frac{2}{(\lambda-2)^2}, \\ r(\lambda) &= \sqrt{\frac{1}{(\lambda-1)^2} + \frac{1}{(\lambda-2)^2}}. \end{aligned}\right\}$$

Последние выражения дают параметрическую запись уравнений гребневых линий. Различные гребневые линии *I*—*IV* соответствуют следующим областям изменения λ:

Гребневая линия *I* для $2 < \lambda < \infty$,

» » *II* » $1{,}5 < \lambda < 2$,

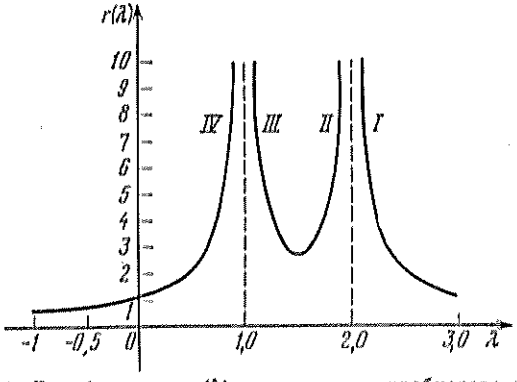
» » *III* » $1 < \lambda < 1{,}5$,

» » *IV* » $-\infty < \lambda < 1$.

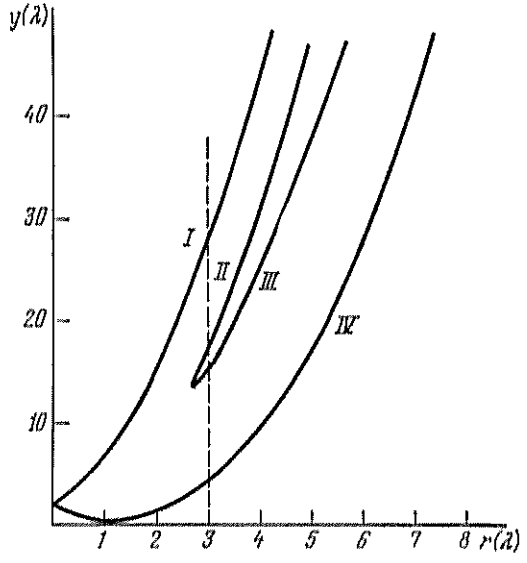
Из фиг. 12.9 видно, что функция $y(x_1, x_2)$ имеет минимум на линии *IV* при $r = 1$, чему соответствует значение $\lambda \approx 0$. При этом координаты минимума таковы:

$$x_1^* = -1, \quad x_2^* = 0{,}5, \quad y^* = 0{,}5.$$

При наличии технологических ограничений на целевую функцию положение максимума последней легко определяется из



Фиг. 12.8. Вид функции $r(\lambda)$ для примера гребневого анализа.



Фиг. 12.9. Линии гребней для примера гребневого анализа.

фиг. 12.9. Пусть, например, $r \leqslant 3$. Тогда максимальное значение достигается на линии *I* при $\lambda = 2{,}345$ и следующих значениях

координат:

$$x_1^{\text{опт}} = 0{,}74, \quad x_2^{\text{опт}} = -2{,}86, \quad y^{\text{опт}} = 26{,}2.$$

12.4. Планирование эксперимента в производственных условиях

В случае отсутствия математической модели управление производственными процессами требует периодического проведения активного эксперимента с целью поиска оптимальных значений целевой величины процесса (выход продукции, доход, расход) при постоянно изменяющихся условиях производства. Возможности экспериментирования на действующих установках, как правило, ограничены. Поэтому для промышленных условий разработаны специальные приемы исследования, требующие небольшого числа опытов и связанные с незначительными отклонениями от номинальных режимов ведения процессов. К таким методам относятся последовательный симплексный метод и метод эволюционного планирования (ЭВОП).

12.4.1. Последовательный симплексный метод планирования эксперимента

Последовательный симплексный метод относится к методам поиска экстремума целевой функции, применение которого требует проведения минимально возможного числа опытов при определении направления движения и связано с весьма незначительными по объему вычислениями. Метод основан на активном эксперименте и может использоваться как в практике научных исследований, так и в управлении технологическими процессами.

Основные понятия последовательного симплексного метода

Симплексом называется простейшая выпуклая геометрическая фигура. Под n-мерным симплексом в n-мерном пространстве E_n понимают фигуру, образованную множеством $(n+1)$ точек, не принадлежащих одновременно ни одному $(n-1)$-мерному подпространству данного пространства E_n. Эти $(n+1)$ точки называются вершинами симплекса. Число вершин симплекса всегда на единицу больше размерности n пространства E_n. В двумерном пространстве (т. е. на плоскости) симплексом является любой треугольник, в трехмерном пространстве — любая треугольная пирамида (тетраэдр).

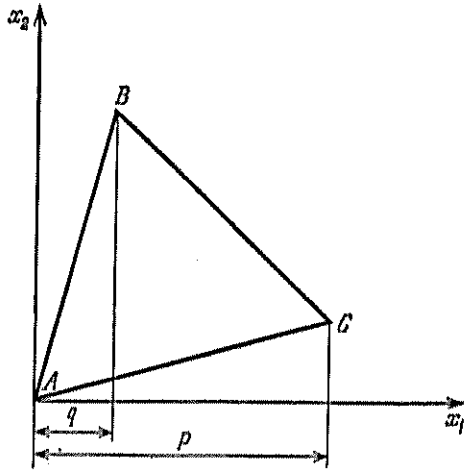
Симплекс называют регулярным, если расстояния между всеми точками, образующими симплекс (т. е. между вершинами симплекса), одинаковы. Из произвольного симплекса всегда можно получить регулярный путем преобразования системы координат.

В дальнейшем мы будем рассматривать только регулярные симплексы.

Построение начального симплекса

Опишем два способа задания координат вершин начального симплекса.

1. Поместим одну из вершин симплекса в начало координат, а остальные вершины расположим так, чтобы ребра, исходящие



Фиг. 12.10. Начальный симплекс с вершиной в начале координат.

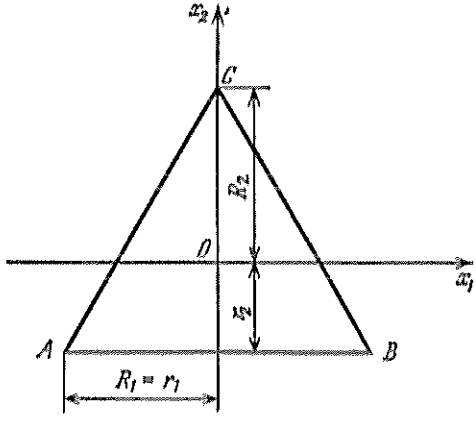
из первой вершины, образовывали одинаковые углы с соответствующими координатными осями (фиг. 12.10). Координаты вершин симплекса в этом случае могут быть представлены следующей матрицей:

Номера вершин	Координаты вершин				
	x_1	x_2	x_3	...	x_n
1	0	0	0	...	0
2	p	q	q	...	q
3	q	p	q	...	q
4	q	q	p	...	q
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
$n+1$	q	q	q	...	p

Здесь

$$p = \frac{1}{n\sqrt{2}}(n - 1 + \sqrt{n+1}),$$
$$q = \frac{1}{n\sqrt{2}}(\sqrt{n+1} - 1). \qquad (12.29)$$

Длина l ребра симплекса (т. е. расстояние между вершинами симплекса) здесь принята равной единице ($l = 1$).



Фиг. 12.11. Начальный симплекс с центром в начале координат.

2. Центр симплекса (центр плана эксперимента) поместим в начало координат, а $(n+1)$-ю вершину — на ось x_n (фиг. 12.11). Остальные вершины располагаются симметрично относительно координатных осей. Координаты вершин в этом случае определяются следующей матрицей:

Номер вершины	Координаты вершин				
	x_1	x_2	$x_3 \ldots x_{n-1}$		x_n
1	$-r_1$	$-r_2$	$-r_3 \ldots -r_{n-1}$		$-r_n$
2	R_1	$-r_2$	$-r_3 \ldots -r_{n-1}$		$-r_n$
3	0	R_2	$-r_3 \ldots -r_{n-1}$		$-r_n$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n	0	0	0	$-R_{n-1}$	$-r_n$
$n+1$	0	0	0	0	R_n

При единичной длине ребра симплекса

$$\left.\begin{aligned} r_l &= \frac{1}{\sqrt{2l\,(l+1)}}, \\ R_l &= \frac{1}{\sqrt{2\,(l+1)}} \end{aligned}\right\} \quad l = 1,2, \ldots, n. \qquad (12.30)$$

Пример 12.5

Матрица координат вершин начального симплекса, построенного по первому способу при $n=3$ ($l=1$), имеет следующие элементы:

Номер эксперимента (вершины)	Координаты точек плана		
	x_1	x_2	x_3
1	0	0	0
2	0,944	0,236	0,236
3	0,236	0,944	0,236
4	0,236	0,236	0,944

Пример 12.6

Матрица координат вершины (матрица планирования эксперимента) начального симплекса, построенного по второму способу, при $n=5$ имеет следующие элементы:

Номер эксперимента (вершины)	Координаты точек плана				
	x_1	x_2	x_3	x_4	x_5
1	−0,5	−0,289	−0,204	−0,158	−0,129
2	0,5	−0,289	−0,204	−0,158	−0,129
3	0	0,578	−0,204	−0,158	−0,129
4	0	0	0,612	−0,158	−0,129
5	0	0	0	0,632	−0,129
6	0	0	0	0	0,645

Применение последовательного симплексного планирования при поиске оптимума целевой функции

Экспериментальное определение оптимума целевой функции осуществляется с помощью следующей процедуры:

1. Преобразование исходных факторов с таким расчетом, чтобы изменение каждого фактора на единицу приводило приблизительно к одинаковому изменению целевой величины.

2. Расчет координат начального симплекса (по способу 1 или 2). Реализация соответствующих опытов.

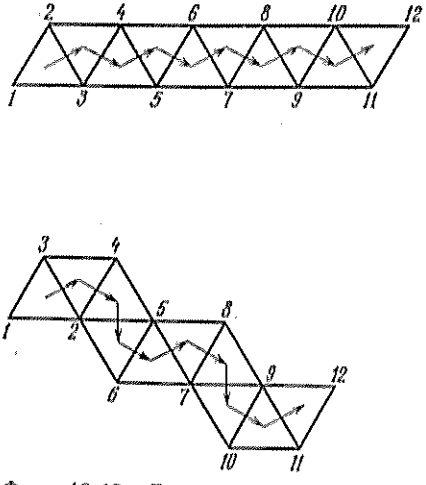
3. Отбрасывание точки плана с наименьшим значением целевой величины и построение нового симплекса. Новый симплекс образуется оставшимися вершинами исходного симплекса и новой вершиной, получаемой путем зеркального отображения отброшенной вершины относительно противоположной ей $(n-1)$-мерной грани исходного симплекса. Координаты новой точки x^* рассчитываются следующим образом (в векторной записи):

$$x^* = \frac{2}{n}(x^1 + x^2 + \ldots + x^{j-1} + x^{j+1} + \ldots + x^{n+1}) - x^j = \\ = \frac{2}{n}\sum_{l=1}^{n+1} x^l - \left[\frac{2}{n} + 1\right] x^j, \qquad (12.31)$$

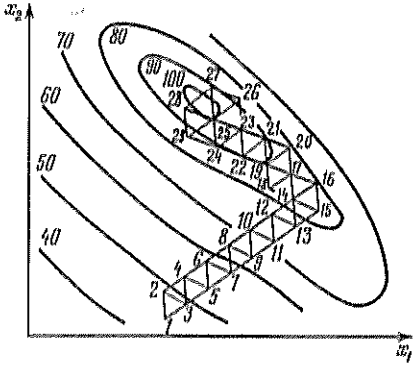
где j — номер вершины исходного симплекса с наименьшим значением целевой функции. Для $\hat{y}^*$ (прогнозируемое значение целевой функции в новой точке) имеем

$$\hat{y}^* = \frac{2}{n}\sum_{l=1}^{n+1} \hat{y}^l - \left[\frac{2}{n} + 1\right] \hat{y}^j. \qquad (12.32)$$

4. Проведение эксперимента в новой точке x^* и получение соответствующего значения $\tilde{y}^*$ целевой величины.



Фиг. 12.12. Виды движения центра симплекса в направлении градиента.



Фиг. 12.13. Поведение симплекса вблизи максимума.

5. Последовательное перемещение симплекса, в процессе которого на каждом шаге происходит отбрасывание вершины симплекса с наихудшим значением целевой величины и реализация опыта в новой вершине. При этом направление перемещения центра симплекса колеблется около направления градиента. Два возможных вида перемещения центра симплекса представлены на фиг. 12.12.

6. Если при перемещении симплекса на протяжении $n+1$ шагов та или иная вершина сохраняет свое положение, то симплекс совершает оборот вокруг этой вершины (фиг. 12.13). Это означает, что либо в данной точке находится оптимум целевой функции, либо значение целевой функции в этой вершине определено неверно. Чтобы уточнить, какая ситуация имеет место, в этой точке вновь проводится эксперимент, и в дальнейшем работают с новым значением целевой величины.

7. Если оказывается, что целевая величина в новой вершине симплекса меньше, чем в остальных вершинах, в соответствии с логикой движения следует возвратиться к предыдущему симплексу. Чтобы предотвратить „зацикливание", в качестве отбрасываемой выбирают вершину, в которой целевая функция имеет величину, следующую по порядку за наихудшей вершиной симплекса. Дальнейшее движение продолжают, реализуя эксперимент в точке, расположенной „зеркально" к отброшенной вершине.

8. Если новая вершина выходит за пределы допустимой области планирования, следует поступать так же, как в п. 7.

9. При достижении области оптимума размер симплекса уменьшают (как правило, на $^1/_4$ начальной величины).

10. Оптимум считается достигнутым, если одна и та же точка входит в последовательные симплексы N раз, где

$$N = 1{,}65n + 0{,}05n^2, \quad 2 \leqslant n \leqslant 30. \tag{12.33}$$

Выражение (12.33) получено эмпирически. Другое условие достижения оптимума

$$\sqrt{\sum_{l=1}^{n+1} \frac{(\tilde{y}^l - \tilde{y}_m)^2}{n}} \leqslant \varepsilon, \tag{12.34}$$

где ε—малая величина, $\tilde{y}_m$—среднее значение целевых величин в вершинах симплекса.

11. Если ошибка эксперимента относительно велика, целесообразно в каждой вершине симплекса ставить несколько опытов и использовать усредненные значения наблюдений целевой функции.

Преимущества и недостатки последовательного симплексного метода

Преимущества:

— Число необходимых опытов при определении направления движения мало по сравнению с другими планами. Каждый дополнительный фактор требует постановки только одного дополнительного опыта.

— Метод чрезвычайно прост с точки зрения объема вычислений.

— Ограничения на область изменения факторов легко учитываются при движении симплекса.

— Чем больше число факторов, тем выше эффективность метода.

— Легко включать или исключать из рассмотрения те или иные переменные в ходе применения метода.

— Направление движения определяется только соотношением величин целевой функции в вершинах симплекса, а не их абсолютными значениями.

— Направление движения может лишь в незначительной степени отличаться от направления крутого восхождения.

— Метод может быть использован в случае дрейфа характеристик объекта.

Недостатки:

— Реализация метода не дает информации о влиянии каждого фактора на целевую функцию.

— Движение по правилам симплексного метода дает ограниченное представление о характере поверхности отклика [1].

Пример 12.7 [10]

Требуется экспериментальным путем, используя последовательный симплексный метод, найти отношение растворителя к основному веществу (x_1), температуру (x_2) и время реакции (x_3), при которых имеет место максимальный выход целевого продукта. В качестве начального выбран симплекс с вершинами в точках $ABCD$ (фиг. 12.14). Координаты этих вершин (в натуральных единицах и в кодовых обозначениях) приведены в табл. 12.6.

---

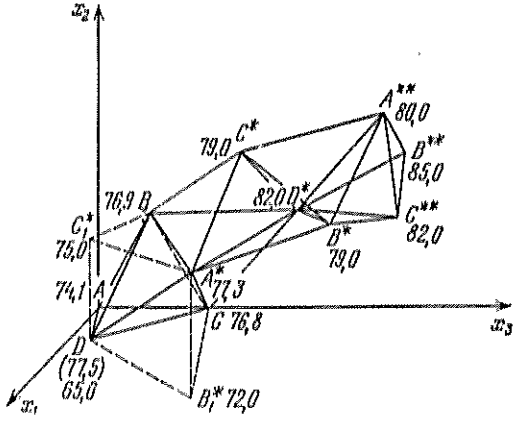
[1] По результатам наблюдений в точках симплексного плана могут быть рассчитаны компоненты градиента, а по наблюдениям в вершинах системы симплексов можно найти описание стационарной области. Подробнее об этом см. статью В. Г. Горского и В. В. Бродского в кн. „Новые идеи в планировании эксперимента", изд-во „Наука", М., 1969, а также Горский В. Г., Адлер Ю. П., Планирование промышленных экспериментов, изд-во „Металлургия", М., 1974. — *Прим. ред.*

Таблица 12.6

План эксперимента и результаты (пример 12.7)

Номер опыта	Исходный симплекс	Вершина симплекса, в которой ставится опыт	Отношение растворителя к основному веществу x_1^*		Температура x_2^*		Время реакции x_3^*		Выход $\tilde{y}$
			г/моль	код	°С	код	мин	код	%
1	*ABCD*	*A*	0,50	0	155	0	40	0	74,1
2	*ABCD*	*B*	0,50	0	164	0,86	50	0,5	76,9
3	*ABCD*	*C*	0,50	0	155	0	60	1,0	76,8
4	*ABCD*	*D*	0,66	0,82	158	0,29	50	0,5	77,5
5	*ABCD*	*A**	0,61	0,55	163	0,77	67	1,33	77,3
6	*A*BCD*	C_1^*	0,68	0,90	168	1,29	51	0,56	75,0
7	*A*BCD*	B_1^*	0,68	0,92	153	—0,16	70	1,49	72,0
8	*A*BCD*	*D*	0,66	0,82	158	0,29	50	0,5	65,0
9	*A*BCD*	*D**	0,41	—0,45	163	0,80	70	1,49	82,0
10	*A*BCD**	*C**	0,51	0,07	171	1,62	64	1,22	79,0
11	*A*BC*D**	*B**	0,52	0,11	168	1,27	84	2,20	79,0
12	*A*B*C*D**	*A***	0,35	—0,73	172	1,69	79	1,94	80,0
13	*A**B*C*D**	*C***	0,34	—0,78	164	0,88	89	2,48	82,0
14	*A**B*C**D**	*B***	0,21	—1,43	165	0,98	75	1,75	85,0

Результаты экспериментов и вытекающие из них решения видны из табл. 12.6. Так как в точке D из-за ошибки эксперимента при первом измерении было получено слишком большое значение выхода, то симплекс начинает поворачиваться вокруг



Фиг. 12.14. Иллюстрация к примеру 12.7.

точки D (симплекс с вершинами A^*BCD). Опыты в точках C_1^* и B_1^* (опыты 6 и 7 в табл. 12.7) не привели к увеличению выхода, поэтому в точке D был вторично проведен эксперимент (опыт 8), который позволил скорректировать значение выхода в этой точке. Результат, полученный в точке B^{**}, удовлетворил экспериментаторов, поэтому после 14-го опыта эксперимент был прекращен и в качестве оптимальных значений факторов были приняты следующие: отношение растворителя к основному веществу 0,21 г/моль, температура 165°С, продолжительность реакции 75 мин.

12.4.2. Метод эволюционного планирования (ЭВОП)

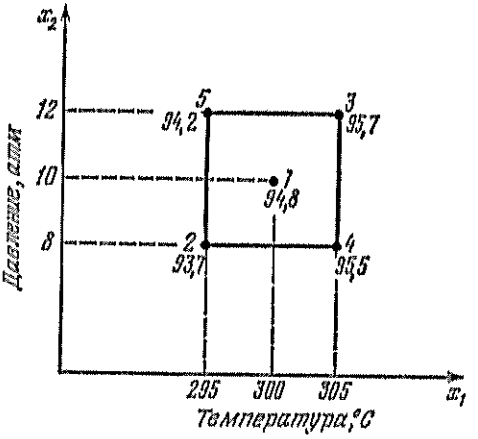
С помощью разработанного Боксом [11] метода эволюционного планирования (ЭВОП) путем незначительных возмущений, постоянно и целенаправленно вносимых в производственный процесс, извлекается информация, используемая для поиска оптимальных режимов процесса.

В промышленных условиях допускаются лишь малые изменения режимов, в то время как ошибки измерения целевых величин могут быть достаточно велики. Поэтому существенной особенностью ЭВОП-метода является возможность надежного выявления малых изменений полезного сигнала на фоне шума. В методе ЭВОП это достигается повторением опытов плана эксперимента p раз. При этом среднеквадратичная ошибка среднего из p независимых измерений оказывается в $\sqrt{p}$ раз меньше среднеквадратичной ошибки единичного измерения. Однократная реализация плана эксперимента в методе ЭВОП называется циклом, а совокупность p реализаций одного и того же плана называют фазой исследования. В одной фазе делается столько циклов, сколько необходимо для того, чтобы на фоне помех обнаружить значимый эффект от воздействия одной или нескольких переменных на целевую функцию и обеспечить возможность значимого улучшения целевой величины.

По окончании фазы выбираются новые базовые значения для варьируемых переменных, составляется и p раз реализуется новый план эксперимента.

В качестве планов эксперимента в методе ЭВОП применяются планы полного и дробного факторного эксперимента, которые дополняются опытом в центре плана. Общее число опытов при использовании полного факторного эксперимента равно $N = 2^n + 1$. Для двух независимых переменных такой план состоит из пяти (фиг. 12.15), а для трех — из девяти опытов. Наиболее удобно пользоваться этим методом, когда одновременно варьируют две или три переменные.

Регистрация и обработка результатов эксперимента проводится с помощью специальных рабочих листков. В заголовке рабочего листка указываются номера точек постановки экспериментов (см. табл. 12.7—12.9). В первую и вторую строки рабочего листка каждой фазы заносятся значения сумм и средних значений целевых величин, полученных для каждой точки эксперимента в ходе реализации предыдущих циклов. В третьей строке приводятся новые значения целевой величины, а в четвертой — разность между средним по предшествующим циклам

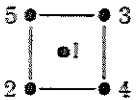


Фиг. 12.15. План эксперимента для метода ЭВОП ($n=2$).

(строка 2) и новым значениями целевой величины. Эффект некоторого фактора равен приращению значения целевой величины в результате перехода фактора с нижнего уровня на верхний. В случае двух факторов (фиг. 12.15) эффект переменной x_1 равен полуразности сумм целевых величин в точках 3, 4 и 2, 5, а эффект переменной x_2 определяется полуразностью сумм целевых величин в точках 3, 5 и 2, 4. Эффект взаимодействия равен полуразности сумм целевых величин точек 2 и 3 и точек 4 и 5. Для определения эффекта изменения среднего значения из суммы целевых величин в точках 2, 3, 4 и 5 вычитают учетверенное значение целевой величины в точке 1 и делят полученный результат на 5. Разность Δm между средним значением целевой величины по наблюдениям во всех точках плана и по всем циклам и средним значениям целевой величины в центре плана ($\tilde{y}_m^1$)

$$\Delta m = \frac{1}{N}\sum_{l=1}^{N}\tilde{y}_m^l - \tilde{y}_m^1 \qquad (12.35)$$

Таблица 12.7



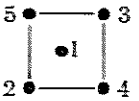
ЭВОП, 1-й цикл ($p=1$)
Отклик: выход продукта

1-я фаза
Дата

Расчет средних значений						Расчет среднеквадратичных отклонений
	Номера точек эксперимента					
	1	2	3	4	5	
1. Сумма предыдущих циклов						Сумма предыдущих циклов $S=$
2. Среднее предыдущих циклов						Среднее предыдущих циклов $s=$
3. Новые наблюдения	94,0	94,5	96,5	94,5	94,5	Размах=
4. Разности (2)—(3)						Новое значение $s=$ Размах$\times f_{N,p}=$
5. Новые суммы (1)+(3)	94,0	94,5	96,5	94,5	94,5	Новая сумма $S=$
6. Новые средние $\tilde{y}_m^i$	94,0	94,5	96,5	94,5	94,5	Новое среднее значение $s=\dfrac{\text{Новая сумма } S}{p-1}$

Вычисление эффектов	Вычисление доверительных интервалов (ДИ)
Эффект температуры $=\frac{1}{2}(\tilde{y}_m^3+\tilde{y}_m^4-\tilde{y}_m^2-\tilde{y}_m^5)=1{,}00$	ДИ для новых средних $\frac{2}{\sqrt{p}}s=$
Эффект давления $=\frac{1}{2}(\tilde{y}_m^2+\tilde{y}_m^5-\tilde{y}_m^3-\tilde{y}_m^4)=1{,}00$	ДИ для эффектов $\frac{2}{\sqrt{p}}s=$
Эффект взаимодействия $=\frac{1}{2}(\tilde{y}_m^2+\tilde{y}_m^3-\tilde{y}_m^4-\tilde{y}_m^5)=1{,}00$	ДИ для изменения среднего $\frac{1{,}78}{\sqrt{p}}s=$
Эффект изменения среднего $=\frac{1}{5}(\tilde{y}_m^2+\tilde{y}_m^3+\tilde{y}_m^4+\tilde{y}_m^5-\tilde{y}_m^1)=0{,}80$	

Таблица 12.8



ЭВОП, 2-й цикл ($p=2$)

1-я фаза

Дата

Расчет средних значений						Расчет среднеквадратичных отклонений
	Номера точек эксперимента					
	1	2	3	4	5	
1. Сумма предыдущих циклов	94,0	94,5	96,5	94,5	94,5	Сумма предыдущих циклов $S=$
2. Среднее предыдущих циклов	94,0	94,5	96,5	94,5	94,5	Среднее предыдущих циклов $s=$
3. Новые наблюдения	96,0	95,0	95,0	95,5	94,0	Размах $=\lvert 1{,}5\rvert+\lvert -2{,}0\rvert=3{,}5$
4. Разности [(2)—(3)]	−2,0	−0,5	1,5	−0,2	0,5	Новое значение $s=$ Размах $\times f_{N,\,p}=3{,}5\times 0{,}3=1{,}05$
5. Новые суммы (1)+(3)	190,0	189,5	191,5	191,0	188,5	Новая сумма $S=1{,}05$
6. Новые средние $\tilde{y}_m^i$	95,0	94,7	95,7	95,5	94,2	Новое среднее значение $s=\dfrac{\text{Новая сумма } S}{p-1}=\dfrac{1{,}05}{1}=1{,}05$

Вычисление эффектов		Вычисление доверительных интервалов (ДИ)
Эффект температуры	$=\frac{1}{2}(\tilde{y}_m^2+\tilde{y}_m^4-\tilde{y}_m^2-\tilde{y}_m^5)=1{,}15$	ДИ для новых средних $\dfrac{2s}{\sqrt{p}}=\dfrac{2\times 1{,}05}{\sqrt{2}}=1{,}48$
Эффект давления	$=\frac{1}{2}(\tilde{y}_m^3+\tilde{y}_m^5-\tilde{y}_m^2-\tilde{y}_m^4)=-0{,}15$	ДИ для эффектов $\dfrac{2s}{\sqrt{p}}=\dfrac{2\times 1{,}05}{\sqrt{2}}=1{,}48$
Эффект взаимодействия	$=\frac{1}{2}(\tilde{y}_m^2+\tilde{y}_m^3-\tilde{y}_m^4-\tilde{y}_m^5)=0{,}35$	ДИ для изменения среднего $\dfrac{1{,}78}{\sqrt{p}}s=\dfrac{1{,}78\times 1{,}05}{\sqrt{2}}=1{,}32$
Эффект изменения среднего	$=\frac{1}{5}(\tilde{y}_m^2+\tilde{y}_m^3+\tilde{y}_m^4+\tilde{y}_m^5-4\tilde{y}_m^1)=0{,}02$	

Таблица 12.9



ЭВОП, 3-й цикл ($p=3$)
Отклик: выход продукта

1-я фаза
Дата

Расчет средних значений						Расчет среднеквадратичных отклонений
	Номера точек эксперимента					
	1	2	3	4	5	
1. Сумма предыдущих циклов	190,0	189,5	191,5	191,0	188,5	Сумма предыдущих циклов $S=1{,}05$
2. Среднее предыдущих циклов	95,0	94,7	95,7	95,5	94,2	Среднее предыдущих циклов $s=1{,}05$
3. Новые наблюдения	94,5	93,5	96,0	97,0	94,0	Размах $=\lvert 1{,}2\rvert+\lvert -1{,}5\rvert=2{,}7$
4. Разности (2)—(3)	0,5	1,2	—0,3	—1,5	0,2	Новое значение $s=$ Размах $\times f_{N,\,p}=2{,}7\times 0{,}35=0{,}95$
5. Новые суммы (1)+(3)	284,5	281,0	287,2	286,5	282,7	Новая сумма $S=1{,}05+0{,}95=2{,}00$
6. Новые средние $\bar{y}_m^i$	94,8	93,7	95,7	95,5	94,2	Новое среднее значение $s=\dfrac{\text{Новая сумма } S}{p-1}=\dfrac{2{,}00}{2}=1{,}00$

Вычисление эффектов	Вычисление доверительных интервалов (ДИ)
Эффект температуры $=\dfrac{1}{2}(\bar{y}_m^3+\bar{y}_m^4-\bar{y}_m^2-\bar{y}_m^5)=1{,}35$	ДИ для новых средних $\dfrac{2s}{\sqrt{p}}=\dfrac{2{,}1}{\sqrt{3}}=1{,}16$
Эффект давления $=\dfrac{1}{2}(\bar{y}_m^2+\bar{y}_m^5-\bar{y}_m^3-\bar{y}_m^4)=0{,}05$	ДИ для эффектов $\dfrac{2s}{\sqrt{p}}=\dfrac{2{,}1}{\sqrt{3}}=1{,}16$
Эффект взаимодействия $=\dfrac{1}{2}(\bar{y}_m^3+\bar{y}_m^5-\bar{y}_m^4-\bar{y}_m^5)=0{,}15$	ДИ для изменения среднего $\dfrac{1{,}78s}{\sqrt{p}}=\dfrac{1{,}78\times 1}{\sqrt{3}}=1{,}02$
Эффект изменения среднего $=\dfrac{1}{5}(\bar{y}_m^2+\bar{y}_m^3+\bar{y}_m^4+\bar{y}_m^5-4\bar{y}_m^1)=0{,}1$	

характеризует вид поверхности в области экспериментирования [1].

Если Δm значимо меньше нуля, мы имеем дело с максимумом; если Δm значимо больше нуля, — с минимумом. Если же Δm незначимо отличается от нуля, то поверхность отклика в данной области либо описывается плоскостью, либо имеет седловую точку.

Для вычисления дисперсий эффектов введем следующие обозначения. После реализации p циклов для N экспериментальных точек получаем следующие результаты измерения целевой величины:

Усредненные значения после $(p-1)$-го цикла $\tilde{y}^1_m \quad \tilde{y}^2_m \ldots \tilde{y}^N_m$, (12.36)

Измеренные значения для p-го цикла $\tilde{y}^{1p} \quad \tilde{y}^{2p} \ldots \tilde{y}^{Np}$, (12.37)

Разность $(\tilde{y}^l_m - \tilde{y}^{lp}) \qquad D^1 \quad D^2 \ldots D^N$, (12.38)

$$D^l = \frac{1}{p-1} \sum_{j=1}^{p-1} \tilde{y}^{lj} - \tilde{y}^{lp}. \tag{12.39}$$

Для дисперсии σ^2_{Dl} получаем

$$\sigma^2_{Dl} = \frac{1}{(p-1)^2} \sum_{j=1}^{p-1} \sigma^2_{\tilde{y}^{lj}} + \sigma^2_{\tilde{y}^{lp}}. \tag{12.40}$$

Так как предполагается, что все измерения целевой величины независимы и имеют одну и ту же дисперсию σ^2, то

$$\sigma^2_D = \frac{1}{(p-1)^2}(p-1)\sigma^2 + \sigma^2 = \frac{p}{p-1}\sigma^2. \tag{12.41}$$

Отсюда следует, что

$$\sigma^2 = \frac{p-1}{p}\sigma^2_D \quad \text{и} \quad \sigma = \sqrt{\frac{p-1}{p}}\,\sigma_D. \tag{12.42}$$

Это означает, что оценка s^2_D дисперсии величин D^l может быть использована для получения оценки s^2 дисперсии ошибки наблюдений. Для упрощения вычислений в ЭВОП-методе оценку s^2_D получают, используя размах величин D:

$$s_D = \frac{\delta D}{d},$$

$$\delta D = \left| \max_l D^l - \min_l D^l \right|. \tag{12.43}$$

Из (12.42) и (12.43) следует, что

$$s = \frac{1}{d}\sqrt{\frac{p-1}{p}}\,\delta D. \tag{12.44}$$

[1] Величина Δm в общем виде оценивает эффект изменения среднего, обсуждавшийся выше для двухфакторного случая. — *Прим. ред.*

Значения

$$\frac{1}{d}\sqrt{\frac{p-1}{p}} = f_{p,\,N}$$

табулированы (см. табл. 12.10).

Таблица 12.10

	Значения для $f_{p,N}$								
Номера циклов p	*Число опытов N*								
	2	3	4	5	6	7	8	9	10
2	0,63	0,42	0,32	0,30	0,28	0,26	0,25	0,24	0,23
3	0,72	0,48	0,40	0,35	0,32	0,30	0,29	0,27	0,26
4	0,77	0,51	0,42	0,37	0,34	0,32	0,30	0,29	0,28
5	0,79	0,53	0,43	0,38	0,35	0,33	0,31	0,30	0,29
6	0,81	0,54	0,44	0,39	0,36	0,34	0,32	0,31	0,30
7	0,82	0,55	0,45	0,40	0,37	0,34	0,33	0,31	0,30
8	0,83	0,55	0,45	0,40	0,37	0,35	0,33	0,31	0,30
9	0,84	0,56	0,46	0,40	0,37	0,35	0,33	0,32	0,31
10	0,84	0,56	0,46	0,41	0,37	0,35	0,33	0,32	0,31
11	0,84	0,56	0,46	0,41	0,38	0,35	0,33	0,32	0,31
12	0,85	0,57	0,47	0,41	0,38	0,35	0,34	0,32	0,31
13	0,85	0,57	0,47	0,41	0,38	0,36	0,34	0,32	0,31
14	0,85	0,57	0,47	0,41	0,38	0,36	0,34	0,32	0,31
15	0,86	0,57	0,47	0,42	0,38	0,36	0,34	0,33	0,31
16	0,86	0,57	0,47	0,42	0,38	0,36	0,34	0,33	0,32
17	0,86	0,57	0,47	0,42	0,38	0,36	0,34	0,33	0,32
18	0,86	0,57	0,47	0,42	0,38	0,36	0,34	0,33	0,32
19	0,86	0,58	0,47	0,42	0,38	0,36	0,34	0,33	0,32
20	0,86	0,58	0,47	0,42	0,38	0,36	0,34	0,33	0,32

В табл. 12.10 в первом столбце указано число циклов p (p изменяется от 2 до 20), а в первой строке — число опытов N в каждом цикле (N изменяется от 2 до 10). Например, для второго цикла и некоторого плана с пятью точками ($2^2 + 1 = 5$) находим $f_{2,\,5} = 0{,}30$. Для оценки s получаем

$$s = \delta D f_{p,\,N}.$$

Для проверки значимости эффектов вычисляем величину доверительного интервала для эффектов факторов и эффектов взаимодействия:

$$\Delta_i = \pm t \frac{2}{\sqrt{N-1}} \frac{s}{\sqrt{p}}, \qquad (12.45)$$

а также для эффекта изменения среднего [1]:

$$\Delta_m = \pm t \sqrt{\frac{N-1}{N}} \frac{s}{\sqrt{p}}. \qquad (12.46)$$

При $N = 2^n + 1 = 5$ получаем для доверительной вероятности $p = 0{,}95$ значение $t \approx 2$ [2] и, следовательно,

$$\Delta_l = \pm 2 \frac{s}{\sqrt{p}},$$

$$\Delta_m = \pm 1{,}78 \frac{s}{\sqrt{p}}. \qquad (12.47)$$

После проведения подобных расчетов может быть принято одно из следующих решений:

1. Начать новую фазу; в качестве центра нового плана выбирают либо точку плана предшествующей фазы с наилучшим значением целевой величины, либо какую-нибудь точку в направлении к оптимуму.

2. Провести изучение поверхности отклика в направлении к оптимуму.

3. Увеличить диапазоны варьирования факторов.

4. Включить в рассмотрение новые факторы, исключив изученные ранее.

Пример 12.8

Требуется увеличить выход химической реакции, проходящей в контактной печи, путем соответствующего выбора температуры T и давления D. Обычно используемая рабочая точка процесса соответствует режиму $T = 300°C$ и $D = 10$ атм. Для первой фазы метода ЭВОП выбраны следующие уровни варьирования: для T — 5°C, для D — 2 атм. Нормированные значения факторов обозначим через x_1 (температура) и x_2 (давление).

На фиг. 12.15 показан план эксперимента для первой фазы. Результаты экспериментов и их обработка по первому циклу первой фазы содержатся в табл. 12.7, а по второму циклу — в в табл. 12.8 [3]. Сравнение величин эффектов, полученных после

---

[1] Эффект считается значимым, если модуль его оценки превосходит модуль величины Δ_l, определяющей границы доверительного интервала.— *Прим. ред.*

[2] Значение t определяется здесь с помощью табл. 15.1 нормированной функции Лапласа из условия $\Phi(t) = (1/2)P$.— *Прим. ред.*

[3] В табл. 12.7—12.9 границы доверительного интервала для средних значений наблюдений в каждой точке плана (ДИ для новых средних) определяются выражением

$$\Delta = \frac{ts}{\sqrt{p}} \text{ при } t = 2.$$

Через S обозначена сумма оценок среднеквадратических ошибок наблюдений, рассчитанных, согласно (12.44), для всех реализованных циклов данной фазы.— *Прим. ред.*

второго цикла, с границами их доверительных интервалов свидетельствует об отсутствии значимых эффектов. Ввиду этого был реализован третий цикл (см. табл. 12.9). Оценка результатов этого цикла свидетельствует о значимом отличии от нуля эффекта температуры.

На этом первая фаза была завершена, и для точек плана были приняты целевые величины, полученные после третьего цикла (фиг. 12.15). Для дальнейшего экспериментирования следует принять одно из четырех перечисленных выше решений. Если предполагается начать новую фазу, то в качестве нового центра эксперимента целесообразно выбрать точку 3 плана первой фазы.

12.5. Методы случайного поиска в задачах экспериментальной оптимизации

Общие замечания

Стохастическими или случайными методами поиска называют методы, в которых для определения оптимума целевой величины или направления движения к экстремуму сознательно используются случайные элементы (случайные числа). Алгоритм поиска содержит при этом следующие три основных элемента, каждый из которых состоит в свою очередь из ряда последовательных этапов:

1. Алгоритм получения начальной точки поиска.
2. Решающее правило, с помощью которого определяется дальнейший ход поиска на основании данных, полученных путем реализации эксперимента.
3. Накопление информации, которая может быть полезной на последующих этапах поиска. Эта информация может различными способами использоваться на фазе обучения, или адаптации, процедуры поиска. Например, на основе накопленных данных может оказаться целесообразным изменить статистические свойства случайных элементов, используемых при поиске, с целью повышения скорости, точности или достоверности поисковой процедуры. Могут быть внесены изменения в значения параметров, характеризующих алгоритм поиска, которые позволят повысить его эффективность.

При применении простейших алгоритмов случайного поиска обычно обходятся без обучения и адаптации.

Конкретные методы случайного поиска

Слепой поиск

Из допустимой области факторного пространства, где по предположению должен находиться максимум целевой функции,

случайным образом выбирается точка, в которой и реализуется эксперимент. Величина целевой функции в этой точке сравнивается с величиной целевой функции в другой точке, полученной аналогично первой точке случайным образом. Наибольшее значение целевой величины и координаты соответствующей ей точки запоминаются. Затем вновь случайным образом выбирается точка из допустимой области, причем всякий раз в качестве начальной точки для последующего поиска используется лучшая из сравниваемых точек. При большой размерности n факторного пространства и при требовании достаточно точного определения положения оптимума использование этого метода приводит к необходимости реализации чрезмерно большого числа опытов. Необходимое число опытов N может быть рассчитано по формуле

$$N = \frac{\ln(1-P_\Delta)}{\ln(1-\Delta^n)} \approx \left(\frac{1}{\Delta}\right)^n \ln\frac{1}{1-P_\Delta}, \qquad (12.48)$$

где P_Δ — вероятность получения положения оптимума с точностью до величины Δ; Δ — граница точности по каждой координате факторного пространства (отношение модуля отклонения действительного положения экстремума x от экспериментально найденного $x_{опт}$ к диапазону изменения фактора [1]):

$$\Delta = \frac{|x - x_{опт}|}{x_{макс} - x_{мин}}.$$

Пример 12.9

Требуется найти положение оптимума в трехмерном пространстве, причем с вероятностью $P_\Delta = 0{,}8$ должна быть обеспечена точность по каждой координате, определяемая величиной $\Delta = 10^{-2}$. В соответствии с уравнением (12.48) необходимое число опытов составляет

$$N = (10^2)^3 \ln 5 \approx 1{,}6 \cdot 10^6.$$

Метод случайных направлений

Из некоторой точки допустимой области с известным значением целевой функции делается шаг величиной λ в случайном направлении и в полученной точке ставится эксперимент. Если целевая величина в этой точке оказывается большей, чем в начальной, шаг считается успешным и следующий шаг в случайном направлении делается из этой новой точки. Если же шаг оказывается неуспешным, осуществляется возврат к исходной точке. Поиск прекращается, если определенное число шагов не приводит к увеличению целевой функции. Число таких шагов

[1] При записи выражения (12.48), а также в примере 12.9 предполагается, что Δ для всех n факторов одинаково, а $|x_{макс} - x_{мин}| = 1$. — *Прим. ред.*

обычно выбирается равным размерности факторного пространства. Модификация метода заключается во введении шага переменной длины.

Метод случайного поиска с шагами в обратном направлении

Изложенный выше метод можно улучшить, если в случае неуспешного шага делать новый шаг из исходной точки в противоположном направлении. Если этот шаг также оказывается неуспешным, следует либо выбрать новое случайное направление, либо уменьшить величину шага. Эффективной модификацией этого метода является метод, предложенный Ноллау и Фюрстом [12].

Стохастический метод поиска Ноллау и Фюрста

Пусть требуется найти максимум функции $y = f(x_1, x_2, \ldots, x_n) > 0$ в некоторой области, заданной неравенствами

$$\begin{aligned} \alpha_1 &\leqslant x_1 \leqslant \beta_1, \\ \alpha_2 &\leqslant x_2 \leqslant \beta_2, \\ &\ldots\ldots \\ \alpha_n &\leqslant x_n \leqslant \beta_n. \end{aligned}$$

Исходя из некоторой точки x^m с координатами $x_1^m, x_2^m, \ldots, x_n^m$ и значением целевой функции $\tilde{y}^m$ определяются координаты следующего опыта путем реализации n-мерного случайного вектора x^{m+1}. Компоненты этого вектора x_k^{m+1}, $k = 1(1)n$, являются независимыми нормально распределенными случайными величинами. Математические ожидания μ_k^{m+1} и дисперсии $(s_k^2)^{m+1}$ этих случайных величин рассчитываются по результатам уже проведенных опытов. Предполагается, что максимальное значение целевой функции ограничено сверху величиной K. Кроме того, считается, что функция $f(x)$ вблизи экстремума изменяется монотонно.

Метод состоит из следующих шагов:

1. Выбор начальной точки.

Начальная точка выбирается из допустимой области на основе априорной информации о положении оптимума. Эту точку обозначаем x^0, а ее компоненты $(x_1^0, x_2^0, \ldots, x_n^0)$.

2. Выбор параметров процедуры.

В качестве математических ожиданий μ_k^1 выбирают значения координат начальной точки x^0, т. е.

$$\mu_1^1 = x_1^0, \quad \mu_2^1 = x_2^0, \quad \ldots, \quad \mu_n^1 = x_n^0. \tag{12.49}$$

Величину s_k^0 определяют по формуле

$$s_k^0 = \frac{1}{6}(\beta_k - \alpha_k). \qquad (12.50)$$

Существует особое правило вычисления вектора средних с компонентами $\mu_1^m, \mu_2^m, \ldots, \mu_n^m$. Этот вектор необходим для получения точки x^m $(m \geqslant 3)$. Правило его вычисления зависит от того, увеличилось или уменьшилось значение целевой функции при переходе из точки x^{m-2} к точке x^{m-1}, т. е. от того, имеет ли место $\tilde{y}^{m-1} > \max \tilde{y}^k + \Delta$ или $\tilde{y}^{m-1} \leqslant \max \tilde{y}^k + \Delta$, $k = 0, 1, \ldots, m-2$, где Δ — корректирующий фактор, зависящий от дисперсии ошибки измерения. Если значение целевой функции увеличилось, то среднее значение μ^m изменяется в направлении вектора $(x^{m-1} - x^*)$ (x^* — вектор координат точки, в которой наблюдался максимальный выход), так как можно ожидать, что и после точки x^{m-1} целевая функция будет расти. Если же значение целевой величины уменьшилось, то целесообразно изменить среднее значение μ^m в направлении вектора $(\mu^{m-1} - x^*)$, приближая его к x^*. Таким образом,

$$\mu^m = x^{m-1} + \frac{K - \tilde{y}^{m-1}}{K - \tilde{y}^0}[x^{m-1} - x^*], \qquad (12.51)$$

если

$$\tilde{y}^{m-1} > \max \tilde{y}^k + \Delta, \quad k = 0, 1, \ldots, m-2,$$

и

$$\mu^m = x^* + \frac{\tilde{y}^*}{K}[\mu^{m-1} - x^*], \qquad (12.52)$$

если

$$\tilde{y}^{m-1} \leqslant \max \tilde{y}^k + \Delta, \quad k = 0, 1, \ldots, m-2.$$

На фиг. 12.16 оба эти случая поясняются геометрически.

Величины $(K - \tilde{y}^{m-1})/(K - \tilde{y}^0)$ и $\tilde{y}^*/K$, используемые в (12.51) и (12.52), обеспечивают малые изменения вектора средних вблизи точки максимума.

Подобным же образом происходит формирование вектора $s^m = (s_1^m, s_2^m, \ldots, s_n^m)$. При значительном удалении от максимума изменение дисперсий происходит с большим шагом, вблизи максимума шаг изменения мал.

Для вектора s^m справедливо выражение

$$s^m = \frac{K - \tilde{y}^{m-1}}{K} s^0, \quad m = 1, 2, \ldots, \qquad (12.53)$$

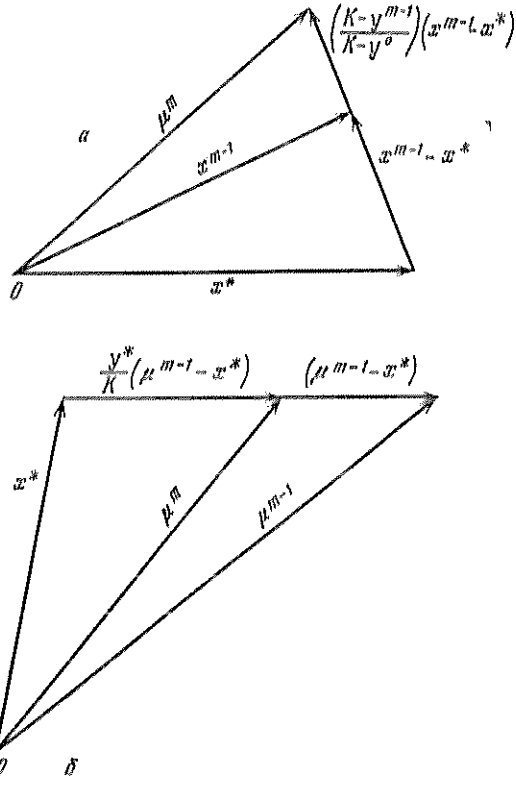
а для компонент соответственно выражение

$$s_i^m = \frac{K - \tilde{y}^{m-1}}{K} s_i^0, \quad i = 1, \ldots, n. \qquad (12.54)$$

Описанный способ вычисления параметров распределения представлен в табл. 12.11.

3. Формирование случайного вектора.

Случайный n-мерный вектор x^m формируется путем реализации случайного вектора ξ^m с независимыми и нормально рас-



Фиг. 12.16. Стратегия метода случайного поиска Фюрста и Ноллау.

a — вычисление среднего при улучшении значения целевой функции; $б$ — вычисление среднего при ухудшении значения целевой функции (под y всюду подразумевается $\tilde{y}$).

пределенными $N(0,1)$ компонентами $\xi_1^m, \xi_2^m, \ldots, \xi_n^m$. Значения $\xi_1^m, \xi_2^m, \ldots, \xi_n^m$ получают либо с помощью специальных таблиц, либо с помощью генераторов случайных чисел. В табл. 12.12, взятой из работы [12], приведена последовательность из 100 нормально распределенных чисел с нулевым математическим ожиданием и единичной дисперсией. Случайный вектор x^m вычисляется по формуле

$$x^m = \mu^m + s^m \xi^m \qquad (12.55)$$

Таблица 12.11

Схема вычислений метода случайного поиска по Фюрсту и Ноллау

Наблюдения	Ситуация	Вычисление μ^{m+1}	Вычисление s^{m+1}	Новая точка плана x^{m+1}
$\bar{y}^0$	—	$\mu^1 = x^0$	$s^1 = (K-\bar{y}^0)s^0/K$	$x^1 = \mu^1 \div s^1\xi^1$
$\bar{y}^1$	$\bar{y}^1 > \bar{y}^0 \div \Delta$ $\bar{y}^1 \leqq \bar{y}^0 \div \Delta$	$\mu^2 = x^1 \div \frac{K-\bar{y}^1}{K-\bar{y}^0}(x^1-x^0)$ $\mu^2 = x^0$	$s^2 = (K-\bar{y}^1)s^0/K$	$x^2 = \mu^2 \div s^2\xi^2$
$\bar{y}^2$	$\bar{y}^2 > \bar{y}^1 \div \Delta \geqq \bar{y}^0 \div \Delta$ $\bar{y}^2 > \bar{y}^0 \div \Delta \geqq \bar{y}^1 \div \Delta$ $\bar{y}^2 \leqq \bar{y}^0 \div \Delta (\bar{y}^0 > \bar{y}^1)$ $\bar{y}^2 \leqq \bar{y}^1 \div \Delta (\bar{y}^1 > \bar{y}^0)$	$\mu^3 = x^2 \div \frac{K-\bar{y}^2}{K-\bar{y}^0}(x^2-x^1)$ $\mu^3 = x^2 \div \frac{K-\bar{y}^2}{K-\bar{y}^0}(x^2-x^0)$ $\mu^3 = x^0 \div \bar{y}^0(\mu^2-x^0)/K$ $\mu^3 = x^1 \div \bar{y}^1(\mu^2-x^1)/K$	$s^3 = (K-\bar{y}^2)s^0/K$	$x^3 = \mu^3 \div s^3\xi^3$
$\bar{y}^3$	$\bar{y}^3 > \bar{y}^2 \div \Delta \geqq \bar{y}^2 \div \Delta \geqq \bar{y}^0 \div \Delta$ *или* $\bar{y}^3 > \bar{y}^2 \div \Delta \geqq \bar{y}^0 \div \Delta \geqq \bar{y}^1 \div \Delta$ $\vdots$	$\mu^4 = x^3 \div \frac{K-\bar{y}^3}{K-\bar{y}^0}(x^3-x^2)$ $\vdots$	$s^4 = (K-\bar{y}^3)s^0/K$ $\vdots$	$x^4 = \mu^4 \div s^4\xi^4$ $\vdots$

Таблица 12.12.

100 случайных чисел, распределенных по нормальному закону $N(0,1)$

1.	−0,47652	2,03453	−1,59887	−0,52145	1,62147
2.	−0,92939	−0,33436	−0,75248	0,47314	0,12502
3.	−0,15245	0,19792	−1,78770	0,89234	1,45258
4.	−0,51204	−0,75004	−0,22851	−0,85170	0,37981
5.	0,41065	−1,77429	−0,68040	−1,46103	0,01638
6.	−0,96600	0,56275	−0,12989	0,13363	−0,17620
7.	−1,57569	−1,41837	−0,37152	2,01655	−0,30411
8.	1,06089	0,88753	0,52225	0,44153	0,82016
9.	0,04392	−0,04455	1,90936	0,90739	0,54857
10.	0,00161	0,41266	−1,51238	−1,57477	0,61823
11.	1,82474	0,33218	−0,72842	−0,84215	0,15574
12.	−0,18540	−1,48422	1,25368	−0,60608	1,51844
13.	0,31464	0,23189	0,29688	0,19582	−0,88678
14.	1,25700	−0,10529	−0,21096	1,00753	0,63598
15.	0,10872	1,45267	0,54192	−1,54240	0,71482
16.	0,90763	−0,52674	−0,29008	0,57970	−0,45479
17.	−0,10819	0,14103	0,60586	−1,15918	−1,35639
18.	0,21693	1,30891	−0,63256	0,92959	−0,54006
19.	1,05750	−0,27029	2,18334	−1,39095	0,30105
20.	1,52464	0,87259	0,10225	1,43431	−0,93119

или для i-й компоненты

$$x_i^m = \mu_i^m + s_i^m \xi_i^m. \tag{12.56}$$

Пример 12.10 [12]

Требуется максимизировать выход формальдегида, получаемого при окислении метанола. Переменные процесса и диапазоны их изменения заданы следующим образом:

температура x_1^*: от 100 до 350° С;

концентрация метанола x_2^*: от 0,105 до 0,305 г/л;

расход воздуха x_3^*: от 10 до 50 л/ч;

ошибка эксперимента не превосходит величины $\Delta = 1\%$;

$K = 100\%$ (полное превращение метанола).

Подставляя приведенные данные в соответствующие формулы, получаем

$$s_1^0 = 40° \text{ С}, \quad s_2^0 = 0{,}03 \text{ г/л}, \quad s_3^0 = 6 \text{ л/ч}.$$

Результаты экспериментов и расчеты, связанные с реализацией метода, представлены в табл. 12.13. После четвертого опыта процесс поиска может быть прекращен, так как практически достигнут теоретический максимум выхода.

Таблица 12.13

Вычисления (пример 12.10)

Наблюдения	*Ситуация*	*Вычисление* μ^{cr}	*Вычисление* s^{cr}	*Новая точка плана*
$\hat{y}^0 = 8\%$	—	$\begin{pmatrix} 170 \\ 0{,}190 \\ 16{,}000 \end{pmatrix}$	$0{,}92 \begin{pmatrix} 40{,}00 \\ 0{,}03 \\ 6{,}00 \end{pmatrix}$	$\begin{pmatrix} 170 \\ 0{,}19 \\ 16{,}0 \end{pmatrix} \div \begin{pmatrix} 36{,}800 \\ 0{,}027 \\ 5{,}520 \end{pmatrix} \begin{pmatrix} 0{,}410 \\ -1{,}774 \\ -0{,}680 \end{pmatrix} = \begin{pmatrix} 185{,}088 \\ 0{,}142 \\ 12{,}246 \end{pmatrix}$
$\hat{y}^1 = 31\%$	$31 > 9$ $(\Delta = 1)$	$\begin{pmatrix} 185{,}088 \\ 0{,}142 \\ 12{,}246 \end{pmatrix} \div \frac{69}{92} \times \begin{pmatrix} 185{,}088 - 170{,}00 \\ 0{,}142 - 0{,}190 \\ 12{,}246 - 16{,}000 \end{pmatrix}$	$0{,}69 \begin{pmatrix} 40{,}00 \\ 0{,}03 \\ 6{,}00 \end{pmatrix}$	$\begin{pmatrix} 196{,}404 \\ 0{,}103 \\ 9{,}644 \end{pmatrix} \div \begin{pmatrix} 27{,}600 \\ 0{,}021 \\ 4{,}140 \end{pmatrix} \begin{pmatrix} -0{,}966 \\ 0{,}562 \\ -0{,}129 \end{pmatrix} = \begin{pmatrix} 169{,}743 \\ 0{,}115 \\ 9{,}110 \end{pmatrix}$
$\hat{y}^2 = 2\%$	$2 < 31$	$\begin{pmatrix} 185{,}088 \\ 0{,}142 \\ 12{,}246 \end{pmatrix} \div 0{,}31 \times \begin{pmatrix} 196{,}404 - 185{,}088 \\ 0{,}103 - 0{,}142 \\ 9{,}644 - 12{,}246 \end{pmatrix}$	$0{,}95 \begin{pmatrix} 40{,}00 \\ 0{,}03 \\ 6{,}00 \end{pmatrix}$	$\begin{pmatrix} 188{,}596 \\ 0{,}131 \\ 11{,}440 \end{pmatrix} \div \begin{pmatrix} 39{,}2 \\ 0{,}029 \\ 5{,}88 \end{pmatrix} \begin{pmatrix} 1{,}060 \\ 0{,}887 \\ 0{,}522 \end{pmatrix} = \begin{pmatrix} 230{,}122 \\ 0{,}157 \\ 14{,}509 \end{pmatrix}$
$\hat{y}^3 = 98\%$	*Останов*			

ЛИТЕРАТУРА

1. Autorenkollektiv, Analyse und Steuerung von Prozessen der Stoffwirtschaft. Akademie-Verlag, Berlin, 1971.
2. Wilde D. J., Optimum Seeking Methods, Prentice-Hall, 1964; русский перевод: Уайлд Д. Дж., Методы поиска экстремума, изд-во «Наука», М., 1967.
3. Hartmann K., диссертация, Ленинградский технологический институт, 1964.
4. Box G.E.P., Wilson K. B., *J. Roy. Stat. Soc.*, Ser. B, **13**, № 1 (1951).
5. Адлер Ю. П., Маркова Е. В., Грановский Ю. В., Планирование эксперимента при поиске оптимальных условий, изд-во «Наука», М., 1971.
6. Gardner D. A., *Ann. Math. Stat.*, 30, № 4, 1082 (1959).
7. Лисенков А. Н. и др., *Заводск. лаб.*, 33, № 10 (1966).
8. Hoerl A. E., *Chem. Eng. Progr.*, 55, № 11, 69 (1959).
9. Spendley W., Next G. R., Himsworth F. R., *Technometrics*, 4, 441 (1962).
10. Маркова Е. В., см. сб. «Планирование эксперимента», изд-во «Наука», М., 1966.
11. Box G.E.P., *Appl. Stat.*, 6, № 2, 81 (1957).
12. Fürst H., Müller H., Nollau V., *Chem. Technik*, 20, № 7, 400 (1968).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Химмельблау Д., Анализ процессов статистическими методами, изд-во «Мир», М., 1973.
2. Горский В. Г., Адлер Ю. П., Планирование промышленных экспериментов, изд-во «Металлургия», М., 1974.
3. Хикс Ч., Основные принципы планирования эксперимента, изд-во «Мир», М., 1967.
4. Круг Г. К., Ермуратский П. В., Экспериментально-статистические методы получения математического описания и оптимизации сложных технологических процессов (симплексный метод планирования экспериментов), вып. 4, НИИТЭХИМ, М., 1968.
5. Налимов В. В., Чернова Н. А., Статистические методы планирования экстремальных экспериментов, изд-во «Наука», М., 1965.
6. Растригин Л. А., Случайный поиск, изд-во «Знание», Рига, 1965.
7. Бояринов А. И., Кафаров В. В., Методы оптимизации в химической технологии, изд-во «Химия», М., 1975.
8. Draper N. R., Ridge Analysis of Response Surfaces, *Technometrics*, 5, № 4 (1963).
9. Слободчикова Р. И., Применение различных методов к исследованию поверхности отклика, *Заводск. лаб.*, 32, № 10 (1966).

13.
МЕТОДЫ КЛАССИФИКАЦИИ

13.1. Введение

Предметы, ситуации и явления, с которыми приходится сталкиваться в практике исследовательской работы, часто оказывается возможным разбить на группы или классы. При этом представители каждого класса обладают некоторыми общими свойствами, которые отсутствуют у представителей других классов. В дальнейшем мы будем предполагать, что классы состоят из „объектов", в качестве которых могут выступать упомянутые выше предметы, ситуации или явления.

Во многих случаях задача исследования состоит в определении класса, к которому относится некоторый изучаемый объект. Непосредственное измерение свойств, характеризующих классы, часто оказывается либо невозможным, либо связано с проведением дорогих или продолжительных испытаний. В то же время исследователь может достаточно просто получить другие данные о рассматриваемом объекте, которые в опосредованном виде содержат информацию о классе объекта. Для использования этой информации исследователь должен построить классифицирующее правило, т. е. правило принятия решения о классе по данным ряда измерений. Это правило строится с помощью методов классификации (или методов „распознавания образов").

Приведем несколько примеров возможного применения методов классификации для исследования и управления технологическими процессами:

1. Прогнозирование ситуаций (например, предсказание качества продукции, предупреждение аварий и т. п.).

2. Техническая диагностика процессов (опознавание вида неисправности, выбор способа ликвидации аварии).

3. Исследование устойчивости сложных процессов или систем регулирования.

Знание классифицирующего правила в этих задачах позволяет оперативно реагировать на изменения режимов работы установок, сократить потери времени при поиске неисправности и при ликвидации аварийных ситуаций, уменьшить затраты на исследование систем [1—4].

Для построения классифицирующего правила проводится экспериментальное исследование процесса, в ходе которого собирается „обучающая выборка". Эта выборка содержит некоторое число объектов, каждый из которых описан набором измеренных значений ряда переменных. В результате эксперимента определяется класс каждого объекта обучающей выборки. Результаты эксперимента используются для построения классифицирующего правила. Эту фазу решения задачи называют „обучением распознаванию", или просто „обучением".

После того как классифицирующее правило построено, класс произвольного объекта можно определить без проведения нового эксперимента.

Перейдем теперь к изложению математической постановки задачи классификации и способов ее решения.

13.2. Основные понятия и определения

Рассмотрим множество G некоторых объектов. Элементами этого множества могут являться, например, всевозможные способы написания букв какого-либо алфавита, физические предметы или явления, производственные ситуации, возникающие при управлении технологическими процессами, и т. д. Каждый объект множества G может быть описан совокупностью значений некоторых переменных. Пространство x размерности n, по осям $x_1, x_2, \ldots, x_n$ которого откладываются значения этих переменных, называется пространством переменных. Объект с номером j можно представить в этом пространстве в виде точки или вектора x^j с координатами $x_1^j, x_2^j, \ldots, x_n^j$. Подмножество объектов в рассматриваемом множестве G, объединенных заданной общностью, будем называть классом. Количество различных классов объектов множества G конечно и равно m. Предполагается, что каждому классу в пространстве x соответствует конечное число областей.

В дальнейшем мы будем рассматривать лишь случаи, когда имеются только два класса, т. е. $m=2$. Эти классы мы обозначим буквами A и B. Подмножества объектов, составляющих классы A и B, обозначим через G_A и G_B соответственно.

Заметим, что задачи с числом классов $m>2$ можно свести к последовательности задач с $m=2$, например, следующим способом:

1. Один из m классов обозначается A, а из остальных образуется „синтетический" класс B.
2. Находится классифицирующее правило для классов A и B.
3. Из класса B выделяется один из $(m-1)$ классов, который обозначается A', а оставшиеся — B'.

4. Находится классифицирующее правило для A' и B'; процедура продолжается до тех пор, пока в классе B не останется только один из исходных классов.

Решение задачи классификации заключается в построении классифицирующего правила, позволяющего по измеренным значениям n переменных x_i любого объекта из G отнести его к классу A или B. Обучающая выборка X_N содержит N объектов ($N = N_A + N_B$) из множества G, причем среди них есть представители обоих классов. Так как число N объектов в обучающей выборке конечно, то, вообще говоря, классифицирующее правило, построенное по этой выборке, может „ошибаться" при определении классов некоторых объектов.

Задачи классификации целесообразно разделить на две группы: детерминистские и статистические. Для задач первой группы характерно то, что в пространстве x классам A и B соответствуют непересекающиеся области. Решение заключается в нахождении в том или ином виде границ этих областей.

В статистических задачах области в пространстве x могут пересекаться. При решении статистических задач обычно рассматривают функции условных плотностей распределения вероятностей объектов классов A и B в пространстве x. Правило классификации ищется здесь с использованием статистических критериев [9—11].

Для каждой из этих групп задач разработаны специфические методы построения классифицирующего правила. Процесс решения обычно состоит из следующих этапов:

1. Выбор совокупности, включающей n переменных, наиболее существенных для классификации.
2. Построение классифицирующего правила.
3. Оценка точности найденного правила.

13.3. Выбор существенных переменных [1]

Как уже указывалось ранее, каждый объект представляется в виде совокупности измеренных значений некоторых переменных. Число переменных, которые могут быть измерены, часто оказывается очень большим, но не все они одинаково важны для решения задачи классификации. Включение большого числа переменных в классифицирующее правило не только затрудняет его нахождение, но и создает большие неудобства в его последующем применении. Кроме того, в некоторых случаях это приводит к снижению точности классификации, особенно при включении в правило так называемых „шумящих" переменных, изменение которых никак не связано с разделением объектов на классы.

[1] См. также разд. 7.

В связи с этим возникает задача выбора из исходных переменных некоторой их совокупности, в которую вошли бы переменные, наиболее существенные для решения поставленной задачи. Для нахождения такой совокупности необходимо уметь оценивать „существенность" различных переменных.

Пусть исходная совокупность состоит из r переменных x_1, x_2, ..., x_r. Допустим далее, что выбрана некоторая мера $U(x_i)$, принимающая меньшие значения для более существенных переменных. Требуется выбрать совокупность из n наиболее важных переменных.

Рассчитаем величины $U(x_i)$, $i = 1, 2, \ldots, r$, и расположим исходные переменные в виде последовательности

$$\{x_{i_1}, x_{i_2}, \ldots, x_{i_r}\} \tag{13.1}$$

в порядке возрастания меры $U(x_i)$, т. е. так, чтобы выполнялись условия

$$U(x_{i_1}) = \min_i U(x_i), \quad i = 1, 2, \ldots, r,$$

$$U(x_{i_2}) = \min_i U(x_i), \quad i = 1, 2, \ldots, i_1 - 1, i_1 + 1, \ldots, r,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

Если бы все переменные изменялись независимо, в искомую совокупность следовало бы включить первые n членов ряда (13.1). На практике, однако, переменные обычно оказываются взаимосвязанными, поэтому при выборе n переменных приходится использовать меры, учитывающие взаимодействие переменных.

Обозначим через W_l некоторую l-ю совокупность, включающую n ($n < r$) переменных x_{l_i}, $i = 1, 2, \ldots, n$, взятых из ряда $x_1, x_2, \ldots, x_r$. Меру, оценивающую эту совокупность, обозначим через $U(W_l)$. Для выбора наилучшей совокупности необходимо рассчитать такие меры для всех возможных совокупностей n переменных. Во многих задачах число таких совокупностей оказывается настолько большим, что исключает возможность их полного перебора. Поэтому приходится применять какие-либо иные способы поиска. Разработано несколько методов решения задачи поиска наилучшей совокупности переменных, отличающихся используемыми мерами оценки их существенности и способами сокращения перебора [5—8].

Рассмотрим один из методов [5], применяемый в статистических задачах классификации. Пусть имеется некоторая совокупность W_l из n переменных $x_{l_1}, x_{l_2}, \ldots, x_{l_n}$. В пространстве x этих переменных каждый объект представляется в виде n-мерного вектора x^l. Мера существенности $U(W_l)$ совокупности переменных W_l вычисляется следующим образом:

1. Высказываются предположения о виде функций плотностей распределения вероятности вектора x для объектов из классов A и B: $p_A(x)$ и $p_B(x)$, а также о величинах вероятностей появления объектов из классов A и B: $P(A)$ и $P(B)$.

2. По обучающей выборке вычисляются оценки неизвестных параметров распределений $p_A(x)$ и $p_B(x)$ (о способах нахождения таких оценок см. разд. 2).

3. Для каждой точки x^j обучающей выборки вычисляется величина $L(x^j)$, равная отношению вероятностей принадлежности объекта x^j к классам A и B:

$$L(x^j) = \frac{P(A)\, p_A(x^j)}{P(B)\, p_B(x^j)}.$$

4. Производится классификация точек обучающей выборки по правилу

$$\begin{aligned} x^j &\in A, \quad \text{если} \quad L(x^j) \geqslant 1, \\ x^j &\in B, \quad \text{если} \quad L(x^j) < 1. \end{aligned} \tag{13.2}$$

Пусть при этом окажется, что неправильно классифицируется N_1 точек класса A и N_2 точек класса B.

5. Вычисляется мера существенности рассматриваемой совокупности переменных

$$U(W_l) = P(A)\frac{N_1}{N_A} + P(B)\frac{N_2}{N_B}, \tag{13.3}$$

или, при равных вероятностях $P(A)$ и $P(B)$,

$$U(W_l) = \frac{1}{2}\left(\frac{N_1}{N_A} + \frac{N_2}{N_B}\right). \tag{13.4}$$

Наконец, если еще $N_A = N_B$, то

$$U(W_l) = \frac{N_1 + N_2}{N}. \tag{13.5}$$

Теперь задача состоит в том, чтобы выбрать такую совокупность W_l, для которой мера $U(W_l)$ минимальна. Как уже отмечалось выше, во многих задачах полный перебор совокупностей W_l оказывается невозможным. Для сокращения перебора применяют случайный поиск наилучшей совокупности с „поощрением" и „наказанием" отдельных переменных.

Перед началом поиска на основании априорных сведений об исследуемом объекте задаются вероятности $P(x_1), \ldots, P(x_r)$ выбора отдельных переменных $x_1, \ldots, x_r \left(\sum_{i=1}^{r} P(x_i) = 1\right)$. Если

такие сведения отсутствуют, полагают

$$P(x_1) = P(x_2) = \ldots = P(x_r) = \frac{1}{r}.$$

В процессе поиска наилучшей совокупности выполняются следующие операции:

1. Производится случайный выбор q совокупностей, содержащих по n переменных. При формировании этих совокупностей каждая переменная x_l выбирается с вероятностью $P(x_l)$. Полученные совокупности обозначаются $W_1, \ldots, W_q$.

2. Для этих совокупностей рассчитываются значения меры $U(W_1), \ldots, U(W_q)$ и находятся совокупности $W^{(1)}_{\text{мин}}$ и $W^{(1)}_{\text{макс}}$, для которых получены минимальная $U^{(1)}_{\text{мин}}$ и максимальная $U^{(1)}_{\text{макс}}$ меры.

3. Вероятности $P(x_l)$ выбора переменных, вошедших в $W^{(1)}_{\text{мин}}$, увеличиваются на заданную величину h за счет остальных переменных, а вероятности выбора переменных, вошедших в $W^{(1)}_{\text{макс}}$, уменьшаются на ту же величину h тоже за счет остальных переменных.

4. Величины $U^{(1)}_{\text{мин}}$ и $U^{(1)}_{\text{макс}}$ заносятся в рабочие ячейки d_1 и d_2 соответственно.

5. Производится случайный выбор новых q совокупностей, причем каждая переменная выбирается в соответствии с измененными вероятностями $P(x_l)$.

6. Рассчитываются меры $U(W_l)$, $l = 1, 2, \ldots, q$, для новых совокупностей и находятся $W^{(2)}_{\text{мин}}$ и $W^{(2)}_{\text{макс}}$ с минимальной $U^{(2)}_{\text{мин}}$ и максимальной $U^{(2)}_{\text{макс}}$ мерами соответственно.

7. Если $U^{(2)}_{\text{мин}} < U^{(1)}_{\text{мин}}$, то эта величина запоминается в ячейке d_1, а вероятности выбора переменных, входящих в $W^{(2)}_{\text{мин}}$, увеличиваются на величину h.

8. Если $U^{(2)}_{\text{макс}} > U^{(2)}_{\text{макс}}$, то $U^{(2)}_{\text{макс}}$ пересылается в ячейку d_2, а вероятности выбора переменных, входящих в $W^{(2)}_{\text{макс}}$, уменьшаются на величину h.

9. Процедура п. 5—8 повторяется до тех пор, пока содержимое ячейки d_1 не будет оставаться неизменным в течение заданного числа t шагов поиска.

После завершения поиска получаются наилучшая совокупность из n переменных, мера ее существенности и вероятности выбора отдельных переменных.

В рассмотренном алгоритме выбор величин t, q и h произволен, однако ясно, что вероятность нахождения наилучшей совокупности повышается с увеличением первых двух из этих величин. К сожалению, при этом увеличивается также и объем вычислений.

Заметим, что вместо (13.2) могут, вообще говоря, использоваться и другие классифицирующие правила.

13.4. Статистические методы классификации

В статистических задачах классификации любой точке пространства переменных могут соответствовать объекты, принадлежащие как классу A, так и классу B. Поэтому при решении таких задач приходится рассматривать функции условных плотностей распределения вероятностей вектора x для классов A и B: $p_A(x)$ и $p_B(x)$. Кроме того, часто считаются известными априорные вероятности $P(A)$ и $P(B)$ принадлежности всякого рассматриваемого объекта к классу A или B соответственно.

Среди статистических методов классификации можно выделить две большие группы. В первую группу входят методы, для применения которых требуется знание функций $p_A(x)$ и $p_B(x)$ [9, 10].

Функции $p_A(x)$ и $p_B(x)$ обычно бывают известны лишь с точностью до значений ряда параметров, величины которых оцениваются по обучающей выборке до решения собственно задачи классификации. В связи с этим методы этой группы иногда называют параметрическими.

Вторую группу образуют методы, применение которых не связано со знанием вида функций $p_A(x)$ и $p_B(x)$. Классифицирующее правило получается в них для всякой точки x' путем рассмотрения точек обучающей выборки в некоторой окрестности x'. Методы этой группы называют локальными. Ниже рассматриваются методы обеих названных групп.

13.4.1. Параметрические методы

При изложении этого раздела будем полагать, что $p_A(x)$ и $p_B(x)$ — плотности вероятностей нормальных законов распределения. Обозначим условные математические ожидания вектора x для классов A и B через m^A и m^B соответственно. Ковариационные матрицы распределения вектора x для классов A и B будем считать одинаковыми, т. е.

$$\mathrm{cov}^A(x) = \mathrm{cov}^B(x) = \mathrm{cov}(x).$$

Тогда функции плотностей вероятности $p_A(x)$ и $p_B(x)$ запишутся в виде

$$\begin{aligned} p_A(x) &= d \cdot \exp\left\{-\frac{1}{2}(x-m^A)' C (x-m^A)\right\}, \\ p_B(x) &= d \cdot \exp\left\{-\frac{1}{2}(x-m^B)' C (x-m^B)\right\}, \end{aligned} \tag{13.6}$$

где d — постоянный множитель; C — матрица, обратная $\mathrm{cov}(x)$:

$$C = \mathrm{cov}^{-1}(x).$$

Согласно теореме Байеса, условные вероятности принадлежности объекта x^j к классам A и B равны

$$P(A \mid x^j) = \frac{P(A) \cdot p_A(x^j)}{P(A) \cdot p_A(x^j) + P(B) \cdot p_B(x^j)},$$
$$P(B \mid x^j) = \frac{P(B) \cdot p_B(x^j)}{P(A) \cdot p_A(x^j) + P(B) \cdot p_B(x^j)}. \tag{13.7}$$

Применим следующее правило классификации. Будем относить объект x^j к тому классу, которому соответствует бо́льшая условная вероятность, т. е.

$$x^j \in A, \quad \text{если} \quad P(A \mid x^j) \geqslant P(B \mid x^j),$$
$$x^j \in B, \quad \text{если} \quad P(A \mid x^j) < P(B \mid x^j). \tag{13.8}$$

Из (13.7) и (13.8) следует, что классификация объекта x^j производится в зависимости от величины

$$L(x^j) = \frac{P(A) \cdot p_A(x^j)}{P(B) \cdot p_B(x^j)} = \frac{P(A)}{P(B)} \cdot l(x^j). \tag{13.9}$$

Величину $l(x^j)$ принято называть отношением правдоподобия. Правило (13.8) при этом можно записать в виде

$$x^j \in A, \quad \text{если} \quad L(x^j) \geqslant 1,$$
$$x^j \in B, \quad \text{если} \quad L(x^j) < 1. \tag{13.10}$$

Рассмотрим граничное значение

$$L(x) = 1. \tag{13.11}$$

Подставив выражения (13.6) для функций плотностей вероятности $p_A(x)$ и $p_B(x)$ в (13.11), после логарифмирования и несложных преобразований получим

$$F(x) = x' C (m^A - m^B) =$$
$$= \frac{1}{2} (m^A + m^B)' C (m^A - m^B) - \ln \frac{P(A)}{P(B)}. \tag{13.12}$$

В пространстве переменных x уравнение (13.12) описывает гиперплоскость, разбивающую его на две области Ω_A и Ω_B, причем в области Ω_A принимается решение об отнесении объектов к классу A, а в области Ω_B — к классу B.

Левая часть уравнения (13.12) является алгебраической функцией от x. Эту функцию обычно называют дискриминантной. Принятие решения о классификации с помощью уравнения (13.12) можно интерпретировать как сравнение значения дискриминантной функции $F(x)$ в некоторой точке x^j с величиной порога y_0,

определяемой правой частью выражения (13.12):

$$\begin{aligned} x' \in A, &\quad \text{если} \quad F(x') \geqslant y_0, \\ x' \in B, &\quad \text{если} \quad F(x') < y_0, \end{aligned} \tag{13.13}$$

где

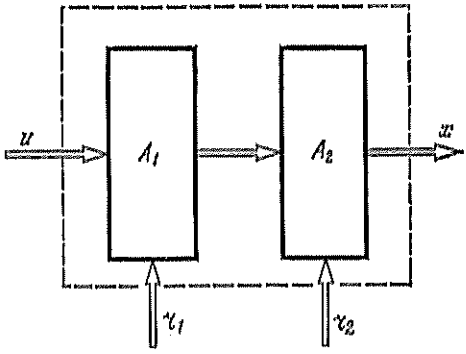
$$y_0 = \frac{1}{2}(m^A + m^B)\, C\, (m^A - m^B) - \ln \frac{P(A)}{P(B)},$$

$$F(x) = x' C\, (m^A - m^B).$$

Заметим, что статистические методы классификации, в которых решающее правило представляется с помощью алгебраической функции, коэффициенты которой зависят от параметров плотностей вероятности $p_A(x)$ и $p_B(x)$, называются методами дискриминантного анализа.

Пример 13.1

Рассмотрим задачу диагностики неисправностей некоторого технологического процесса, состоящего из двух последовательно включенных аппаратов A_1 и A_2 (фиг. 13.1). Обозначим через u



Фиг. 13.1. Схема технологической системы в примере 13.1.

вектор, характеризующий свойства исходного сырья, поступающего на вход процесса; r_1 и r_2 — векторы возмущений в аппаратах A_1 и A_2; x — вектор переменных, характеризующих выходной продукт. Предполагается, что измерены могут быть лишь u и x. Для упрощения положим, что в процессе работы вектор u остается неизменным, и потому для выявления неисправностей следует рассматривать лишь x. По величинам компонент вектора x можно различать три состояния системы: 1) исправная работа; 2) неисправность одного из аппаратов, 3) неисправность двух аппаратов. Наибольшие трудности вызывает определение номера неисправного аппарата во втором случае. В силу несимметрич-

ности системы неисправности в A_1 и A_2 по-разному влияют на значение вектора x. Однако наличие неконтролируемых возмущений r_1 и r_2 приводит к тому, что по величине x нельзя однозначно установить, какой именно аппарат вышел из строя.

Применим для решения задачи выявления неисправного аппарата методы классификации. Будем считать вектор x' объектом класса A, если неисправен аппарат A_1, и объектом класса B, если неисправен аппарат A_2. Пусть известно, что объекты обоих классов распределены нормально, а априорные вероятности классов заданы: $P(A) = 0{,}4$, $P(B) = 0{,}6$.

Параметры распределений $p_A(x)$ и $p_B(x)$ (здесь размерность x $n = 2$):

$$m^A = \begin{pmatrix} -1{,}5 \\ 3{,}4 \end{pmatrix}, \quad m^B = \begin{pmatrix} 0{,}7 \\ 7{,}4 \end{pmatrix},$$

$$\operatorname{cov}(x) = \begin{pmatrix} 3{,}367 & 3{,}740 \\ 3{,}740 & 8{,}936 \end{pmatrix}.$$

Обратная матрица

$$C = \begin{pmatrix} 0{,}556 & -0{,}233 \\ -0{,}233 & 0{,}210 \end{pmatrix}.$$

Требуется найти классифицирующее правило и отнести с его помощью точку x^1 (2,0; 7,0) к тому или иному классу.

Рассчитаем пороговое значение

$$y_0 = \frac{1}{2}(-0{,}8 \;\; 10{,}8) \begin{pmatrix} 0{,}556 & -0{,}233 \\ -0{,}233 & 0{,}210 \end{pmatrix} \begin{pmatrix} -2{,}2 \\ -4{,}0 \end{pmatrix} - \ln \frac{0{,}4}{0{,}6} = -1{,}25.$$

Найдем дискриминантную функцию

$$y = F(x) = -0{,}29x_1 - 0{,}33x_2.$$

Классифицирующее правило имеет вид

$$x' \in A, \text{ если } -0{,}29x_1 - 0{,}33x_2 \geqslant -1{,}25,$$
$$x' \in B, \text{ если } -0{,}29x_1 - 0{,}33x_2 < -1{,}25.$$

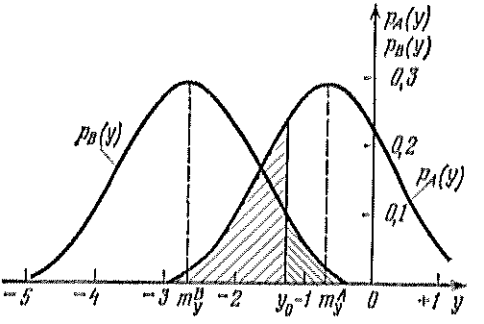
Для классифицируемой точки x^1

$$F(x^1) = -2{,}89 < -1{,}25.$$

Следовательно, точка x' должна быть отнесена к классу B. На фиг. 13.2 показаны распределения $p_A(y)$ и $p_B(y)$ для $y = F(x)$, а также пороговое значение y_0 для данной задачи.

При выводе уравнения граничной поверхности (13.12) предполагалось, что ковариационные матрицы для классов A и B одинаковы. Если это предположение не выполняется, граничная

поверхность перестает быть линейной. Решение задачи в этом случае становится более сложным, но все же может быть найдено с помощью процедуры, аналогичной изложенной выше. Гораздо труднее найти решение в задачах, в которых вид законов распределений вероятности появления объектов из классов A и B отличается от нормального. Для таких случаев разработан ряд специальных методов классификации, рассмотренных, например, в [10, 11].



Фиг. 13.2. Распределения вероятностей величины $y = F(x)$ для классов A и B.

Значения параметров распределений $p_A(x)$ и $p_B(x)$, если они не заданы, могут быть оценены по обучающей выборке. Этому могут сопутствовать различного рода статистические расчеты, связанные с проверкой гипотез о нормальности законов распределения, равенстве ковариационных матриц и т. п. В случае многомерных задач классификации для проверки таких гипотез используются методы многомерного статистического анализа [9].

13.4.2. Локальные методы

В рассмотренных в предыдущем разделе статистических методах классификации необходимо было заранее знать вид распределений в пространстве x объектов классов A и B. Во многих случаях, однако, весьма трудно высказать какую-либо гипотезу о виде этих распределений. В таких ситуациях могут быть использованы локальные методы [10, 11], в которых решение о классификации произвольной точки x' принимается после рассмотрения точек обучающей выборки, попадающих в малую окрестность x'. Если в методах первой группы для получения функций $p_A(x)$ и $p_B(x)$ необходимо было по обучающей выборке оценить неизвестные параметры распределений, то локальные методы по существу имеют дело с построением некоторой кусочной аппроксимации функций $p_A(x)$ и $p_B(x)$.

Рассмотрим алгоритм одного из наиболее простых локальных методов классификации, обычно называемого методом Фикса — Ходжеса. Метод этот заключается в следующем:

1. Вводится понятие „расстояния" между любыми двумя точками x' и x^t в пространстве переменных x. В частном случае это может быть евклидово расстояние

$$r(x', x^t) = \left[\sum_{i=1}^{n} (x_i' - x_i^t)^a\right]^{1/a}. \tag{13.14}$$

2. В зависимости от числа N объектов в обучающей выборке X_N и априорных сведений о задаче выбирается некоторое нечетное число $N' < N$.

3. Строится сфера наименьшего радиуса r_0 с центром в изучаемой точке x', такая, что в нее попадает не менее N' точек из выборки X_N. Пусть из них N'_A точек принадлежит классу A и N'_B — классу B.

4. Точка x' классифицируется по правилу

$$\begin{aligned} x' &\in A, \quad \text{если } N'_A \geqslant N'_B, \\ x' &\in B, \quad \text{если } N'_A < N'_B. \end{aligned} \tag{13.15}$$

В более сложных модификациях локальных методов используются другие меры расстояний между точками и иные правила классификации. Так, в работе [10] получено выражение оценки отношения правдоподобия

$$l(x') = \frac{p_A(x')}{p_B(x')}$$

в некоторой точке x', использующее лишь информацию о классификациях точек, попадающих в малую сферическую окрестность точки x'. Эта оценка имеет вид

$$l(x') = \frac{(N'_A - 1) \cdot N_B}{(N'_B - 1) \cdot N_A} \cdot \left(\frac{r_A}{r_B}\right)^n. \tag{13.16}$$

Здесь r_A и r_B — расстояния от точки x' до наиболее удаленных точек классов A и B из обучающей выборки в сферической окрестности точки x'; n — размерность пространства x. Решение для точки x' при этом принимается по правилу (13.10). При $N_A = N_B$ и $r_A = r_B$ правило (13.16) совпадает с правилом (13.15).

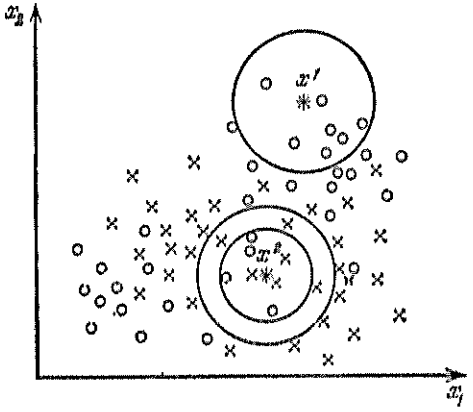
Пример 13.2

В двумерном пространстве переменных точки x' обучающей выборки X_N имеют расположение, показанное на фиг. 13.3. Выборка X_N содержит $N = 70$ точек с известными классификациями, причем $N_A = N_B = 35$. Объекты класса A обозначены

кружками, класса B — крестиками. Необходимо найти классификации точек x^1 и x^2, указанных на рисунке.

Решение. Распределения объектов из классов A и B нам не известны. Из фиг. 13.3 следует, что они могут иметь достаточно сложный, многомодальный характер. Методы классификации, рассмотренные в разд. 13.4.1, здесь окажутся неэффективными.

Применим для классификации точек x^1 и x^2 метод Фикса — Ходжеса. Расстояние между точками будем при этом рассчитывать по формуле (13.14). Число N' зададим равным семи. Рассмотрим точку x^1. Постепенно увеличивая радиус сферы, описанной



Фиг. 13.3. Расположение точек обучающей выборки в пространстве переменных (пример 13.2).

вокруг этой точки, добиваемся, чтобы в нее попало семь точек. Поскольку все эти точки принадлежат классу A, относим x^1 также к классу A.

Точно так же поступаем при определении классификации точки x^2. Из семи точек, попавших в ее сферическую окрестность, к классу A принадлежит $N'_A = 4$ точки, к классу B $N'_B = 3$ точки. Если принимать решение по правилу (13.15), x^2 должна быть отнесена к классу A. Однако расстояние до самой дальней точки класса B оказывается значительно меньше, чем до самой дальней точки класса A в этой окрестности. Поэтому правилом (13.16) точка x^2 будет отнесена к классу B. Заметим, что при незначительном увеличении радиуса сферической окрестности точки x^2 в нее попадает сразу три точки класса B и x^2 будет отнесена к классу B также и правилом (13.15).

В локальных методах классификации затруднение вызывает выбор числа N', определяющего размеры сферических окрестностей классифицируемых точек. Выбор этого числа никак не фор-

мализован и должен производиться в зависимости от условий каждой конкретной задачи: числа N объектов в обучающей выборке, примерного вида распределений $p_A(x)$ и $p_B(x)$, размерности n пространства переменных x, используемого правила классификации.

Практическая реализация методов локальной классификации требует, как правило, использования цифровых вычислительных машин. При классификации любого объекта здесь необходимо рассматривать всю обучающую выборку, которая, следовательно, должна постоянно находиться в памяти вычислительной машины.

13.5. Детерминистские методы решения задачи классификации

Детерминистская постановка задачи классификации предполагает, что объекты с тождественными значениями переменных относятся всегда к одному и тому же классу. Следовательно, в пространстве x классам A и B соответствуют непересекающиеся области. Такие задачи в принципе могут рассматриваться и как статистические, если считать, что каждая точка относится к одному из классов с вероятностью 1, к другому — с вероятностью 0. Однако имеется большое число методов (например, описанные в работах [12—16]), которые предназначены для решения именно детерминистских задач, и их применение в этих случаях часто оказывается более простым и эффективным. Рассмотрим некоторые из этих методов.

13.5.1. Интерполяционные методы построения классифицирующего правила

Введем понятие классифицирующей функции:

$$y^j = \begin{cases} +1, & \text{если } x^j \in A, \\ -1, & \text{если } x^j \in B. \end{cases} \tag{13.17}$$

Заметим, что для объектов обучающей выборки значения этой функции известны и их можно записать в виде вектора $\tilde{Y}'_N = (\tilde{y}^1, \ldots, \tilde{y}^N)$. Используя эту информацию, можно с помощью тех или иных интерполяционных методов найти значения этой функции в точках пространства x, не входящих в обучающую выборку.

Рассмотрим решение задачи с помощью метода наименьших квадратов. Будем искать интерполирующую функцию для y в виде

$$y = a'f(x),$$

где a — вектор неизвестных коэффициентов,

$$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_k \end{bmatrix},$$

$f(x)$ — вектор заданных функций, зависящих лишь от x:

$$f(x) = \begin{bmatrix} f_0(x) \\ f_1(x) \\ f_2(x) \\ \cdot \\ \cdot \\ \cdot \\ f_k(x) \end{bmatrix}.$$

Предполагается, что выполняется условие

$$k + 1 \leqslant N.$$

В качестве критерия построения интерполирующей функции используем критерий метода наименьших квадратов:

$$Q(a) = \sum_{j=1}^{N} [\tilde{y}^j - a' f(x^j)]^2, \qquad (13.18)$$

где $x^j (j = 1, 2, \ldots, N)$ — объекты обучающей выборки.

Оптимальным с точки зрения критерия (13.18) является вектор $\hat{a}$, такой, что

$$Q(\hat{a}) = \min_a Q(a).$$

После того как найден этот вектор, решение о классификации произвольного объекта x^l принимается по знаку интерполирующей функции в точке x^l.

Приведенный метод решения не позволяет в ряде случаев получить необходимую точность классификации. Это вызвано тем, что классифицирующая функция часто имеет сложный вид. Из (13.17) следует, что y претерпевает разрыв при переходе из области одного класса в область другого класса. Кроме того, здесь для построения интерполирующей функции используется критерий, отличный от того, по которому обычно оценивают точность классифицирующего правила, например вероятность ошибочной классификации. В результате может найтись интер-

полирующая функция, „хорошая" с точки зрения критерия (13.18), но непригодная для классификации объектов, не входящих в обучающую выборку.

13.5.2. Методы построения разделяющей поверхности

Предположим, что в пространстве x существует по крайней мере одна поверхность

$$F(x) = 0, \tag{13.19}$$

такая, что

$$\begin{aligned} F(x^j) &\geqslant 0 \text{ для любого } x^j \in G_A, \\ F(x^t) &< 0 \text{ для любого } x^t \in G_B. \end{aligned} \tag{13.20}$$

Это означает, что все объекты класса A находятся по одну сторону от этой поверхности, а все объекты класса B — по другую ее сторону. Такая поверхность называется разделяющей. Из свойств (13.20) следует, что разделяющая поверхность и классифицирующая функция связаны между собой соотношением

$$y = \operatorname{sign}\{F(x)\}, \tag{13.21}$$

где

$$\operatorname{sign} Z = \begin{cases} +1, & \text{если } Z \geqslant 0, \\ -1, & \text{если } Z < 0. \end{cases}$$

Таким образом, если бы удалось найти хорошую аппроксимацию разделяющей поверхности (АРП)

$$F(x, a) = 0,$$

где a — вектор коэффициентов, то решение о классификации того или иного объекта x^j можно было бы принимать в зависимости от знака $F(x^j, a)$:

$$\begin{aligned} x^j \in A, &\quad \text{если} \quad F(x^j, a) \geqslant 0, \\ x^j \in B, &\quad \text{если} \quad F(x^j, a) < 0. \end{aligned} \tag{13.22}$$

Пусть, например, требуется построить аппроксимацию разделяющей поверхности в линейном относительно вектора a виде

$$F(x, a) = a' f(x) = 0. \tag{13.23}$$

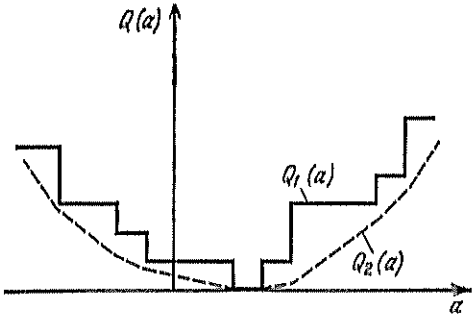
Здесь $f(x)$ обозначает вектор функций, зависящих только от x. Будем стремиться так выбрать вектор a, чтобы знак $F(x^j, a)$ совпадал со знаком $F(x^j)$ для возможно большего числа объектов x^j из обучающей выборки. Тогда в качестве величины, характеризующей точность аппроксимации разделяющей функции,

естественно выбрать следующую:

$$Q_1(a) = \frac{1}{2}\sum_{j=1}^{N}\{1 - \operatorname{sign}[\tilde{y}^j F(x^j, a)]\}. \qquad (13.24)$$

Очевидно, что $Q_1(a) \geqslant 0$.

Вектор $\hat{a}$, обеспечивающий минимум целевой функции (13.24), соответствует поверхности (13.23), разделяющей объекты обучающей выборки с наименьшим числом ошибок. В том случае, когда разделение производится безошибочно, имеем



Фиг. 13.4. Вид функций $Q_1(a)$ и $Q_2(a)$, используемых в критерии построения разделяющей поверхности.

$$Q_1(\hat{a}) = 0.$$

На практике использование целевой функции (13.24) вызывает определенные затруднения, так как она имеет кусочно постоянный характер. На фиг. 13.4 показан пример такой функции для одномерной задачи. За минимум функции легко принять точку a, лежащую на одной из площадок. Кроме того, применение широко распространенных методов поиска экстремума, основанных на вычислении градиента, оказывается здесь невозможным. Для решения здесь могут быть использованы методы, в которых для выбора направления движения не требуется вычислять градиент функции, например симплексный или комплексный методы [17, 18].

Видоизменим функцию, оценивающую точность аппроксимации разделяющей поверхности, так, чтобы сделать ее кусочно линейной. При этом желательно придать линейным участкам наклон в сторону минимума функции. Для этого зададим ее в таком виде:

$$Q_2(a) = r\sum_{j=1}^{N} R\{\tilde{y}^j F(x^j, a)\}, \qquad (13.25)$$

где r — коэффициент пропорциональности, а

$$R\{Z\} = \begin{cases} 0, & \text{если } Z \geqslant 0, \\ |Z|, & \text{если } Z < 0. \end{cases}$$

На фиг. 13.4 показан вид функции $Q_2(a)$ для уже упоминавшейся одномерной задачи. Величина $Q_2(a)$ зависит не только от числа несовпадений знаков функций $F(x^j, a)$ и $F(x^j)$ в точках обучающей выборки X_N, но и от значения $F(x^j, a)$ для неправильно классифицируемых точек. Если с помощью некоторой АРП все точки из X_N классифицируются правильно, то

$$Q_2(a) = 0;$$

в остальных случаях

$$Q_2(a) > 0.$$

Поскольку $Q_2(a)$ является непрерывной и кусочно линейной функцией по a, то для поиска ее минимума может быть использован практически любой из известных методов, например градиентный метод. Градиент функции $Q_2(a)$ по a равен

$$\nabla_a Q_2(a) = -\frac{1}{2} r \sum_{j=1}^{N} \{1 - \operatorname{sign}[\tilde{y}^j F(x^j, a)]\} \tilde{y}^j \nabla_a F(x^j, a).$$

Поскольку процесс поиска может прекратиться до достижения экстремального значения целевой функции, поиск повторяют несколько раз, используя разные начальные приближения $\hat{a}^{(0)}$. За искомое решение принимается точка, в которой за время поиска получена наименьшая величина $Q_2(a)$. Заметим, что поиск можно прекратить и раньше, если в некоторой точке $\hat{a}^{(l)}$ будет получено

$$Q_2(\hat{a}^{(l)}) = 0,$$

так как значения критериев (13.24) и (13.25) не могут быть меньше нуля. Такой результат означает, что найденная поверхность

$$F(x, \hat{a}^{(l)}) = 0$$

безошибочно разделила точки обучающей выборки X_N на два класса. Если в результате поиска минимума функции $Q_2(a)$ окажется

$$\min_a Q_2(a) > 0,$$

то это значит, что разделяющая поверхность имеет вид, отличающийся от рассматриваемого. В этом случае нужно либо изменить вид рассматриваемых поверхностей (13.23), либо перейти

к другим методам построения аппроксимации разделяющей поверхности, специально предназначенным для нахождения АРП сложного вида.

Пример 13.3

Рассмотрим задачу нахождения границы устойчивости экзотермической реакции, протекающей в проточном реакторе с внешним охлаждением [20]. При некоторых условиях охлаждения и концентрациях реагентов наблюдаются периодические колебания температуры реакции. Это явление будем называть неустойчивой реакцией. Взаимосвязь температуры, концентрации реагентов и условий охлаждения описывается двумя дифференциальными уравнениями вида

$$\frac{dc}{dt} = a_1 (c_0 - c) - r(c, T),$$

$$\frac{dT}{dt} = a_1 (T_0 - T) + a_2 r(c, T) - a_1 K (T - T_R),$$

где c — концентрация реагентов в реакторе; c_0 — концентрация реагентов в исходном продукте; T — температура в реакторе; T_0 — температура исходного продукта; T_R — температура теплоносителя на входе системы охлаждения; K — переменная, зависящая от условий охлаждения; $r(c, T)$ — скорость реакции; a_1, a_2 — константы, зависящие от объема реактора, расхода продукта на его входе, плотности и теплоемкости смеси реагентов.

Обозначим установившееся значение концентрации реагентов через c_s и введем новую переменную $V = c_s/c_0$. Требуется построить границу области устойчивости реакции в пространстве переменных V и K. Величины c_0, T_0, T_R, a_1, a_2, а также вид функции $r(c, T)$ считаются известными и неизменными во времени.

Решение. Для любых заданных значений V и K решение уравнений, описывающих реакцию, может быть найдено на аналоговой или цифровой вычислительной машине. Пусть, например, эти уравнения решены для ряда точек пространства переменных x. Будем считать точки, в которых реакция оказалась устойчивой, относящимися к классу A, а остальные — к классу B. Тогда разделяющей поверхностью для этих классов будет служить граница области устойчивости реакции.

Пусть в результате экспериментов получена обучающая выборка X_{16} (табл. 13.1) [1], в которую вошло 8 объектов класса A и 8 объектов класса B. Будем искать уравнение границы области устойчивости в виде полинома второй степени от V и K. Тогда, минимизируя функцию (13.25), можно найти, например, такую

[1] При составлении этой выборки использовались результаты, приведенные в работе Х. Брандеса [20].

поверхность, безошибочно разделяющую точки выборки X_{16}:

$$-1{,}0-26{,}74V+0{,}8452K+39{,}66V^2-0{,}0171K^2-0{,}5258KV=0.$$

Вид этой поверхности показан на фиг. 13.5.

Таблица 13.1

Обучающая выборка (пример 13.3)

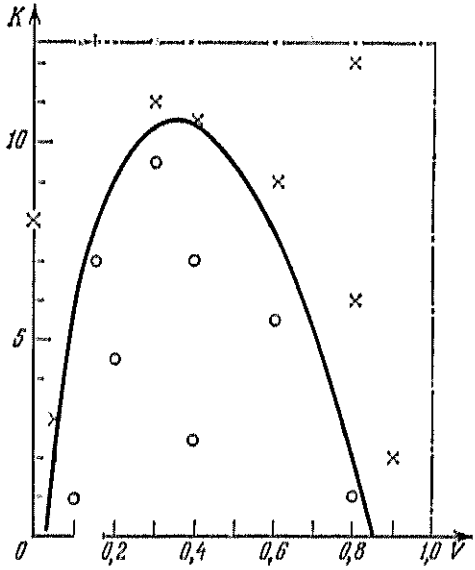
l	V^l	K^l	$\tilde{y}^l$	l	V^l	K^l	$\tilde{y}^l$
1	0,00	8,0	+1	9	0,40	7,0	−1
2	0,05	3,0	+1	10	0,60	5,5	−1
3	0,30	11,0	+1	11	0,20	4,5	−1
4	0,80	12,0	+1	12	0,40	2,5	−1
5	0,60	9,0	+1	13	0,10	0,5	−1
6	0,80	6,0	+1	14	0,80	0,5	−1
7	0,90	2,0	+1	15	0,40	10,5	+1
8	0,30	9,5	−1	16	0,15	7,0	−1

Применение методов классификации для решения задач, подобных рассмотренной выше, оказывается особенно выгодным в тех случаях, когда объект имеет сложную структуру, но достаточно легко могут быть реализованы эксперименты по определению устойчивости в каких-либо точках пространства переменных. Заметим, что эти эксперименты могут проводиться как на самом объекте, так и на его модели.

13.6. Оценка точности классифицирующего правила

При решении задачи классификации мы располагаем не всем множеством объектов G, а лишь обучающей выборкой X_N. Поэтому правило, являющееся наилучшим по этой выборке, может оказаться недостаточно хорошим для всего множества G. Это обстоя-

тельство делает необходимой оценку точности найденного правила, причем такую оценку следует производить по объектам, не входящим в обучающую выборку. На основании этой оценки может решаться вопрос о том, правильно ли был выбран вид классифицирующего правила.



Фиг. 13.5. Расположение точек обучающей выборки и аппроксимации разделяющей поверхности в пространстве переменных (пример 13.3).

Будем характеризовать качество классифицирующего правила с помощью вероятности ошибочной классификации:

$$P_{\text{ош}} = P^{A}_{\text{ош}} + P^{B}_{\text{ош}},$$

где $P^{A}_{\text{ош}}$ и $P^{B}_{\text{ош}}$—вероятности ошибочной классификации объектов класса A и B соответственно.

Рассмотрим два способа нахождения оценок $P_{\text{ош}}$, пригодных как для статистических, так и для детерминистских задач. Первый и наиболее часто применяемый из них состоит в следующем:

1. Исходная обучающая выборка X_N разделяется случайным образом на две подвыборки X_{N_1} и X_{N_2}.

2. По подвыборке X_{N_1} производится построение классифицирующего правила заданного вида.

3. По найденному правилу производится классификация объектов из X_{N_2} и результаты ее сравниваются с известной истинной классификацией.

4. Точность классифицирующего правила оценивается величиной

$$\hat{P}_{\text{ош}} = P(A)\frac{n^A_{\text{ош}}}{N_A} + P(B)\frac{n^B_{\text{ош}}}{N_B}, \qquad (13.26)$$

где $n^A_{\text{ош}}$ и $n^B_{\text{ош}}$—число ошибочных классификаций объектов подвыборки X_{N_2} из классов A и B соответственно; N_2—число объектов в подвыборке X_{N_2} ($N_2 = N_A + N_B$). Заметим, что для достоверности оценки (13.26) необходимо увеличивать число N_2 объектов в подвыборке X_{N_2}. Однако общее число объектов для обучения ограничено и увеличение N_2 приводит к уменьшению числа N_1 объектов в выборке X_{N_1}, по которой строится классифицирующее правило. Это вызывает в свою очередь ухудшение точности правила. Поэтому для оценки точности классифицирующего правила по описанному методу важно правильно задать числа N_1 и N_2. В то же время для большинства задач невозможно заранее дать какие-либо рекомендации относительно выбора этих чисел.

От указанных недостатков в значительной мере свободен второй способ оценки точности найденных правил [19]. Он заключается в следующем:

1. Исходная обучающая выборка делится, как и ранее, на подвыборки X_{N_1} и X_{N_2}, но теперь в X_{N_2} включается только один объект, выбранный из X_{N_2} ($N_2 = 1$).

2. По подвыборке X_{N_1} находится классифицирующее правило.

3. С его помощью классифицируется объект, включенный в X_{N_2}.

4. Выборки X_{N_1} и X_{N_2} вновь объединяются.

5. Процедура, соответствующая п. 1—4, повторяется заданное число N раз, так что в выборке X_{N_2} побывают все объекты исходной обучающей выборки.

6. Точность классифицирующего правила оценивается по формуле (13.26).

Неудобство применения последнего способа оценки точности классифицирующего правила состоит в необходимости больших затрат вычислительного времени. Поэтому при выборе первого или второго из рассмотренных способов следует исходить из конкретных условий задачи.

13.7. Планирование эксперимента в задаче классификации

Так же как и в задачах построения моделей объектов, планируемый эксперимент является важным средством повышения надежности классифицирующего правила. Здесь мы рассмотрим один из способов планирования эксперимента при уточнении параметров разделяющей поверхности [22].

Пусть все переменные x_i, $i = 1, 2, \ldots, n$, являются управляемыми и область планирования Ω_x задана неравенствами

$$x_{i-} \leqslant x_i \leqslant x_{i+}, \quad i = 1, 2, \ldots, n. \tag{13.27}$$

Предположим, что в пространстве x существует единственная разделяющая поверхность и ее можно описать уравнением

$$F(x, \bar{a}) = \bar{a}_0 + \bar{a}_1 x_1 + \bar{a}_2 x_2 + \ldots + \bar{a}_n x_n = 0, \tag{13.28}$$

т. е. она представляет собой гиперплоскость в пространстве x. Это предположение необходимо для упрощения последующих выводов, которые затем без большого труда могут быть распространены на поверхности более сложного вида.

Допустим далее, что гиперплоскость (13.28) не проходит через начало координат, т. е. $\bar{a}_0 \neq 0$. Выполнения этого условия легко добиться, например, переносом начала координат в одну из точек обучающей выборки. Поделив все параметры $\bar{a}_i$, $i = 0, 1, \ldots, n$, на абсолютную величину $\bar{a}_0$, получим вектор параметров $\bar{a}$, в котором компонент $\bar{a}_0$ принимает лишь значения $+1$ или -1. Соответствующим заданием классифицирующей функции (13.17) можно добиться [исходя из условия (13.21)], чтобы было $\bar{a}_0 = +1$. Наконец, предположим, что заранее могут быть заданы сколь угодно широкие, но конечные пределы, ограничивающие область возможных значений $\bar{a}_i$:

$$a_{i-} \leqslant \bar{a}_i \leqslant a_{i+}, \quad i = 1, 2, \ldots, n. \tag{13.29}$$

Пусть проведено N экспериментов и составлена обучающая выборка X_N. В общем случае для этой выборки существует бесконечное множество поверхностей вида (13.28), которые будут безошибочно разделять объекты из X_N в соответствии с их классификациями $\tilde{Y}_N$. Эти поверхности, как и ранее, будем называть аппроксимациями разделяющей поверхности (АРП). Множеству АРП в пространстве параметров a соответствует некоторая область $\omega(X_N, \tilde{Y}_N)$, называемая областью решений обучающей выборки X_N. Размеры и конфигурация этой области зависят от числа и расположения точек выборки X_N в пространстве x и их классификаций $\tilde{Y}_N$.

Любой вектор из $\omega(X_N, \tilde{Y}_N)$ может в конечном счете оказаться вектором параметров истинной разделяющей поверхности. В практических задачах обычно отсутствует информация о предпочтительности тех или иных векторов из $\omega(X_N, \tilde{Y}_N)$. Поэтому можно считать, что все векторы $a \in \omega(X_N, \tilde{Y}_N)$ с равными вероятностями могут быть векторами параметров разделяющей поверхности.

Тогда точность решения, получаемого по выборке X_N, можно характеризовать величиной объема $V(X_N, \tilde{Y}_N)$ области $\omega(X_N, \tilde{Y}_N)$.

Пусть к выборке X_N добавлена некоторая точка x^{N+1}. До проведения эксперимента в этой точке ее классификация y^{N+1} неизвестна. Можно говорить лишь об априорных вероятностях принадлежности этой точки к тому или другому классу. Поэтому в качестве характеристики целесообразности проведения очередного эксперимента в точке x^{N+1} естественно выбрать величину математического ожидания объема области решения после реализации эксперимента в этой точке. Тогда критерий планирования можно сформулировать следующим образом:

$$E_{y^{N+1}}\{V(X_N, x^{N+1}; \tilde{Y}_N, y^{N+1})\} = \min_{x \in \Omega_x} E_y\{V(X_N, x; \tilde{Y}_N, y)\}. \quad (13.30)$$

Последовательный выбор точек постановки эксперимента в соответствии с критерием (13.30) составляет содержание задачи последовательного планирования эксперимента в детерминистской задаче классификации.

Рассмотрим один метод планирования эксперимента, в основу которого положен критерий (13.30).

Пусть эксперимент предполагается провести в некоторой точке x. Гиперплоскость

$$1 + \sum_{i=1}^{n} a_i x_i = 0 \quad (13.31)$$

в пространстве параметров a делит область решений $\omega(X_N, \tilde{Y}_N)$ на две части ω_1 и ω_2 с объемами V_1 и V_2 (если гиперплоскость не пересекает эту область, то один из этих объемов равен нулю). При этом

$$V_1 = V(X_N, \tilde{Y}_N) - V_2. \quad (13.32)$$

После реализации эксперимента в x и определения значения $\tilde{y}$ новая область решений будет иметь объем V_1, если $\bar{a} \in \omega_1$, и V_2, если $\bar{a} \in \omega_2$. При отсутствии априорных сведений о предпочтительности того или иного значения a из области решений $\omega(X_N, \tilde{Y}_N)$ разумно принять, что

$$P(\bar{a} \in \omega_1) = \frac{V_1}{V(X_N, \tilde{Y}_N)}. \quad (13.33)$$

Математическое ожидание объема области решений при выборе точки x для постановки очередного эксперимента равно

$$E_y\{V(X_N, x; \tilde{Y}_N, y)\} = P(\bar{a} \in \omega_1) V_1 + P(\bar{a} \in \omega_2) V_2. \quad (13.34)$$

С учетом (13.33) из (13.34) получаем

$$E_y\{V(X_N, x; \tilde{Y}_N, y)\} = \frac{1}{V(X_N, \tilde{Y}_N)}\{V_1^2 + [V(X_N, \tilde{Y}_N) - V_1]^2\}. \quad (13.35)$$

Нетрудно убедиться, что минимум величины (13.35) достигается в тех точках x^{N+1}, для которых

$$V_1 = V_2 = \frac{1}{2}V(X_N, \tilde{Y}_N). \quad (13.36)$$

Итак, очередной эксперимент следует проводить в такой точке x^{N+1}, чтобы гиперплоскость (13.31) разделила область решений $\omega(X_N, \tilde{Y}_N)$ на две части с равными объемами. Вообще говоря, может существовать бесконечное множество точек, удовлетворяющих этому условию. Для получения единственного решения можно использовать дополнительные условия (например, потребовать, чтобы объем сферы, в которую вписывается область решения, был минимален) или выбирать точку x^{N+1} случайным образом из числа тех, для которых выполняется (13.36).

После выбора точки x^{N+1} и проведения в ней эксперимента эта точка присоединяется к обучающей выборке, а описанная выше процедура повторяется с целью нахождения точки x^{N+2} и т. д.

На каждом шаге такой процедуры объем области решений уменьшается в 2 раза, и, следовательно, после реализации m экспериментов объем области решений будет в 2^m раз меньше объема исходной области решений.

Реализация процедуры последовательного планирования эксперимента в соответствии с критерием (13.30) требует многократного расчета объемов многомерных областей решения и потому связана с большими затратами вычислительного времени. На практике более удобным оказывается упрощенный алгоритм последовательного планирования, состоящий из следующих операций:

1. По обучающей выборке X_N находится АРП с параметрами a^*, удовлетворяющая условию

$$Q_3(a^*) = \min_a Q_3(a), \quad (13.37)$$

где

$$Q_3(a) = \begin{cases} Q_3(a), & \text{если } Q_3(a) > 0, \\ -\min\limits_{x^j \in X_N} [|F(x^j, a)|/\|a\|], & \text{если } Q_3(a) = 0, \end{cases}$$

$$\|a\| = \left[\sum_{i=1}^{n} a_i^2\right]^{1/2},$$

$Q_3(a)$ рассчитывается по формуле (13.25).

В пространстве x гиперплоскость с параметрами a^* является разделяющей для точек x^j из обучающей выборки X_N и наиболее удалена от ближайших к ней точек этой выборки.

2. Точка x^{N+1} для проведения следующего эксперимента выбирается случайным образом из числа тех, для которых выполняются условия

$$\begin{cases} x^{N+1} \in \Omega_x, \\ 1 + \sum_{i=1}^{n} a_i^* x_i^{N+1} = 0. \end{cases}$$

Таким образом, точка x^{N+1} должна находиться на гиперплоскости с параметрами a^* и принадлежать области планирования.

Упрощенный алгоритм планирования легко реализуем и не требует громоздких вычислений. Вместе с тем этот алгоритм оказывается весьма эффективным в смысле повышения точности оценок параметров разделяющей поверхности. Моделирование этого алгоритма показало, что средний проигрыш в точности по сравнению с алгоритмом, основанным на критерии (13.30), составляет около 10%.

ЛИТЕРАТУРА

1. Брановский В. Л., Лунц А. Л., Наткович Ю. С., Об одной многоэтапной процедуре прогнозирования качества технологических процессов, см. сб. „Проблемы планирования эксперимента", изд-во „Наука", М., 1969.
2. Морозов В. С., Морозова В. А., Маркова Е. В., Герасимова М. Ф., Векслер М. А., Князева М. Г., Филатова В. В., Хейфец С. Е., Распознавание образов при классификации качества продукта по входным параметрам, см. сб. „Проблемы планирования эксперимента", изд-во „Наука", М., 1969.
3. Лецкий Э. К., Ермуратский П. В., Дьякова Н. С., Федин В. А., Решение некоторых задач управления сложными производственными процессами с применением табличного метода, см. сб. „Приборостроение, средства автоматизации и системы управления", изд-во „Наука", М., 1967.
4. Вольтер Б. В., Дорофеюк А. А., Касавин А. Л., Торговицкий И. Ш., Анализ химико-технологических процессов методами автоматической классификации данных, *Теоретич. основы хим. технол.*, № 4 (1970).
5. Ивов Г. С., Выбор эффективной системы зависимых признаков, см. сб. „Вычислительные системы", вып. 19, Ин-т математики СОАН СССР, Новосибирск, 1965.
6. Барабаш Ю. Л., Варский Б. В., Зиновьев В. Т., Кириченко В. С., Сапегин В. Ф., Вопросы статистической теории распознавания, изд-во „Советское радио", М., 1967.
7. Marill T., Green D. M., On the Effectiveness of Receptors in Recognition Systems, *IRE Trans. Inform. Theory*, IT-9, № 1, 11—17 (1963).
8. Kudo A., Mahalanobis' Measure as a Criterion for the Selection of Variables, *Mem. Fac. Kyushu Univ.*, A17, № 1, 63—75 (1963).
9. Anderson T. W., An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958.
10. Статистические методы классификации, сб. статей под. ред. Ю. Н., Благовещенского, вып. 1, Изд-во МГУ, 1969.

11. Fix E., Hodges J. L., Nonparametric Discrimination II. Small Sample Performance, Techn. Rep. 11, Randolph Field, Texas, 1952.
12. Sebestien G. S., Decision Making Process in Pattern Recognition, Macmillan, New York, 1962; русский перевод: Себастиан Г. С., Принятие решений при распознавании образов, изд-во „Техника", Киев, 1965.
13. Цыпкин Я. З., Адаптация и обучение в автоматических системах, изд-во „Наука", М., 1968.
14. Айзерман М. А., Браверман Э. М., Розоноэр Л. И., Метод потенциальных функций в теории обучения машин, изд-во „Наука", М., 1970.
15. Вапник В. Н., Червоненкис А. Я., Теория распознавания образов, изд-во „Наука", М., 1974.
16. Nilsson N. J., Learning Machines, McGraw-Hill, New York, 1965; русский перевод: Нильсон Н., Обучающиеся машины, изд-во „Мир", М., 1967.
17. Box M. J., A new method of constrained optimization and a comparison with other methods, *Computer J.*, 8, № 1 (1965).
18. Nelder J. A., Mead R. M., A simplex method for function minimization, *Computer J.*, 8, № 1 (1965).
19. Lachenbruch P. A., Mickey M. R., Estimation of error rates in discriminant analysis, *Technometrics*, 10, № 1 (1968).
20. Brandes H., Das Verhalten des Rührreaktors im *C*, *T*-Schaubild und die Ermittlung der stabilisierenden Kühlbedingungen, см. Grundlagen der chemischen Prozessregelungen, Oldenbourg, München/Wien, 1964.
21. Цыпкин Я. З., Основы теории обучающихся систем, изд-во „Наука", М., 1970.
22. Лецкий Э. К., Фомин Г. А., Последовательное планирование эксперимента в задачах распознавания образов, *Изв. АН СССР*, Техническая кибернетика, № 1 (1972).

14.

ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ ОПРЕДЕЛЕНИЯ ДИНАМИЧЕСКИХ ХАРАКТЕРИСТИК ОБЪЕКТОВ УПРАВЛЕНИЯ

Промышленными объектами часто нельзя эффективно управлять, не зная их характеристики. При построении систем автоматического управления необходимо тщательно исследовать динамические свойства объекта и использовать результаты исследования как на этапе формулирования задачи, так и при разработке алгоритма управления. Динамические характеристики следует принимать во внимание при априорной оценке ожидаемого экономического эффекта от внедрения системы управления. Эффективность действия автоматических регуляторов, равно как и выбор его структуры и параметров немыслимы без анализа динамических свойств объекта регулирования.

При построении математической модели статики динамические свойства объекта необходимо учитывать при планировании эксперимента, а также при его проведении и обработке результатов. Динамические характеристики должны быть известны и при экспериментировании на объекте с целью его оптимизации или определения наилучшего управления. Математическая модель динамики, или, короче, динамическая модель, описывающая зависимость выходных переменных объекта от входных переменных как функцию времени, позволяет в этих случаях повысить точность решения задачи, а также путем специальных мер ускорить исследование.

Рассматриваемые в данном разделе экспериментальные методы построения динамических моделей обладают по сравнению с аналитическими методами тем преимуществом, что не требуют глубокого анализа физико-химических процессов, протекающих в объекте. Но в этом же заключается и их недостаток, поскольку перенос результатов эксперимента на другой объект оказывается невозможным или трудно выполнимым.

14.1. Динамические характеристики объектов управления

Известно большое число способов, с помощью которых можно описать динамические свойства объекта управления. Как правило, объекты оказываются нелинейными, т. е. описываются нелинейными дифференциальными или интегральными уравнениями.

Однако в окрестности рабочих режимов во многих случаях возможна линеаризация. В дальнейшем будут рассматриваться именно такие объекты. Методы идентификации динамических характеристик достаточно хорошо разработаны только для линейных объектов.

Другой существенной особенностью рассматриваемых здесь объектов является сосредоточенность их параметров. Описание многих объектов, явно имеющих распределенную структуру, с помощью обыкновенных дифференциальных уравнений, т. е. представление их в виде структур с сосредоточенными параметрами, оправдано многолетней практикой решения задач автоматизации. Такое упрощение устраняет многие затруднения, связанные с исследованием дифференциальных уравнений в частных производных. Поэтому дальнейшее изложение будет ограничено рассмотрением только линейных объектов с сосредоточенными параметрами.

Рассмотрим одномерный объект, т. е. объект, имеющий один вход и один выход. Обобщение на многомерный случай не вызывает принципиальных затруднений.

Обозначим входную переменную объекта $x(t)$, выходную $y(t)$. Эти переменные связаны некоторым дифференциальным уравнением. Точно определить это уравнение по экспериментальным данным оказывается возможным лишь в простейших случаях. Кроме того, использование дифференциальных уравнений не всегда целесообразно, так как существуют иные характеристики динамических объектов, более удобные для анализа и синтеза систем управления.

Удобной формой математического описания звеньев систем управления является передаточная функция $W(p)$. Ее определяют как отношение преобразований Лапласа выходного сигнала к входному при нулевых начальных условиях [2]:

$$W(p) = \frac{Y(p)}{X(p)} \tag{14.1}$$

при $x(0) = \dot{x}(0) = \ldots = x^{n-1}(0) = y(0) = \dot{y}(0) = \ldots = y^{(m-1)}(0) = 0$,

$$\dot{x} = \frac{dx}{dt}, \quad \dot{y} = \frac{dy}{dt}, \quad x^{(k)} = \frac{d^k x}{dt^k}, \quad y^{(k)} = \frac{d^k y}{dt^k},$$

где $X(p)$ и $Y(p)$ — преобразования Лапласа функций времени $x(t)$ и $y(t)$. Для рассматриваемого здесь класса объектов предполагается, что передаточные функции могут быть представлены в следующем виде:

$$W(p) = \frac{K(p)}{D(p)} = \frac{b_m p^m + b_{m-1} p^{m-1} + \ldots + b_0}{a_n p^n + b_{n-1} p^{n-1} + \ldots + a_0}, \tag{14.2}$$

где всегда $m \leqslant n$.

Непосредственное отыскание передаточной функции в виде отношения преобразований Лапласа выходных и входных сигналов объекта оказывается практически неосуществимым. Во-первых, найти преобразование Лапласа по наблюдениям некоторой функции во времени чрезвычайно сложно. Кроме того, по определению требуется, чтобы начальные условия для наблюдаемых процессов были нулевыми. Некоторые весьма сложные для реализации способы нахождения коэффициентов передаточной функции (14.2) при заданном порядке полиномов в числителе и знаменателе приведены в [1].

На практике нахождение передаточных функций можно выполнить с помощью других динамических характеристик: импульсной переходной функции (весовой функции), частотных характеристик (амплитудно-фазовых характеристик) и переходных функций (кривых разгона).

Весовой функцией называют оригинал (обратное преобразование Лапласа) передаточной функции $W(p)$:

$$w(t) = \frac{1}{2\pi j}\int_{\sigma - j\infty}^{\sigma + j\infty} W(p)\, e^{pt}\, dp \qquad (14.3)$$

или

$$W(p) = \int_0^{\infty} w(t)\, e^{-pt}\, dt,$$

где $j = \sqrt{-1}$, $\sigma = \text{const}$.

Один из возможных способов нахождения функции $w(t)$ состоит в том, что на вход объекта при нулевых начальных условиях подается единичный импульс $\delta(t)$ (δ-функция). При этом на выходе получается сигнал, совпадающий с $w(t)$. Единичный импульс определяется с помощью предельного перехода

$$\delta(t_1) = \lim_{\tau \to 0} \eta(t),$$

где

$$\eta(t) = \begin{cases} 0, & \text{если } t < t_1, \\ 1/\tau, & \text{если } t_1 \leqslant t \leqslant t_1 + \tau, \\ 0, & \text{если } t > t_1 + \tau. \end{cases}$$

К сожалению, реализовать этот способ практически не удается, так как амплитуда импульса должна быть бесконечно большой.

Весовая функция позволяет непосредственно связать выходную и входную переменные объекта с помощью интеграла Дюамеля:

$$y(t) = \int_0^{\infty} w(\tau)\, x(t - \tau)\, d\tau. \qquad (14.4)$$

Это выражение играет важную роль при математическом описании динамических систем. Весовая функция $h(t)$ связана простым соотношением с переходной функцией $w(t)$:

$$h(t) = \int_0^t w(\tau)\, d\tau. \qquad (14.5)$$

Переходная функция описывает изменение сигнала на выходе объекта при подаче на его вход единичного скачка $1(t)$:

$$1(t) = \begin{cases} 0, & \text{если } t \leqslant 0, \\ 1, & \text{если } t > 0. \end{cases}$$

Существует целый ряд методов, позволяющих перейти от переходной функции к передаточной функции и частотной характеристике.

При описании динамических свойств объектов часто используется комплексный коэффициент усиления $W(j\omega)$, который определяется как отношение комплексной амплитуды выходного сигнала к комплексной амплитуде входного сигнала при подаче на вход синусоидального воздействия. Выражение для $W(j\omega)$ можно получить формальной заменой переменной p на $j\omega$ в передаточной функции объекта. Комплексный коэффициент усиления полностью характеризует динамические свойства объекта при изменении частоты от нуля до бесконечности.

Геометрическое место конца вектора $W(j\omega)$ при изменении частоты называется амплитудно-фазовой характеристикой, комплексной частотной характеристикой или частотным годографом комплексного коэффициента усиления.

Экспериментальное измерение $W(j\omega)$ можно выполнить путем подачи на вход объекта сигнала синусоидальной формы, имеющего определенную амплитуду и частоту. При этом на выходе измеряются амплитуда и фаза сигнала. Если, например,

$$x = X_m \sin \omega t,$$
$$y = Y_m \sin (\omega t + \varphi),$$

то комплексная амплитуда входного сигнала $\dot{x}_m = X_m$, а комплексная амплитуда выходного сигнала

$$\dot{y}_m = Y_m \exp (j\varphi).$$

Тогда

$$W(j\omega) = \frac{\dot{y}_m}{\dot{x}_m} = A(\omega) \exp (j\varphi).$$

На практике применяют различные модификации частотных характеристик: амплитудно-частотные, фазочастотные, логарифмические амплитудно-частотные и т. д. Все они определяются

с помощью комплексного коэффициента усиления, который в дальнейшем будет называться частотной характеристикой. Эта характеристика позволяет достаточно просто найти передаточную функцию объекта.

14.2. Выбор способа идентификации объекта управления

При постановке задачи идентификации динамического объекта необходимо учитывать целый ряд аспектов:

1. Какая форма и какая интенсивность допустимы для изучающего воздействия? Допустимы ли вообще специальные изучающие воздействия?
2. Какова степень зашумленности выходного сигнала объекта?
3. Какие технические средства для идентификации [1] имеются в распоряжении исследователей?
4. Каковы допустимые продолжительность и стоимость исследований?
5. Насколько высокой должна быть точность идентификации?
6. В какой мере получаемые результаты являются репрезентативными?

При выборе метода идентификации во многих случаях приходится принимать во внимание и иные аспекты.

Если в системе отсутствуют случайные возмущения, то наиболее эффективными будут детерминированные методы идентификации, связанные с подачей на вход объекта детерминированных сигналов. В тех случаях, когда допускается применение скачкообразных входных сигналов, наблюдаемые переходные функции позволяют исчерпывающе описать динамические свойства объекта. При проведении подобных исследований для записи выходного процесса применяются типовые регистрирующие приборы. Эти методы хорошо описаны в литературе (см., например, [3]).

При экспериментальном построении частотной характеристики проводятся многократные эксперименты на разных частотах. Из-за этого продолжительность исследований здесь оказывается намного большей, чем при получении переходных функций. Кроме того, при практической реализации этих методов приходится преодолевать большое число технических трудностей. Это связано с тем обстоятельством, что для большинства промышленных объектов часто́ты, на которых необходимо производить эксперименты, являются очень низкими: они составляют около 1/1000 Гц. Аппаратура для генерации и измерения таких сигналов оказывается весьма сложной.

[1] Идентификацией называют построение математической модели (статической или динамической) объекта на основе экспериментальных данных. Это понятие в настоящее время преимущественно связывают с задачей построения динамических моделей.— *Прим. ред.*

Класс объектов, допускающих интенсивные воздействия, является весьма ограниченным. В то же время амплитуда сигнала должна быть достаточно большой, чтобы исключить методические погрешности измерений.

Для большинства промышленных объектов, в том числе и процессов химической технологии, характерно наличие интенсивных случайных возмущений. Они вызываются многими причинами: случайными изменениями параметров объекта и управляющих воздействий, случайными погрешностями приборов, помехами в каналах передачи данных и т. д. Случайный характер возмущений обусловливает применение статистических методов обработки сигналов. В дальнейшем будут рассматриваться исключительно статистические методы идентификации динамических характеристик.

При определении переходной функции приходится усреднять большое число измеренных кривых разгона. На практике этот способ реализовать трудно, так как все кривые должны сниматься при одинаковых условиях. Статистическая обработка сигналов при расчете частотных характеристик выполняется сравнительно просто. Это объясняется тем, что в стационарном режиме при синусоидальном входном сигнале на выходе объекта получается также синусоидальный сигнал с той же частотой. Необходимо лишь определить параметры сигнала, т. е. его амплитуду и фазу. Если на сигнал наложены случайные возмущения, задача состоит в оценке параметров сигнала известного вида. В этом случае оптимальным (в смысле минимума суммы квадратов ошибок) способом обработки является когерентное интегрирование. Этот способ заключается в интегрировании произведения исследуемого сигнала и нормированного сигнала известного вида. Для иллюстрации рассмотрим пример нахождения параметров частотной характеристики.

Предположим, что входной сигнал объекта в некоторой точке частотной характеристики, соответствующей частоте ω, может быть представлен в виде

$$x(t) = a \sin \omega t.$$

В объекте без возмущений на выходе будет получаться также синусоидальный сигнал

$$y_0(t) = A \sin(\omega t + \varphi).$$

Этот сигнал аддитивно смешивается с возмущающей функцией $f(t)$. Следовательно, измерительный прибор покажет величину

$$y(t) = y_0(t) + f(t).$$

Вычислим теперь для одного и того же интервала времени $0 \leqslant t \leqslant T$ два интеграла:

$$I_s = \frac{1}{T}\int_0^T y(t)\sin\omega t\,dt,$$
$$I_c = \frac{1}{T}\int_0^T y(t)\cos\omega t\,dt. \qquad (14.6)$$

Время проведения эксперимента кратно периоду частоты ω ($T = 2\pi M/\omega$), где M — достаточно большое целое число. При этом условии из (14.6) получаем

$$I_s = \frac{1}{2}A\cos\varphi + \delta_s(M),$$
$$I_c = \frac{1}{2}A\sin\varphi + \delta_c(M), \qquad (14.7)$$

где δ_s и δ_c — случайные величины. Их средние значения равны нулю, а дисперсия убывает по мере увеличения продолжительности эксперимента. Если выбрать величину M так, чтобы значениями δ_s и δ_c можно было пренебречь, то из уравнений (14.7) следует

$$\varphi = \operatorname{arctg}(I_c/I_s),$$
$$A = 2\sqrt{I_c^2 + I_s^2}.$$

Затраты на определение частотной характеристики при действии возмущений оказываются очень большими.

В последнее время большое распространение при идентификации динамических характеристик нашли корреляционные методы, предназначенные для определения весовых функций. Эти методы будут рассмотрены в следующем разделе.

14.3. Общая характеристика корреляционных методов идентификации

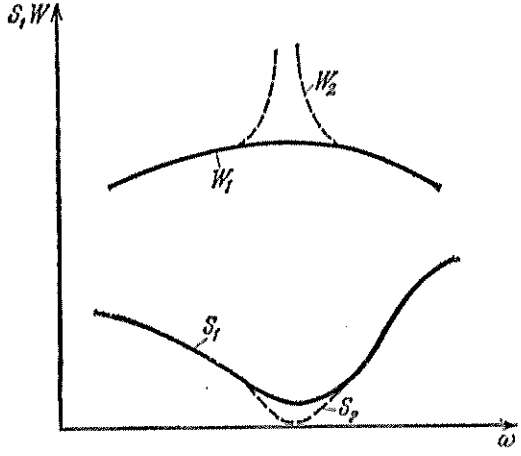
Известно большое число методов идентификации [4, 5], не использующих оценки статистических характеристик случайных процессов на входе и выходе объекта (например, корреляционные функции и спектральные плотности).

Весовая функция $w(t)$ стационарного линейного объекта при условии некоррелированности внутренних возмущений с входными переменными объекта определяется интегральным уравнением Винера — Хопфа

$$R_{xy}(\tau) = \int_0^\infty R_{xx}(t-\tau)\,w(t)\,dt, \qquad (14.8)$$

где $R_{xx}(t-\tau)$ — автокорреляционная функция входного сигнала $x(t)$; $R_{xy}(\tau)$ — взаимно корреляционная функция входного и выходного сигналов.

Прежде чем использовать уравнение (14.8) для идентификации, необходимо оценить влияние погрешности в определении корреляционных функций $R_{xy}(\tau)$ и $R_{xx}(t-\tau)$ на погрешность в расчете весовой функции $w(t)$. При этом оказывается, что



Фиг. 14.1. Влияние ошибок на результат идентификации.

отыскание функции $w(t)$ с помощью уравнения (14.8) является некорректно поставленной задачей. Это означает, что малые погрешности в корреляционных функциях $R_{xy}(\tau)$ и $R_{xx}(t-\tau)$ приводят к сколь угодно большим ошибкам в оценке $w(t)$.

Рассмотрим эту проблему несколько подробнее. Для этого исследуем решение уравнения (14.8), полученное с помощью частотного (спектрального) метода. Из уравнения (14.8) с помощью преобразования Фурье можно получить следующее уравнение [4]:

$$S_{xy}(j\omega) = W(j\omega) S_{xx}(\omega), \qquad (14.9)$$

где $S_{xy}(j\omega)$ и $S_{xx}(\omega)$ — спектральные плотности сигналов на входе и выходе объекта; $W(j\omega)$ — комплексный коэффициент усиления (частотная характеристика объекта). Отсюда следует, что

$$W(j\omega) = \frac{S_{xy}(j\omega)}{S_{xx}(\omega)}.$$

Применение спектральных характеристик для решения уравнения (14.8) весьма привлекательно, так как в результате сразу

получается частотная характеристика. Кроме того, спектральные характеристики могут быть найдены экспериментальным путем.

При конечной длине реализаций $x(t)$ и $y(t)$ для некоторых значений ω спектральная плотность $S_{xx}(\omega)$ близка к нулю. В этом случае малые изменения в величине $S_{xx}(\omega)$, вызванные помехами или вычислительными ошибками, могут привести к сколь угодно большим изменениям в функции $W(j\omega)$ (фиг. 14.1). На этой фигуре пунктирной линией показано, как небольшие изменения $S_{xx}(\omega)$ приводят к большим изменениям $W(\omega)$.

Еще раз отметим, что среднеквадратичное отклонение весовой функции, найденной из уравнения (14.8), от действительной весовой функции может быть малым, но ценность результата тем не менее будет невелика, так как эта функция не соответствует физическому смыслу протекающих в объекте процессов. Физический смысл имеют только гладкие решения уравнения Винера—Хопфа.

Для нахождения гладких решений уравнения (14.8) используются методы регуляризации. Один из них заключается в таком определении весовой функции, чтобы достигался минимум функционала

$$Q = E\left\{\left[y(t) - \int_0^\infty w(t-\tau)\,x(\tau)\,d\tau\right]^2\right\} + \alpha\Omega[w(t)], \quad (14.10)$$

где α — положительное число; $\Omega[\ldots]$ — сглаживающий функционал, имеющий вид

$$\Omega[w(t)] = \int_0^T [k(\tau)\,\dot{w}^2(\tau) + p(\tau)\,w^2(\tau)]\,d\tau,$$
$$k(\tau) > 0, \quad p(\tau) > 0, \quad \dot{w}(\tau) = \frac{d}{d\tau}w(\tau). \quad (14.11)$$

Условие минимума функционала Q можно записать в такой форме [6, 7]:

$$R_{xy}(\tau) = \alpha\left\{\frac{d}{dt}\left[k\frac{\partial w}{\partial \tau}\right] + pw\right\} + \int_0^\infty R_{xx}(\tau - t)\,w(t)\,dt, \quad (14.12)$$
$$\dot{w}(0) = \dot{w}(\infty) = 0.$$

Уравнение (14.12) отличается от уравнений Винера—Хопфа дополнительным членом в правой части, обеспечивающим существование, единственность и дифференцируемость решения.

Рассмотрим еще один способ отыскания весовой функции. Пусть искомая функция может быть представлена в виде ряда

$$w(t) = \sum_{i=0}^{N} c_i\varphi_i(t), \quad (14.13)$$

где $\varphi_l(t)$ — некоторые известные функции (базисная система функций); c_l — коэффициенты разложения $w(t)$ по функциям $\varphi_l(t)$. Подставим (14.13) в первое слагаемое выражения (14.10):

$$\varepsilon^2 = E\left\{\left[y(t) - \int_0^\infty x(t-\tau)\sum_{l=0}^{N} c_l\varphi_l(\tau)\,d\tau\right]^2\right\}. \qquad (14.14)$$

При этом ε^2 есть функция вектора коэффициентов c, компонентами которого являются $c_0, c_1, \ldots, c_N$. Введем обозначения

$$\begin{aligned} c &= \{c_0, c_1, \ldots, c_N\}, \\ z_l(t) &= \int_0^\infty \varphi_l(\tau)\,x(t-\tau)\,d\tau. \end{aligned} \qquad (14.15)$$

Величины $z_l(t)$ называют выходными реакциями фильтров с весовыми функциями $\varphi_l(t)$ на входной сигнал $x(t)$. С учетом обозначений (14.15) запишем (14.14) в таком виде:

$$\varepsilon^2 = E\left\{\left[y(t) - \sum_{l=0}^{N} c_l z_l(t)\right]^2\right\}. \qquad (14.16)$$

Вектор c, минимизирующий величину ε^2, находится из $(N+1)$ уравнений

$$\frac{\partial \varepsilon^2}{\partial c_j} = 0, \quad j = 0, 1, \ldots, N. \qquad (14.17)$$

Дифференцируя (14.16) по c_j и подставляя результат в (14.17), получим систему уравнений для определения вектора $\hat{c}$:

$$E\{y(t)\,z_j(t)\} = \sum_{l=0}^{N} \hat{c}_l\,E\{z_j(t)\,z_l(t)\}, \quad j = 0, \ldots, N. \qquad (14.18)$$

Здесь $y(t)$ и $z_j(t)$ — случайные функции. В связи с этим числа $E\{y(t)\,z_j(t)\}$ и $E\{z_j(t)\,z_l(t)\}$ представляют собой начальные значения взаимно корреляционных функций сигналов $y(t)$ и $z_j(t)$ или $z_l(t)$ и $z_j(t)$ соответственно. Для упрощения записи обозначим их так:

$$\begin{aligned} R_{yj} &= E\{y(t)\,z_j(t)\}, \\ R_{lj} &= E\{z_l(t)\,z_j(t)\}. \end{aligned} \qquad (14.19)$$

Заметим, что из (14.19) следует равенство $R_{lj} = R_{jl}$. С учетом (14.19) перепишем теперь систему уравнений (14.18) в такой форме:

$$R_{yj} = \sum_{l=0}^{N} \hat{c}_l R_{lj}, \quad j = 0, \ldots, N. \qquad (14.20)$$

Левые части уравнений (14.20) рассчитываются по экспериментальным данным, и поэтому они в наибольшей степени подвержены влиянию помех на выходе объекта.

Учитывая, что R_{lj} представляет собой начальное значение взаимно корреляционной функции сигналов $z_l(t)$ и $z_j(t)$, можно записать на основе известной связи между корреляционной функцией и спектральной плотностью

$$R_{lj} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_l(j\omega)\, \Phi_j(-j\omega)\, S_{xx}(\omega)\, d\omega, \qquad (14.21)$$

где $\Phi_l(j\omega)$ — частотная характеристика фильтра с весовой функцией $\varphi_l(t)$; $S_{xx}(\omega)$ — спектральная плотность входного сигнала $x(t)$. Достаточное условие существования решения системы уравнений (14.20) имеет вид

$$R_{lj} = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases}$$

В соответствии с определением (14.19) матрица R_{lj} будет единичной, если выходные сигналы фильтров базисной системы функций ортонормированны. Из (14.21) следует, что ортонормированность может быть обеспечена лишь совместным выбором $\Phi_l(j\omega)$ и $S_{xx}(\omega)$ или соответственно $\varphi_l(t)$ и $x(t)$.

При выборе базисной системы функций $\varphi_l(t)$ нужно учитывать, что с вычислительной точки зрения удобно пользоваться системой ортогональных функций. Во временно́й области из ортогональности функций $\varphi_l(t)$ следует

$$\int_0^{\infty} \varphi_l(t)\, \varphi_j(t)\, dt = \begin{cases} A, & \text{если } i = j, \\ 0, & \text{если } i \neq j, \end{cases}$$

а в частотной области

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_l(j\omega) \Phi_j(-j\omega)\, d\omega = \begin{cases} A, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases} \qquad (14.22)$$

Сравнивая выражения (14.21) и (14.22), можно заметить, что при ортогональной базисной системе функций имеет место также ортогональность выходных сигналов фильтров, если спектральная плотность выходного сигнала $x(t)$ постоянна, т. е. если

$$S_{xx}(\omega) = S_0. \qquad (14.23)$$

Это означает, что входным сигналом объекта должен быть белый шум; тогда матрица R_{lj} оказывается единичной при ортогональной базисной системе функций. Однако белый шум в принципе

является лишь математической абстракцией. Вопросы реализации сигналов, близких к белому шуму, рассмотрены в работе [8].

Необходимо отметить, что корреляционная функция белого шума представляет собой δ-функцию. Если эту δ-функцию подставить в качестве $R_{xx}(\tau)$ в уравнение Винера—Хопфа, получается тривиальное решение

$$R_{xy}(\tau) = \int_0^\infty w(\tau)\,\delta(t-\tau)\,dt = w(\tau).$$

Благодаря этому метод идентификации упрощается, так как из экспериментальных данных необходимо определять только $R_{xy}(\tau)$.

Известно немало систем ортогональных функций. Поэтому при выборе конкретной базисной системы принимаются во внимание дополнительные физические соображения. Как уже отмечалось выше, на практике большинство объектов может быть описано передаточной функцией в виде правильной дроби

$$W(p) = \frac{K(p)}{D(p)},$$

где

$$D(p) = \prod_{k=1}^{r} (p + p_k)^{n_k},$$

$$\sum_{k=1}^{r} n_k = n.$$

Для весовой функции таких объектов можно записать следующее выражение:

$$w(t) = \sum_{k=1}^{r} \sum_{j=1}^{n_k} \frac{K_{kj}}{(n_k-j)!}\, t^{n_k-j} \exp(p_k t), \qquad (14.24)$$

где

$$K_{kj} = \frac{1}{(j-1)!}\left[\frac{d^{j-1}}{dD^{j-1}} \frac{(p+p_k)^{n_k} K(p)}{D(p)}\right]_{p=-p_k}.$$

Сравнивая (14.24) и (14.13), можно заметить, что система функций должна при этом иметь такой вид:

$$\{\varphi_l(t)\} = \{t^{n_l} \exp(p_l t)\}. \qquad (14.25)$$

При выборе системы функций в таком виде величины n_l и p_l должны быть заранее известны, что при решении задач идентификации обычно не выполняется. Если, однако, учесть, что характер изменения функции $w(t)$ при больших значениях t в основном определяется членом $e^{p_1 t}$, где $|p_1| = \min_l |p_l|$, то после предварительной оценки величины p_1 можно использовать сле-

дующую систему функций:

$$\{\varphi_i(t)\} = \{t^n e^{-\alpha t}\}, \tag{14.26}$$

где $1/\alpha$ — параметр системы функций, по своему смыслу аналогичный постоянной времени.

Таким способом находится система ортонормированных функций, элементы которой определяются выражением (14.26). Рассмотрим систему функций Лагерра, отвечающую этому выражению. Эти функции во временной области и через изображение по Лапласу определяются следующим образом:

$$l_i(t) = \sqrt{2\alpha}\, e^{-\alpha t} \sum_{k=0}^{i} \frac{i!\,(-2\alpha)^k}{(i-k)!\,(k!)^2}\, t^k, \tag{14.27}$$

$$L_i(p) = \frac{\sqrt{2\alpha}}{p+\alpha}\left(\frac{p-\alpha}{p+\alpha}\right)^i. \tag{14.28}$$

В дальнейшем в уравнении (14.13) вместо $\varphi_i(t)$ будем писать $l_i(t)$, если в качестве базисной системы принята система функций Лагерра.

Обобщая некоторые результаты этого раздела, можно сделать следующие выводы:

1. При идентификации динамической системы рекомендуется отыскивать весовую функцию в виде ряда (14.13)

$$w(t) = \sum_{i=0}^{N} c_i l_i(t).$$

Это позволяет корректно и относительно просто решить задачу идентификации.

2. В качестве базисной системы функций $l_i(t)$ целесообразно применять ортонормированную систему функций Лагерра (14.27) и (14.28).

3. Коэффициенты c_i определяются из решения системы уравнений (14.20), матрица которой образована коэффициентами корреляции сигналов на выходах фильтров Лагерра и сигнала на выходе объекта.

4. В качестве входного сигнала для целей идентификации рекомендуется использовать белый шум.

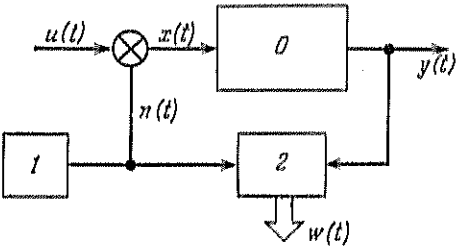
14.4. Некоторые особенности активного корреляционного метода при использовании функций Лагерра

14.4.1. Общие замечания

Одним из важнейших требований к методу идентификации является недопустимость существенного нарушения нормального режима эксплуатации объекта. Кроме того, для многих технологических объектов длительное активное экспериментирование

оказывается невозможным по экономическим причинам. Поэтому идентификация часто проводится в режиме нормальной эксплуатации. Под этим термином понимается проведение пассивного эксперимента (наблюдение за входом и выходом объекта) с последующей его обработкой.

Практика, однако, показывает, что во многих случаях пассивный эксперимент не позволяет получить хороших результатов. Это связано с тем, что, во-первых, в естественном входном сиг-



Фиг. 14.2. Общая схема применения активного корреляционного метода.

нале представлен далеко не весь необходимый спектр частот, во-вторых, при проектировании технологических объектов принимаются специальные меры по стабилизации выходных переменных.

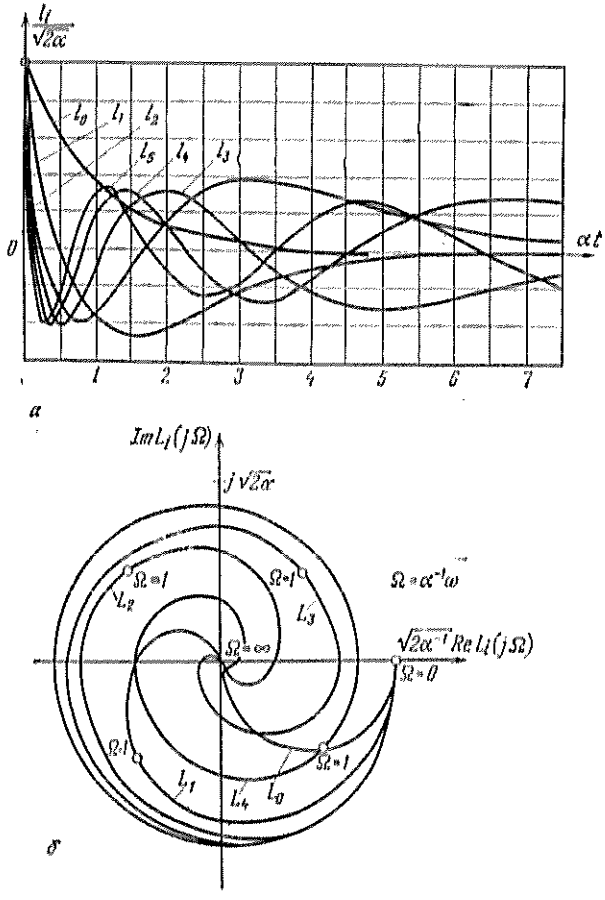
Выходом из такой ситуации является применение специальных изучающих сигналов, подаваемых на объект в дополнение к рабочим сигналам. Спектральные характеристики этих изучающих сигналов заранее задаются исходя из соображений точности. На фиг. 14.2 изображена структурная схема наиболее часто применяемого активного метода идентификации. Здесь $u(t)$ — рабочие управляющие воздействия на объект (0); $n(t)$ — изучающий сигнал, синтезируемый генератором изучающего сигнала (1); $x(t)$ — входной сигнал объекта, представляющий собой сумму двух сигналов: $x(t) = u(t) + n(t)$. Выходной сигнал объекта $y(t)$ и изучающий сигнал передаются в вычислительное устройство (2), где в соответствии с некоторым алгоритмом идентификации определяются динамические характеристики [например, $w(t)$].

Для метода, рассмотренного в разд. 14.3, вычислительное устройство должно рассчитать вектор коэффициентов c разложения весовой функции объекта по системе ортонормированных функций Лагерра. При этом в соответствии с формулой (14.19) рассчитываются значения R_{ij} и R_{yi} с последующим решением системы уравнений (14.20).

Если же сигнал $n(t)$ задан так, что допускается тривиальное решение уравнения Винера—Хопфа, то вычислительное устройство реализует процедуру расчета взаимно корреляционной функции $R_{nn}(\tau)$.

14.4.2. Исследование точности аппроксимации весовой функции объекта с помощью системы функций Лагерра

В разд. 14.3 было предложено в качестве базисной системы функций использовать функции Лагерра, которые во временной области описываются выражением (14.27). На фиг. 14.3, *б* показаны частотные характеристики фильтров с весовыми функциями



Фиг. 14.3. Частотные характеристики фильтров Лагерра.

$l_i(t)$ (в дальнейшем мы будем называть их фильтрами Лагерра). Аналитическое выражение для частотных характеристик получается из (14.28) заменой p на $j\omega$:

$$L_i(j\omega) = \frac{\sqrt{2\alpha}}{\alpha + j\omega}\left(\frac{j\omega - \alpha}{j\omega + \alpha}\right)^i. \qquad (14.29)$$

Характеристики, изображенные на фиг. 14.3, *б*, соответствуют относительным частотам Ω:

$$L_l(j\Omega) = \frac{\sqrt{2/\alpha}}{1+j\Omega}\left(\frac{j\Omega-1}{j\Omega+1}\right)^l. \qquad (14.30)$$

Выражение (14.30) получается из (14.29) заменой

$$\frac{\omega}{\alpha} = \Omega.$$

Система функций Лагерра обладает следующими важными свойствами:

1. Эти функции достаточно просто могут быть рассчитаны с помощью электронных вычислительных машин, так как значения функции с номером i получаются путем несложных операций, исходя из значений функции с номером $(i-1)$.

2. Система функций $l_l(t)$ обладает тем свойством, что любую функцию $w(t)$, удовлетворяющую условию

$$\int_0^\infty w^2(t)\,dt < \infty,$$

можно как угодно точно представить в виде ряда (14.13):

$$w(t) = \sum_{l=0}^{N} c_l l_l(t). \qquad (14.31)$$

Это справедливо по крайней мере при $N \to \infty$, поскольку

$$\lim_{N\to\infty} \int_0^\infty \left[w(t) - \sum_{l=0}^{N} c_l l_l(t)\right]^2 dt = 0. \qquad (14.32)$$

3. Система функций $l_l(t)$ является ортонормированной на отрезке $[0, \infty]$, т. е.

$$\int_0^\infty l_i(t)\,l_j(t)\,dt = \begin{cases} 1, & \text{если } j = i, \\ 0, & \text{если } j \neq i. \end{cases} \qquad (14.33)$$

Второе свойство делает систему функций $l_l(t)$ универсальной и применимой для представления динамических характеристик объектов весьма широкого класса. Третье свойство приводит к существенному упрощению вычисления коэффициентов разложения (14.31), так как при этом и в случае специального выбора входного сигнала добавление новых членов не требует повторного расчета ранее найденных коэффициентов.

С возрастанием числа членов разложения по рассматриваемой системе функций N довольно быстро обеспечивается заданная точность ε (свойство сходимости). Для многих технологических объектов оказывается достаточным $N \leqslant 5$.

После того как выбраны число членов ряда N и параметр функций Лагерра α, остается открытым вопрос о достигаемом качестве аппроксимации. Для суждения о точности аппроксимации часто используют величину относительной квадратичной ошибки

$$\delta_N^2 = \frac{\int_0^\infty [w_0(t) - w(t)]^2 dt}{\int_0^\infty w_0^2(t)\, dt}, \qquad (14.34)$$

где $w_0(t)$ — истинная весовая функция объекта; $w(t) = \sum_{l=0}^{N} c_l l_l(t)$ — ее аппроксимация.

Коэффициенты c_l, минимизирующие (14.34) при использовании ортогональной системы функций $l_l(t)$, являются коэффициентами Фурье функции $w_0(t)$ и могут быть вычислены по формуле

$$c_l = \int_0^\infty w_0(t)\, l_l(t)\, dt$$

или по теореме Парсеваля в частотной области:

$$c_l = \frac{1}{2\pi} \int_{-\infty}^{+\infty} W_0(j\omega)\, L_l(-j\omega)\, d\omega. \qquad (14.35)$$

Если задать передаточную функцию объекта $W_0(p)$, формула (14.35) позволяет найти коэффициенты ряда (14.31) и затем по формуле (14.34) оценить ошибку аппроксимации. Величину δ_N^2 также можно выразить с использованием равенства Парсеваля через передаточные функции объекта и фильтров Лагерра.

Для многих технологических объектов, аппроксимируемых моделью с сосредоточенными параметрами, передаточная функция может быть представлена в виде дроби

$$W_0(p) = \frac{K_0}{(1 + T_0 p)^n}, \qquad n = 1, 2, \ldots. \qquad (14.36)$$

Введем в рассмотрение параметр

$$\beta = \alpha T_0, \qquad (14.37)$$

характеризующий соотношение постоянных времени объекта и фильтра Лагерра. Вычисляя коэффициенты c_l, $l = 0, \ldots, N$, по формуле (14.35) и подставляя их в соответствующее выражение для расчета δ_N^2, можно получить величину среднеквадратичного

отклонения (14.34) в следующем виде [8]:

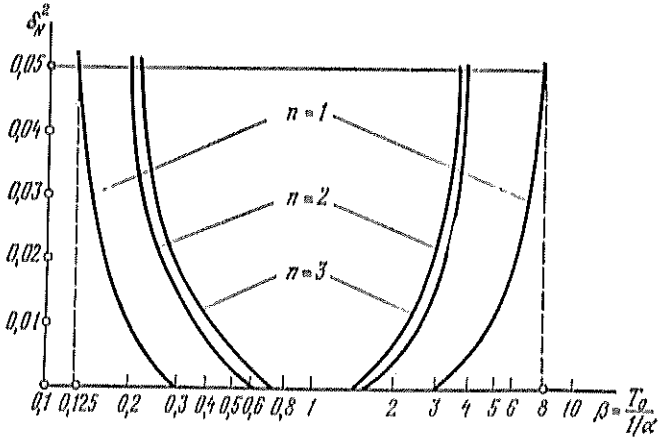
$$\delta_N^2 = 1 - \beta \frac{2^{2n}}{(2n-2)!} \sum_{i=0}^{N} (i!) \left\{ \sum_{j=0}^{i} \frac{(n+j-1)!\,(-2\beta)^j \left(\frac{1}{1+\beta}\right)^{n+j}}{(n-j)!\,(j!)^2} \right\}^2. \quad (14.38)$$

Из (14.38) следует, что

$$\delta_N^2(\beta) = \begin{cases} 0 \text{ при } \beta \to 1 \text{ и } N \geqslant n, \\ 1 \text{ при } \beta \to 0 \text{ или } \beta \to \infty \text{ и } N < \infty, \end{cases} \quad (14.39)$$

$$\lim_{N \to \infty} \delta_N^2(\beta) = 0 \text{ при } 0 < \beta < \infty. \quad (14.40)$$

При ограниченном числе членов ряда ($N < \infty$) наивысшая точность ($\delta_N^2 = 0$) аппроксимации для объекта (14.36) достигается



Фиг. 14.4. Зависимость ошибки аппроксимации от коэффициента β при различных n.

при $\beta = 1$ (т. е. при $T_0 = 1/\alpha$), когда постоянные времени объекта и фильтра Лагерра совпадают.

На фиг. 14.4 изображен график $\delta_N^2(\beta)$ для $N = 5$ (см. также табл. 14.1).

При практическом решении задачи идентификации необходимо так задавать значение параметра β, чтобы полученный результат обладал наивысшей точностью. Для этого в первом приближении предполагают, что объект является инерционным звеном первого порядка. Из априорных сведений о времени переходного процесса определяют величину T. Затем задают $\alpha = 1/T$, что соответствует $\beta = 1$.

На самом деле объект может не быть инерционным звеном первого порядка, а время переходного процесса известно лишь

Таблица 14.1

Верхняя и нижняя границы параметра β для инерционных объектов *n*-го порядка

n	$\beta_{мин}$	$\beta_{макс}$
1	0,125	8,0
2	0,23	3,8
3	0,35	3,0

приближенно. Поэтому в процессе идентификации приходится корректировать значение α.

Подводя итог, отметим, что с помощью функций Лагерра можно при заданном α с достаточной точностью аппроксимировать динамические характеристики объектов в широком диапазоне изменения постоянных времени.

14.4.3. Изучающий сигнал для активного метода идентификации

С точки зрения точности идентификации наилучшим оказывается сигнал с равномерной спектральной плотностью, т. е. сигнал типа „белый шум“ [1].

В практических задачах речь может идти только о некоторой аппроксимации белого шума. Помимо достаточно равномерной спектральной плотности, сигнал, аппроксимирующий белый шум, должен быть по технологическим причинам ограничен по амплитуде и должен относительно просто вычисляться на электронных вычислительных машинах, если при идентификации используются средства вычислительной техники.

Легко показать, что изучающий сигнал должен обладать следующими свойствами:

1)
$$n(l) = \begin{cases} +a \\ \text{или} \\ -a, \end{cases} \qquad (14.41)$$

где a — предельно допустимое значение амплитуды изучающего сигнала;

2) $n(l)$ может изменять свои значения только в дискретные моменты времени $l_k = kT_0$, где k — целое положительное число, T_0 — время такта.

---

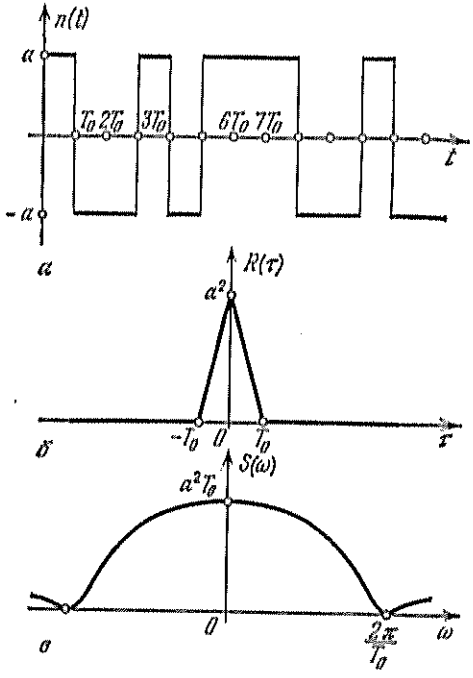
[1] Другие подходы к выбору наилучшего типа и параметров входного сигнала рассматриваются в работах [1—12], указанных в списке дополнительной литературы к данному разделу. — *Прим. ред.*

Если вероятности состояний $+a$ и $-a$ равны между собой и значения, взятые в разные моменты времени, некоррелированны, то такой сигнал называют двоичным белым шумом. Двоичный белый шум $n(t)$ имеет следующие статистические характеристики:

$$R_{nn}(\tau) = \begin{cases} a^2(1-|\tau|/T_0), & \text{если } |\tau| \leqslant T_0, \\ 0, & \text{если } |\tau| > T_0, \end{cases} \tag{14.42}$$

$$S_{nn}(\omega) = a^2 T_0 \left(\frac{\sin 0{,}5\omega T_0}{0{,}5\omega T_0}\right)^2. \tag{14.43}$$

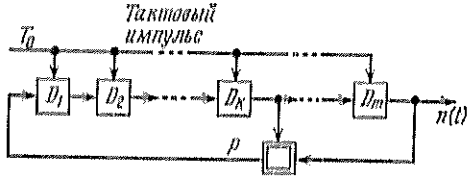
Пример реализации сигнала $n(t)$, его автокорреляционная функция $R_{nn}(\tau)$ и спектральная плотность $S_{nn}(\omega)$ показаны на



Фиг. 14.5. Двоичный белый шум и его характеристики.

фиг. 14.5. Для получения сигналов, близких по своим свойствам к двоичному белому шуму, часто применяют псевдослучайные периодические последовательности, называемые M-последовательностями [1, 8, 9]. В качестве генераторов M-последовательностей используются регистры с обратной связью. Эта обратная связь охватывает два или более элементов регистра (фиг. 14.6). В обратную связь включен логический элемент, выход которого обозначен P. На фиг. 14.6 выходная величина

i-го элемента обозначена D_i, $i=1, \ldots, m$, причем D_i может принимать значения 0 и 1 или $+1$ и -1. Такая схема получения псевдослучайной последовательности была предложена Хаффманом. Поэтому соответствующий сигнал иногда называют сигналом Хаффмана. Элемент в цепи обратной связи должен реализовывать для последовательности чисел 0 и 1 одну из двух функ-



Фиг. 14.6. Генератор для получения псевдослучайного двоичного сигнала.

ций алгебры логики: функцию сложения по модулю два или функцию эквивалентности.

Функция сложения по модулю два ($P=D_k \oplus D_m$) определяется таблицей

D_k	0	1	0	1
D_m	0	0	1	1
$P=D_k \oplus D_m$	0	1	1	0

(14.44)

Функция эквивалентности ($P=D_k \sim D_m$) определяется таблицей

D_k	0	1	0	1
D_m	0	0	1	1
$P=D_k \sim D_m$	1	0	0	1

(14.45)

Из сравнения таблиц следует, что функция сложения по модулю два и функция эквивалентности связаны соотношением

$$D_k \oplus D_m = \overline{D_k \sim D_m}.$$

Введем дискретное время и обозначим через $D_i(n)$ и $P(n)$ состояния элементов регистра в моменты времени $t=nT_0$, где n — целое положительное число или нуль. Алгоритм получения последовательности запишем в виде

$$\begin{aligned} &D_i(n)=D_{i-1}(n-1), \quad i=2, \ldots, m, \\ &D_1(n)=P(n-1), \\ &\text{Выход: } N(n)=D_m(n). \end{aligned} \tag{14.46}$$

Таблица 14.2

Данные, необходимые для разработки программ получения двоичных последовательностей

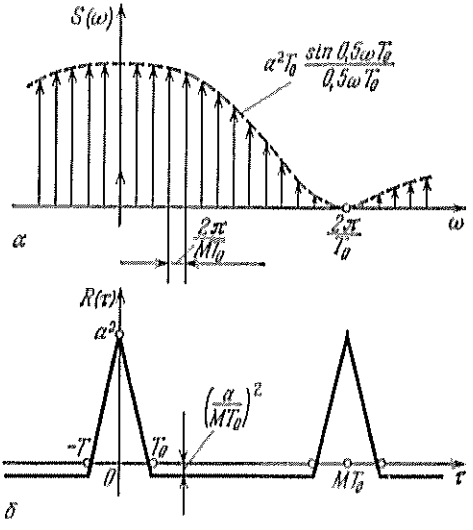
Число элементов в регистре m	Период последовательности $M = 2^m - 1$	Номера элементов в цепи обратной связи	
		k	m
3	7	2	3
4	15	3	4
5	31	2 3	5 5
6	63	5	6
7	127	2 3 5	7 7 7
9	511	4 5	9 9
10	1 023	3 7	10 10
11	2 047	2 9	11 11
15	32 767	4 7 8 11 13	15 15 15 15 15
17	131 071	3 5 6 11 14	17 17 17 17 17

Последовательности, состоящие из 0 и 1, легко могут быть получены на электронных вычислительных машинах с помощью программ, реализующих логические функции (14.44) или (14.45).

При использовании последовательностей, состоящих из -1 и $+1$, функция обратной связи реализуется как операция обычного умножения. В этом случае такая функция совпадает с функцией эквивалентности. Последовательности, состоящие из -1 и

$+1$, удобны для различных задач идентификации, так как с их помощью можно непосредственно получить центрированный и нормированный (с единичной амплитудой) двоичный белый шум. Как правило, для получения этих последовательностей используются высокопроизводительные вычислительные машины. Все данные, необходимые для применения программ такого рода, приведены в табл. 14.2.

Существенными параметрами двоичного изучающего сигнала, получаемого с помощью такой последовательности, являются



Фиг. 14.7. Корреляционная функция и спектральная характеристика псевдослучайного двоичного сигнала.

период $T_R = MT_0$, где $M = 2^m - 1$, и длительность такта T_0. Рассмотрим вопрос о наилучшем выборе этих параметров для решения задачи идентификации.

Прежде всего заметим, что из-за периодичности двоичного сигнала его спектральная характеристика оказывается дискретной:

$$S_{ии}(\omega) = \frac{a^2 T_0}{M}\left\{\sum_{k=-\infty}^{k=+\infty}(M+1)\left(\frac{\sin 0{,}5\omega T_0}{0{,}5\omega T_0}\right)^2 \delta(\omega-\omega_k) + \frac{M}{T_R}\delta(\omega)\right\}, \tag{14.47}$$

где $k = \pm 1, \pm 2, \ldots; k \neq 0$; $\omega_k = (2\pi/T_R)k$.

На фиг. 14.7, *а* показана одна из таких спектральных характеристик. Автокорреляционная функция изучающего сигнала изображена на фиг. 14.7, *б*. Эта корреляционная функция является периодической (с тем же периодом $T_R = MT_0$, что и

Таблица 14.3

Время памяти $T_п$ для фильтров Лагерра

$$L_l(p) = \frac{1}{p+\alpha}\left(\frac{p-\alpha}{p+\alpha}\right)^l$$

Точность установления	Номер фильтра l						
	0	1	2	3	4	5	6
$\Delta = 0{,}01$	$\frac{4{,}63}{\alpha}$	$\frac{7{,}37}{\alpha}$	$\frac{10{,}0}{\alpha}$	$\frac{12{,}4}{\alpha}$	$\frac{14{,}6}{\alpha}$	$\frac{17{,}0}{\alpha}$	$\frac{19{,}2}{\alpha}$
$\Delta = 0{,}001$	$\frac{6{,}95}{\alpha}$	$\frac{10{,}0}{\alpha}$	$\frac{12{,}7}{\alpha}$	$\frac{15{,}4}{\alpha}$	$\frac{17{,}9}{\alpha}$	$\frac{20{,}2}{\alpha}$	$\frac{22{,}8}{\alpha}$

у самого сигнала) и имеет некоторую постоянную составляющую. В остальном корреляционная функция изучающего сигнала идентична корреляционной функции идеального белого шума.

Для того чтобы сформулировать условия применимости такого изучающего сигнала, определим понятие „времени памяти системы". Под этим термином мы будем понимать интервал времени, в течение которого необходимо запоминать входную переменную системы с тем, чтобы в момент t можно было вычислить сигнал y на выходе с требуемой точностью. Очевидно, что время памяти совпадает с временем $T_п$ установления переходного процесса. Временем памяти для инерционного звена является время, в течение которого выходной процесс входит в 5%-ную зону нового стационарного состояния. В этом случае время установления переходного процесса равно трем постоянным времени. Рассмотрим следующий пример.

Пусть передаточная функция инерционного звена

$$W(p) = \frac{K}{1+Tp} \quad \text{и} \quad \Delta = 0{,}05\frac{K}{T}; \quad T_п = 3T.$$

Для этого инерционного звена можно составить следующую таблицу:

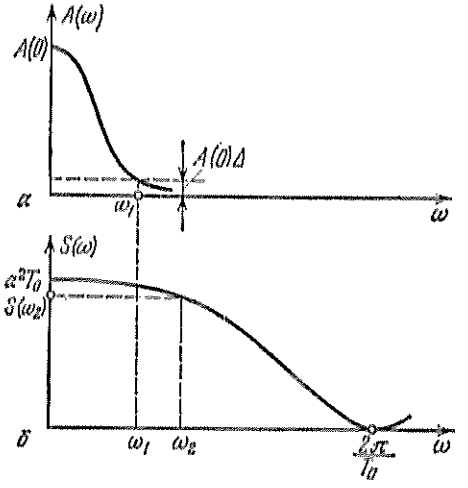
$\Delta/(K/T)$	0,05	0,01	0,001
$T_п/T$	3	4,63	6,95

Аналогичные данные были получены и для фильтров Лагерра (табл. 14.3).

Время памяти объекта позволяет оценить необходимый период изучающего сигнала. Для того чтобы этот сигнал воспринимался объектом как случайный двоичный шум, необходимо выполнение следующего простого соотношения:

$$T_R > T_п, \tag{14.48}$$

т. е. период последовательности должен быть больше времени памяти объекта. Строго говоря, высказанные здесь соображения и вытекающее из них неравенство (14.48) обеспечивают необхо-



Фиг. 14.8. К определениям частоты среза ω_1 и граничной частоты ω_2 при заданном значении Δ.

димую равномерность спектральной характеристики изучающего сигнала лишь в области низких частот.

Время памяти объекта до решения задачи идентификации, как правило, бывает не известно. Поскольку объект аппроксимируется суммой функций Лагерра, можно предположить, что наименьшее время памяти фильтров определяет также и общую величину $T_п$. Если при разложении в ряд искомой характеристики ограничиться шестью фильтрами Лагерра, то из табл. 14.3 следует

$$T_п > \frac{22{,}8}{\alpha}. \tag{14.49}$$

Из фиг. 14.5 и 14.7, а также из соответствующих формул (14.42) и (14.47) следует, что равномерность спектральной плотности изучающего сигнала в области высоких частот определяется величиной T_0. Конечно, нет необходимости требовать, чтобы спектральная плотность была равномерной за пределами полосы

Таблица 14.4

Частоты среза для фильтров Лагерра при заданных значениях Δ

Δ	0,16	0,09	0,07	0,05	0,03	0,01
$\frac{\omega_1}{2\pi\alpha}$	0,17	0,28	0,36	0,51	0,84	2,52

Таблица 14.5

Зависимость неравномерности спектра от ширины полосы частот (0, ω_2)

Δ, %	100	60	26	19	13	9	7	6	5	4	2,5	1,5
$\omega_2 \times \frac{T_0}{2\pi}$	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{9}$	$\frac{1}{10}$	$\frac{1}{16}$	$\frac{1}{20}$

пропускания объекта. Поскольку объект аппроксимируется функциями Лагерра, рассмотрим частотную характеристику соответствующих фильтров. Из определения амплитудно-частотной характеристики имеем

$$A_l(\omega) = |L_l(j\omega)| = \frac{\sqrt{2\alpha}}{\sqrt{\alpha^2 + \omega^2}}, \qquad (14.50)$$

т. е. $A_l(\omega)$ одинаковы для всех фильтров.

Частотой среза называют такое значение ω_1, для которого при $\omega > \omega_1$ коэффициент $A(\omega)/A(0)$ меньше заданной величины Δ (фиг. 14.8, *а*):

$$\frac{A(\omega)}{A(0)} \leqslant \Delta.$$

В табл. 14.4 приведены значения ω_1 для различных Δ, вычисленные с учетом условия (14.48).

Введем следующий коэффициент, характеризующий неравномерность спектра:

$$\Delta S(\omega) = \frac{S_{\text{макс}} - S(\omega)}{S_{\text{макс}}} \times 100\%. \qquad (14.51)$$

Граничной частотой ω_3 при заданном значении Δ назовем такую, для которой из условия $\omega \geqslant \omega_3$ следует $\Delta S(\omega) \geqslant \Delta$ (фиг. 14.8, *б*). В табл. 14.5 представлены значения ω_3, вычисленные с применением формулы (14.51) для различных Δ. Очевидно, должно

выполняться неравенство (фиг. 14.8)

$$\omega_2 \geqslant \omega_1. \tag{14.52}$$

Задаваясь допустимыми значениями коэффициентов $A(\omega)/A(0)$ и $\Delta S(\omega)$, из табл. 14.4 и 14.5, а также из соотношения (14.52) можно найти связь между величинами T_0 и α в виде соответствующего неравенства. Например, задавая $A(\omega)/A(0) = 0{,}05$, находим $\omega_1 = 0{,}51 \cdot \alpha \cdot 2\pi$. Пусть задано $\Delta S = 5\%$. Тогда из табл. 14.5 получаем $\omega_2 = 2\pi/9T_0$.

Из неравенства (14.52) следует

$$\frac{2\pi}{9T_0} \geqslant 0{,}51\alpha \cdot 2\pi$$

или

$$\frac{1}{\alpha T_0} \geqslant 4{,}5.$$

Для $A(\omega)/A(0) = 0{,}03$ и $\Delta S = 2{,}5\%$ аналогично получаем $1/(\alpha T_0) \geqslant 13{,}4$.

К сожалению, нет достаточно обоснованных критериев для выбора допустимых коэффициентов $A(\omega)/A(0)$ и $\Delta S(\omega)$. Кроме того, не только эти величины определяют выбор T_0. Критерии для оптимального выбора значений этих коэффициентов можно было бы получить после длительной работы по моделированию алгоритма при воздействии помех и после анализа ошибок этого метода идентификации [8]. Такие исследования показали, что алгоритм идентификации в основном можно характеризовать следующим соотношением:

$$\frac{1}{\alpha T_0} = 8. \tag{14.53}$$

Применяя формулы (14.53) и (14.49) для вычисления периода изучающего сигнала, получим такое неравенство: $T_R > 22{,}8/\alpha$. Из (14.53) следует: $1/\alpha = 8T_0$. Поэтому имеем

$$T_R > 192\, T_0$$

и

$$\frac{T_R}{T_0} = M > 192. \tag{14.54}$$

Из табл. 14.2 видно, что ближайшим подходящим значением является $M = 511$.

11.4.4. Основные рекомендации по применению активного метода идентификации с псевдослучайным двоичным сигналом

При использовании псевдослучайного двоичного сигнала необходимо определять параметр α функций Лагерра. Если этот параметр известен, с его помощью можно найти все остальные

параметры. В частности, величина такта изучающего сигнала с учетом (14.53) будет равна

$$T_0 = \frac{1}{8\alpha}. \qquad (14.55)$$

Период M изучающего сигнала, как уже отмечалось выше, должен быть больше 192. Для двоичного псевдослучайного сигнала могут быть взяты следующие значения: $M = 511$; 1023; 2047 и т. д. Чем большим будет период, тем меньшей окажется ошибка метода. Однако одновременно возрастает время, требуемое для исследования объекта. Это время должно охватывать целое число периодов изучающего сигнала. Кроме того, было показано, что при разложении искомой весовой функции обычно требуется учитывать от 5 до 7 функций Лагерра ($N = 4 \div 6$).

Поскольку время, отводимое для анализа данных, ограничено, точное значение коэффициента корреляции R_{uj} не известно и используют его оценку $\hat{R}_{uj}$. Но если в правой части системы (14.18) будут стоять оценки $\hat{R}_{uj}$, то и решение системы позволит получить лишь оценки коэффициентов $\hat{c}_j$:

$$\sum_{l=0}^{N} \hat{c}_l R_{lj} = \hat{R}_{uj}. \qquad (14.56)$$

Можно предложить следующий алгоритм выбора параметра α:

1. На основании ориентировочной оценки времени переходного процесса $T_п$ определяется эквивалентная постоянная времени $T_1 = T_п/4$. При этом в первом приближении предполагается, что объект можно представить в виде инерционного звена первого порядка.

2. Постоянная времени фильтров Лагерра полагается равной T_1, т. е. $1/\alpha_1 = T_1$. Следовательно, на первом этапе значение параметра α_1 при идентификации задается равным

$$\alpha_1 = \frac{1}{T_1}.$$

Длительность такта изучающего сигнала, с помощью которого будет проводиться первый этап идентификации, рассчитывается по формуле

$$T_{01} = \frac{1}{8\alpha_1} = \frac{T_п}{32}. \qquad (14.57)$$

3. После того как будут определены параметры T_{01} и α_1, на объект подается изучающий сигнал и проводится идентификация. В результате получаются оценки коэффициентов $\hat{c}_{1l}$.

4. В комплексном пространстве строится изображение следующей функции для различных ω:

$$\hat{W}_1(j\omega) = \sum_{l=0}^{N} \hat{c}_{1l} \frac{(j\omega - \alpha_1)^l}{(j\omega + \alpha_1)^{l+1}}, \qquad (14.58)$$

где $0 \leqslant \omega \leqslant 3{,}2\,\alpha_1$.

Значение $3{,}2\alpha_1$ соответствует частоте среза объекта при $\Delta = 0{,}05$ (см. табл. 14.4).

5. Наилучшая аппроксимация функции (14.58) отыскивается в следующем виде:

$$W(j\omega) = \frac{K}{(1 + j\omega T)^m}, \qquad (14.59)$$

где m последовательно принимает значения 1, 2, 3. Метод аппроксимации заключается в расчете обратной частотной характеристики с использованием оценок $\hat{c}_{1l}$:

$$[\hat{W}_1(j\omega)]^{-1/m}. \qquad (14.60)$$

Уравнение (14.59) может быть записано в такой форме:

$$[W(j\omega)]^{-1/m} = \sqrt[m]{1/K}\,(1 + j\omega T). \qquad (14.61)$$

Тогда задача нахождения $(K)^{-1/m}$ и T в уравнении (14.61) может быть решена с помощью стандартных программ метода наименьших квадратов [1]. Для каждой из трех пар найденных с помощью этих программ значений K и T, соответствующих $m = 1, 2, 3$, определяется наибольшее абсолютное отклонение (14.61) от (14.60). В качестве окончательного решения выбирается та пара значений K и T, для которой наибольшее отклонение окажется минимальным.

---

[1] Положим

$$[\hat{W}_1(j\omega)]^{-1/m} = y_1(\omega) + jy_2(\omega),$$

а

$$\sqrt[m]{\frac{1}{K}}\,(1 + j\omega T) = b_1 + j\omega b_2.$$

Тогда оценки $\hat{b}_1$ и $\hat{b}_2$ можно найти из условия

$$\min_{b_1,\, b_2} \sum_{l=1}^{L} [(y_1(\omega^l) - b_1)^2 + (y_2(\omega^l) - \omega^l b_2)^2],$$

где L — число рассчитанных точек функции $\hat{W}(j\omega)$, по которым проводится аппроксимация. Получаем

$$\hat{b}_1 = \sum_{l=1}^{L} y_1(\omega^l)/L, \quad \hat{b}_2 = \sum_{l=1}^{L} y_2(\omega^l)\,\omega^l \Big/ \sum_{l=1}^{L} (\omega^l)^2.$$

Отсюда $K = \hat{b}_1^{-m}$, $T = \hat{b}_2/\hat{b}_1$. — *Прим. ред.*

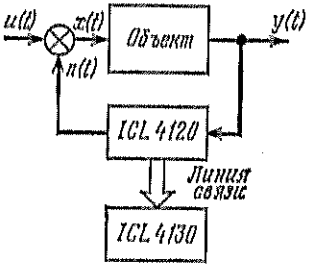
6. Найденное значение T используется для расчета α. Затем вновь проводится идентификация. В разд. 14.4.2 (см. также фиг. 14.4) было показано, что наибольшая точность аппроксимации весовой функции суммой функций Лагерра достигается при $\alpha = 1/T$. Поэтому новая идентификация проводится при

$$\alpha_2 = \frac{1}{T}. \tag{14.62}$$

Длительность такта изучающего сигнала устанавливается равной

$$T_{02} = \frac{1}{8T}.$$

Наиболее просто представленный здесь метод исследования промышленных объектов можно реализовать с помощью управляющих вычислительных машин. В качестве примера укажем



Фиг. 14.9. Пример идентификации объекта управления.

на иерархическую систему управляющих ЭВМ, с помощью которой этот метод применялся для идентификации объектов в режиме реального времени на нефтехимическом комбинате в г. Шведте-на-Одере (ГДР) (фиг. 14.9). Вычислительная машина нижнего уровня генерирует псевдослучайный двоичный изучающий сигнал и подает его на вход объекта. При этом период сигнала был принят равным $M = 2047$. Параметр сигнала T_0 мог задаваться как целое число из интервала $T_0 = 1 - 30$ мин. Сигнал на выходе объекта измеряется ЭВМ нижнего уровня и по системе связи передается в ЭВМ верхнего уровня. Естественно, что вычислительная система не только проводит идентификацию, но и решает другие задачи. В ЭВМ верхнего уровня выполняется обработка поступающего сигнала с целью расчета оценок начальных значений взаимно корреляционных функций $\hat{R}_{uj}$ при $0 \leqslant j \leqslant 6$.

Такая обработка производится периодически через каждые 2 ч и занимает 30 с вычислительного времени. Результаты расчетов [включая коэффициенты R_{ij} системы (14.20)], поступившая информация и программа идентификации хранятся на магнитных

дисках. Коэффициенты R_{ij} системы (14.20) рассчитываются только один раз и с опережением, поскольку заранее известен алгоритм генерации изучающего сигнала. Кроме того, известна и величина $T_0 = 1/8\alpha$. Одновременно со значениями $\hat{R}_{ij}$ рассчитываются также дисперсии σ^2_{Rj} оценок $\hat{R}_{ij}$, которые уменьшаются по мере увеличения времени идентификации [1]. Если дисперсия σ^2_{Rj} становится меньше некоторого заданного значения для всех j, то идентификация считается завершенной.

С помощью алгоритма идентификации в указанной вычислительной системе было проведено исследование динамических характеристик процесса каталитической конверсии метана. Особое внимание при этом было уделено каналу «Расход метана — Температура в зоне реакции». Из предварительных исследований было известно ориентировочное время переходного процесса: $T_п \approx 60 - 75$ мин. Длительность такта изучающего сигнала была выбрана, исходя из (14.57), равной

$$T_{01} = \frac{T_п}{32} = 1{,}9 - 2{,}3 \text{ мин}.$$

Для исследования было принято $T_{01} = 2$ мин. Тогда значение α_1 оказалось равным $\alpha_1 = 1/8T_{01} = 1/16$. Идентификация продолжалась 35 ч, и в результате были рассчитаны и выведены на печать первые приближения оценок коэффициентов $\hat{c}_{1l}$:

$\hat{c}_{10}$	$\hat{c}_{11}$	$\hat{c}_{12}$	$\hat{c}_{13}$	$\hat{c}_{14}$	$\hat{c}_{15}$	$\hat{c}_{16}$
2,95	−3,105	0,96	0,446	0,481	0,273	0,142

С применением этих оценок коэффициентов были найдены значения $\hat{W}(j\omega)$ частотной характеристики (14.58). При этом частота изменялась в следующем диапазоне:

$$0 \leqslant \omega \leqslant 3{,}2 \times \frac{1}{16} = 0{,}2 \text{ мин}^{-1}.$$

Для $\hat{W}_1(j\omega)$ справедливо выражение

$$\hat{W}_1(j\omega) = \sum_{l=0}^{6} \hat{c}_{1l} \frac{(j\omega - 1/16)^l}{(j\omega + 1/16)^{l+1}}.$$

Затем с помощью специальной программы на ЭВМ были вычислены значения $[\hat{W}_1(j\omega)]^{-1/m}$ для $m = 1, 2, 3$ в соответствии с формулой (14.61). После этого методом наименьших квадратов была

[1] В данном случае в силу ортогональности сигналов на выходе фильтров оценки коэффициентов оказываются пропорциональными оценке $\hat{R}_{ij}$. Поэтому дисперсия оценки $\hat{R}_{ij}$ пропорциональна дисперсии оценки $\hat{c}_j$. — *Прим. ред.*

произведена аппроксимация полученных результатов функцией вида $\sqrt[m]{1/K}\,(1+j\omega T)$.

Значение m и соответствующая пара значений (K, T), для которых оказалось минимальным наибольшее отклонение величины $[\hat{W}(j\omega)]^{-1/m}$ от найденной аппроксимирующей зависимости, были выведены на печать. В рассмотренном примере они оказались равными

$$m = 3,$$
$$T = 7{,}31 \text{ мин},$$
$$K = 104{,}4.$$

Это означает, что наилучшей аппроксимацией найденных оценок частотной характеристики является следующая функция:

$$\hat{W}_1(j\omega) = \frac{104{,}4}{(1+j\,7{,}31\omega)^3}. \qquad (14.63)$$

Результаты аппроксимации были использованы для коррекции параметров α и T_0. В соответствии с (14.62) получено значение

$$\alpha_2 = \frac{1}{7{,}3} = 0{,}137.$$

Соответствующая длительность такта изучающего сигнала равна

$$T_{02} = \frac{1}{8\alpha_2} = 0{,}92 \text{ мин}.$$

Округленно принимается $T_{02} = 1$ мин. Это значение используется программой для генерации изучающего сигнала. В программе обработки данных полагается

$$\alpha_2 = \frac{1}{8T_{02}} = \frac{1}{8}.$$

Затем проводится новая идентификация. Оценки коэффициентов, рассчитанные спустя 42 ч, оказались такими:

$\hat{c}_{20}$	$\hat{c}_{21}$	$\hat{c}_{22}$	$\hat{c}_{23}$	$\hat{c}_{24}$	$\hat{c}_{25}$	$\hat{c}_{26}$
4,21	−7,982	3,365	0,427	0,210	0,133	0,084

Эти оценки считаются окончательными и могут быть использованы для решения различных задач управления.

Представляет интерес аппроксимировать найденную динамическую характеристику с помощью последовательного включения трех инерционных звеньев с одинаковыми постоянными времени. Результат при этом хорошо согласуется с (14.63):

$$\hat{W}_2(j\omega) = \frac{106{,}9}{(1+j\,7{,}55\omega)^3}. \qquad (14.64)$$

Такое хорошее соответствие указывает на малую чувствительность этого метода идентификации по отношению к значению параметра α.

ЛИТЕРАТУРА

1. Himmelblau D. M., Process Analysis by Statistical Methods, Wiley, New York, 1970; русский перевод: Химмельблау Д., Анализ процессов статистическими методами, изд-во «Мир», М., 1974.
2. Теория автоматического управления, под ред. А. В. Нетушила, ч. 1, изд-во «Высшая школа», М., 1968.
3. Дудников Е. Г. и др., Построение математических моделей химико-технологических объектов, изд-во «Химия», М., 1970.
4. Солодовников В. В., Статистическая динамика линейных систем автоматического управления, Физматгиз, М., 1960.
5. Райбман Н. С. и др., Адаптивные модели в системах управления, изд-во «Советское радио», М., 1966.
6. Тихонов А. Н., О решении некорректно поставленных задач и методе регуляризации, *ДАН СССР*, **151**, № 3 (1963).
7. Тихонов А. Н., Гласко В. Б., О приближенном решении интегральных уравнений Фредгольма I рода, *Журн. вычисл. мат. и мат. физ.*, **4**, № 3 (1964).
8. Александровский Н. М., Егоров С. В., Кузин Р. Е., Адаптивные системы управления сложными технологическими объектами, изд-во «Энергия», М., 1972.
9. Pittermann F., Schweizer G., Erzeugung und Verwendung von binären Rauschen bei Flugversuchen, *Regelungstechnik*, **14**, № 2 (1966).
10. Теория автоматического управления, под ред. А. В. Нетушила, ч. 2, изд-во «Высшая школа», М., 1972.
11. Müller J. A., Regelstreckenanalyse mittels adaptiver Modelle, T. 1 u. 2, *Messen, Steuern, Regeln*, № 2 u. 4 (1968).
12. Strobel H., Systemanalyse mit determinierten Testsignalen, Technik, Berlin, 1968.
13. Материалы симпозиума IFAC «Identification and Parameter Estimation», Прага, 1968 и 1970.
14. Егоров С. В., Способ определения динамических характеристик сложных объектов, *Автоматика и телемеханика*, № 12 (1966).
15. Wilfert H.-H., Signal- und Frequenzanalyse, Technik, Berlin, 1967.
16. Materialen der Arbeitstagung «Algorithmisierte Prozessanalyse», Zentralinstitut für technische Kybernetik und Informationsprogramme, DAW, Berlin u. DGMA, KdT, Dresden, 13—14, Mai 1971.
17. Балакирев В. С., Дудников Е. Г., Цирлин А. М., Экспериментальное определение динамических характеристик промышленных объектов управления, изд-во «Энергия», М., 1967.
18. Materialen des XV. Internationalen Wissenschaftl. Kolloquiums über Technische und Biomedizinische Kybernetik, Ilmenau, 28. Sept.—2. Okt. 1970.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Levadi V. S., Design of Input Signals for Parameter Estimation, *IEEE Trans. on Autom. Control*, AC-11, № 2 (1966).
2. Aoki M., Staley R. M., On the Signal Synthesis in Parameter Identification, *Automatica*, № 6 (1970).
3. Keviczky L. K., Banyasz Cs., Optimal Identification by Simulation of the Information Obtained from Processes, Summer Simulation Conf., San Diego, 1972.

4. Журавлев Л. В., Цирлин А. М., О выборе формы пробного воздействия для определения характеристики объектов при наличии шума, см. сб. „Идентификация и аппаратура для статистических исследований", изд-во „Наука", М., 1970.
5. Бахвалов Л. А., Лецкий Э. К., Об одной задаче синтеза входного сигнала при статистической идентификации линейных динамических систем, *Техн. киберн.*, № 3 (1974).
6. Планирование и автоматизация эксперимента в научных исследованиях, сб. статей под ред. Г. К. Круга, изд-во "Советское радио", М., 1974.
7. Бородюк В. П., Кириченко А. В., Кузнецов В. Е., Один подход к планированию оптимального эксперимента на динамических моделях, см. сб. „Применение планирования эксперимента в радиоэлектронике и смежных областях техники", МДНТП им. Ф. Э. Дзержинского, М., 1975.
8. Arimoto S., Kimura H., Optimum Input Test Signals for Systems Identification—an Information-Theoretical Approach, *Int. J. Syst. Sci.*, 1, № 3 (1971).
9. Litman S., Higgins W. H., Growing Exponentials as a Probing Signal for System Identification, *Proc. IEEE*, 51 (June 1963).
10. Кузьмин И. В., Оценка эффективности и оптимизация АСКУ, изд-во „Советское радио", М., 1971.
11. Лившиц К. И., Тренугов А. Ф., О выборе сигналов при идентификации линейных систем по методу наименьших квадратов, *Техн. киберн.*, № 5 (1974).
12. Лецкий Э. К., О применении автоматизированных систем при экспериментальных исследованиях объектов управления, Труды МИНТ, вып. 503, М., 1975.

15.

МАТЕМАТИЧЕСКОЕ ПРИЛОЖЕНИЕ

А. СПЕЦИАЛЬНЫЕ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Разделы книги, предшествующие данному, предполагают знакомство читателя с рядом математических дисциплин, таких, как линейная алгебра, теория вероятностей и математическая статистика, теория случайных процессов и т. п. Изложение всех этих вопросов в рамках данной книги невозможно даже в весьма краткой форме. Ряд основополагающих положений математической статистики содержится во 2-м и 3-м разделах книги. Другие из вышеназванных дисциплин излагаются в различных книгах [1—9]. В этом приложении будут рассмотрены только некоторые специальные распределения вероятностей, которые используются во всех предшествующих разделах книги.

Некоторые определения

При аналитической записи выражений для некоторых распределений понятие „факториал"

$$n! = 1 \cdot 2 \ldots n \tag{15.1}$$

используется для нецелых величин. Такое расширение понятия „факториал" приводит к гамма-функции.

Определение 5.1

Пусть x — некоторое действительное число, причем $x \neq 0, -1, -2, -3, \ldots$. Тогда интеграл

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\, dt \tag{15.2}$$

называют гамма-функцией [1].

[1] Отметим, что $\Gamma(n+1) = n!$ при $n = 0, 1, 2, \ldots$. — *Прим. ред.*

15.1. Биномиальное распределение

Пусть $(A_1, A_2, \ldots, A_k)$—полная система событий, причем

$$P(A_l) = p_l, \quad \sum_{l=1}^{k} p_l = 1. \tag{15.3}$$

Тогда вероятность $P_n(n_1, n_2, \ldots, n_k)$ того, что при n испытаниях событие A_l произойдет точно n_l раз $\left(\sum_{l=1}^{k} n_l = n\right)$ определяется выражением

$$P_n(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1!\, n_2! \ldots n_k!}\, p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}. \tag{15.4}$$

Выражение (15.4) описывает так называемое полиномиальное распределение, причем термин „полиномиальное" связан с тем, что

$$(p_1 + p_2 + \ldots + p_k)^n = \sum P_n(n_1, n_2, \ldots, n_k),$$

где сумма берется по всем возможным наборам значений

$$n_1, n_2, \ldots, n_k \text{ при } \sum_{l=1}^{k} n_l = n.$$

Пусть теперь возможны только два события A и $\bar{A}$, причем $P(A) = p$, а $P(\bar{A}) = 1 - p$. Тогда, пользуясь (15.4), можно записать выражение для вероятности $P_n(m)$ того, что событие A в n испытаниях произойдет ровно m раз:

$$P_n(m) = \frac{n!}{m!\,(n-m)!}\, p^m (1-p)^{n-m} = \binom{n}{m} p^m (1-p)^{n-m}. \tag{15.5}$$

Выражение (15.5) описывает биномиальное распределение вероятностей. Если m—случайная величина, распределенная по биномиальному закону (15.5), то математическое ожидание равно

$$E\{m\} = np, \tag{15.6}$$

а дисперсия

$$E\{(m-np)^2\} = \sigma^2 = np(1-p). \tag{15.7}$$

15.2. Нормальное распределение

Биномиальный закон (15.5) описывается дискретным симметричным распределением вероятностей [значения $P_n(m)$ одинаковы для всех m, одинаково удаленных от среднего значения np]. Из (15.5) предельным переходом при $n \to \infty$ можно получить симметричное непрерывное распределение, называемое нормальным (или гауссовым) распределением. Функция нормального

распределения имеет вид

$$N_{m,\sigma}(x) = \int_{-\infty}^{x} n_{m,\sigma}(x)\,dx, \tag{15.8}$$

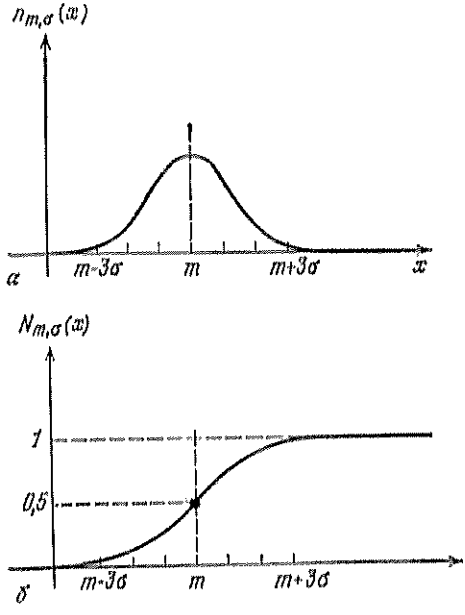
где $n_{m,\sigma}(x)$ — функция плотности нормального распределения, определяемая выражением

$$n_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\}. \tag{15.9}$$

Если случайная величина X распределена согласно (15.9), то

$$E\{X\} = m,\ D\{X\} = E\{(X-m)^2\} = \sigma^2. \tag{15.10}$$

Нормальный закон при среднем значении $m = 0$ и дисперсии $\sigma^2 = 1$ называют нормированным нормальным распределением.



Фиг. 15.1. Функция плотности (*а*) и функция распределения (*б*) вероятностей для нормального закона.

Его функция распределения

$$\Phi(x) = N_{0,1}(x) = \int_{-\infty}^{x} \varphi(x)\,dx, \tag{15.11}$$

где $\varphi(x)$ — функция плотности нормированного нормального распределения,

$$\varphi(x) = n_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}. \qquad (15.12)$$

Между нормированным и ненормированным нормальными распределениями существует следующая взаимосвязь:

$$N_{m,\sigma}(x) = \Phi\left(\frac{x-m}{\sigma}\right). \qquad (15.13)$$

Поэтому достаточно табулировать только функцию $\Phi(x)$.

Нормальное распределение занимает центральное место в вероятностных расчетах и в математической статистике. Так, часто полагают, что результаты наблюдений в экспериментальных исследованиях подчиняются нормальному закону. На фиг. 15.1 показаны функция плотности и функция распределения для нормального закона. Табл. 15.1 содержит функцию плотности $\varphi(x)$ нормированного нормального распределения. В силу симметрии нормального закона $\varphi(x) = \varphi(-x)$, и поэтому $\varphi(x)$ табулирована только для $x \geqslant 0$.

Кроме того, таблица содержит значения

$$\Phi(x) = \int_0^x \varphi(x)\,dx = \Phi(x) - 0{,}5, \qquad x \geqslant 0.$$

Функция $\Phi(x)$ называется функцией Лапласа [1]. В силу симметрии нормального закона имеем

$$\Phi(-x) = 0{,}5 - \overline{\Phi}(x).$$

Например,

$$\varphi(-0{,}35) = \varphi(0{,}35) = 0{,}3752,$$

$$\Phi(-0{,}35) = 0{,}5 - \overline{\Phi}(0{,}35) = 0{,}5 - 0{,}136831 = 0{,}363169.$$

15.3. χ^2-распределение

Пусть случайные переменные $X_1, X_2, \ldots, X_k$ подчинены нормированному нормальному распределению, причем n переменных независимы [остальные $(k-n)$ линейно связаны с этими переменными]. Тогда случайная величина

$$X = X_1^2 + X_2^2 + \ldots + X_k^2 \qquad (15.14)$$

подчинена χ^2-распределению с числом степеней свободы n.

[1] Иногда функцией Лапласа называют функцию $\overline{\Phi}(x) = \int_0^x \varphi(x)\,dx$. — *Прим. ред.*

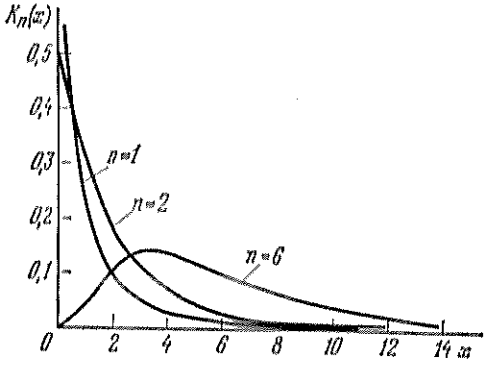
Функция распределения величины X

$$K_n(x) = \int_{-\infty}^{x} k_n(x)\,dx, \tag{15.15}$$

где $k_n(x)$ — функция плотности χ^2-распределения

$$k_n(x) = \begin{cases} 0 & \text{для } x < 0, \\ 2^{-n/2}\Gamma^{-1}\left(\frac{n}{2}\right) x^{(n/2)-1}\exp\left(-\frac{x}{2}\right) & \text{для } x \geqslant 0. \end{cases} \tag{15.16}$$

В табл. 15.2 приведены критические значения $e = \chi^2_{\text{кр}}$ χ^2-распределения для числа степеней свободы n, изменяющегося от 1



Фиг. 15.2. χ^2-распределение.

до 30, и различных вероятностей $P(\chi^2 < e)$. Значение $e = \chi^2_{\text{кр}}$ соответствует решению относительно e уравнения [1]

$$P(\chi^2 < e) = 1 - \alpha. \tag{15.17}$$

Для $n \geqslant 30$ χ^2-распределение достаточно хорошо представляется нормальным законом с математическим ожиданием n и дисперсией $2n$ [2]. На фиг. 15.2 показано, как изменяется характер χ^2-распределения при увеличении числа степеней свободы n.

[1] Величина e, найденная из (15.17), является квантилем порядка $(1-\alpha)$ χ^2-распределения с соответствующим числом степеней свободы. В тексте книги эта величина обозначается также через $\chi^2_{\text{кр}}$. Отметим, что (15.17) означает следующее: вероятность того, что $\chi^2 < e$, равна $(1-\alpha)$. — *Прим. ред.*

[2] Чаще пользуются аппроксимацией нормальным законом с центром $\sqrt{2n-1}$ и дисперсией 1 величины $\sqrt{2\chi^2}$. При этом квантиль $\chi^2_{\text{кр}}$ порядка $P = 1-\alpha$ при $n > 30$ вычисляется по формуле

$$\chi^2_{\text{кр}} \approx \frac{1}{2}\left(\sqrt{2n-1} + e\right),$$

15.4. t-распределение Стьюдента

Пусть случайные переменные X, X_1, X_2, ..., X_k подчинены нормальному распределению с нулевым средним и произвольной дисперсией σ^2. Пусть далее величина X не зависит от остальных X_i, а среди X_i имеется ровно n линейно независимых величин. Тогда случайная величина

$$T = \frac{X}{\sqrt{\frac{1}{n}(X_1^2 + X_2^2 + \ldots + X_k^2)}} \tag{15.18}$$

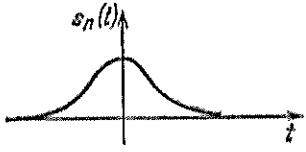
имеет функцию распределения вероятностей

$$P(T < t) = S_n(t) = \int_{-\infty}^{t} s_n(t)\,dt, \tag{15.19}$$

где $s_n(t)$ — функция плотности распределения:

$$s_n(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}. \tag{15.20}$$

Распределение, описываемое функциями (15.19) и (15.20), называется t-распределением Стьюдента с n степенями свободы.



Фиг. 15.3. t-распределение.

Заметим, что t-распределение не зависит от σ^2 (фиг. 15.3). Через $\Gamma(x)$ в (15.20) обозначена гамма-функция, определенная выше (см. определение (15.1)). При $n \to \infty$ t-распределение стремится к нормированному нормальному закону. В табл. 15.3 приведены

где ε находится с помощью таблиц нормированного нормального распределения из условия

$$\Phi(\varepsilon) = 0{,}5 - \alpha.$$

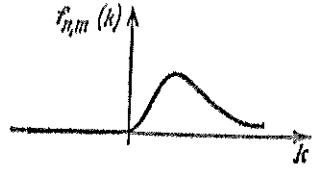
При близких к 0 или 1 величинах α более точную аппроксимацию дает следующее выражение:

$$\chi^2_{\text{кр}} \approx n\left(1 - \frac{2}{9n} + \varepsilon\sqrt{\frac{2}{9n}}\right)^3. — \textit{Прим. ред.}$$

значения e, соответствующие решению уравнения $P\{|T|<e\}=1-\alpha$ при различных значениях числа степеней свободы n и вероятности $(1-\alpha)$ [1].

15.5. F-распределение Фишера

Пусть случайные величины $X_1, X_2, \ldots, X_k$ и $Y_1, Y_2, \ldots, Y_l$ подчинены нормальному закону распределения с нулевым средним и произвольной дисперсией σ^2. Величины X_i не зависят



Фиг. 15.4. F-распределение.

от X_j. Кроме того, пусть среди X_i имеется n, а среди Y_j — m линейно независимых величин. Тогда случайная переменная

$$F=\frac{(1/n)\left(X_1^2+X_2^2+\ldots+X_k^2\right)}{(1/m)\left(Y_1^2+Y_2^2+\ldots+Y_l^2\right)} \tag{15.21}$$

с законом распределения вероятностей

$$P(F<k)=F_{n,m}(k)=\int\limits_{-\infty}^{k} f_{n,m}(k)\,dk \tag{15.22}$$

имеет функцию плотности распределения $f_{n,m}(k)$ вида

$$f_{n,m}(k)=\begin{cases} 0 & \text{для } k<0, \\ \dfrac{\Gamma\left(\dfrac{n+m}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)}\left(\dfrac{n}{m}\right)^{n/2}k^{(n/2)-1}\left(\dfrac{n}{m}k+1\right)^{-\frac{n+m}{2}} & \text{для } k\geqslant 0. \end{cases} \tag{15.23}$$

Распределение (15.22), (15.23) называют F-распределением Фишера [2] с числами степеней свободы n и m (фиг. 15.4). В табл. 15.4

[1] В тексте книги величина e, соответствующая решению уравнения $P\{|T|<e\}=1-\alpha$ при n степенях свободы, называется критическим значением t-распределения при уровне значимости α и обозначена через $t_{кр}$.— *Прим. ред.*

[2] Величина F, вычисляемая согласно (15.21), представляет собой отношение двух независимых оценок одной и той же дисперсии. Английский статистик Р. Фишер впервые исследовал распределение этого отношения, однако им было получено распределение не F, а величины $z=(1/2)\ln F$. Позднее Спе-

содержатся значения e, получаемые из условия $P(F < e) = 0{,}95$ для различных сочетаний чисел степеней свободы n и m [1].

В табл. 15.5 приведены значения e, соответствующие вероятности 0,99.

Б. МЕТОД НАИМЕНЬШИХ АБСОЛЮТНЫХ ОТКЛОНЕНИЙ И ЧЕБЫШЕВСКАЯ АППРОКСИМАЦИЯ [2]

15.6. Постановка задачи

Существует ряд практических задач, удовлетворительное решение которых нельзя получить с помощью метода наименьших квадратов (МНК). В этих случаях часто может быть использован метод чебышевской аппроксимации (ЧА) или связанный с ним метод наименьших абсолютных отклонений (МНАО) [3].

В случае МНАО параметры a модели

$$y = y(a, x) \tag{15.24}$$

находятся из условия

$$A = \sum_{l=1}^{N} |y(a, x^l) - \tilde{y}^l| = \min, \tag{15.25}$$

а в случае ЧА — из условия

$$T = \max_l |y(a, x^l) - \tilde{y}^l| = \min. \tag{15.26}$$

Использование методов ЧА и МНАО по сравнению с МНК связано с тремя существенными трудностями:

1) минимизируемые функции (15.25) и (15.26) недифференцируемы, вследствие чего оказывается невозможным использование классических методов оптимизации;

2) задачи (15.25) и (15.26) принципиально имеют бесконечное множество решений;

3) статистический анализ существенно усложняется по сравнению с МНК.

декор предложил перейти непосредственно к распределению величины F. В связи с этим введенное здесь F-распределение называют также распределением Снедекора. — *Прим. ред.*

[1] В тексте книги величина e, получаемая из условия $P(F < e) = 1 - \alpha$, называется критическим значением F-распределения при уровне значимости α и числах степеней свободы $\varphi_1 = n_1 = n$ и $\varphi_2 = n_2 = m$ и обозначается через $F_{кр}$. — *Прим. ред.*

[2] Данный раздел основан на работах автора, опубликованных в 1975—1976 гг. в ряде научных отчетов Высшей инженерной школы в г. Лейпциге.

[3] См. также разд. 10.4.5 и 10.4.6. — *Прим. ред.*

Эти трудности в значительной степени устраняются, если удается путем подходящего выбора экспериментальных точек x^i получить решения задач (15.25) и (15.26) в явном виде.

Прежде всего отметим, что задачи (15.25) и (15.26) с помощью уравнений

$$|y(a, x^i) - \tilde{y}^i| = \min_{v_i} \{y(a, x^i) - \tilde{y}^i + 2v_i \,|\, y(a, x^i) - \tilde{y}^i + v_i \geqslant 0, \quad v_i \geqslant 0\} \tag{15.27}$$

и

$$\max_i |y(a, x^i) - \tilde{y}^i| = \min_c \{c \,|\, c + y(a, x^i) - \tilde{y}^i \geqslant 0, \quad c - y(a, x^i) + \tilde{y}^i \geqslant 0, \quad i = 1, \ldots, N\} \tag{15.28}$$

могут быть сведены к следующим задачам оптимизации при ограничениях:

МНАО-задача:

$$A = A(a, v) = \sum_{i=1}^{N} (y(a, x^i) + 2v_i) = \min, \tag{15.29}$$

$$y(a, x^i) + v_i \geqslant \tilde{y}^i, \quad v_i \geqslant 0, \quad i = 1, \ldots, N. \tag{15.30}$$

ЧА-задача:

$$T = T(c, a) = c = \min, \tag{15.31}$$

$$c + y(a, x^i) \geqslant \tilde{y}^i, \quad c - y(a, x^i) \geqslant -\tilde{y}^i, \quad i = 1, 2, \ldots, N. \tag{15.32}$$

В общем случае это задачи невыпуклого нелинейного программирования. Однако, используя их специальную структуру, путем подходящего выбора плана эксперимента можно в определенных случаях получить полное и явное решение задачи. Ниже приведены некоторые результаты, относящиеся к случаю, когда модель линейная, т. е.

$$y = y(a, x) = a_0 f_0(x) + a_1 f_1(x) + \ldots + a_k f_k(x). \tag{15.33}$$

При этом задачи (15.29)—(15.31) являются задачами линейного программирования (ЛП).

МНАО-ЛП-задача [1]:

$$A = c'(Fa + 2v) = \min, \tag{15.34}$$

$$Fa + v \geqslant \tilde{Y}, \quad v \geqslant 0, \tag{15.35}$$

где

$$c = (1, 1, 1, \ldots, 1)'. \tag{15.35а}$$

ЧА-ЛП-задача:

$$T = c = \min, \tag{15.36}$$

$$cc + Fa \geqslant \tilde{Y}, \quad cc - Fa \geqslant -\tilde{Y}. \tag{15.37}$$

[1] Здесь F — матрица значений функций $f_l(x)$ в N опытах размера $N \times (k+1)$; a — вектор-столбец размера $(k+1)$; v — вектор-столбец размера N; $\tilde{Y}$ — вектор-столбец результатов наблюдений. — *Прим. ред.*

15.7. Линейные модели и симметричные планы

Для линейной модели

$$y = y(a_0, a_1, x) = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n = a_0 + a'x, \quad (15.38)$$

где

$$a = (a_1, a_2, \ldots, a_n)', \quad (15.39)$$

в случае когда план D объема $N = 2n$ является симметричным, т. е.

$$D: x^i = -x^{i+n}, \quad i = 1, 2, \ldots, n, \quad (15.40)$$

задачи линейного программирования (15.33)—(15.37) принимают следующий вид:
МНАО-задача:

$$A = 2na_0 + 2\sum_{i=1}^{n}(v_i + v_{i+n}) = \min! \quad (15.41)$$

$$ea_0 + Xa + v^1 \geqslant \tilde{Y}_1, \quad v^1 \geqslant 0, \quad (15.42)$$
$$ea_0 - Xa + v^2 \geqslant \tilde{Y}_2, \quad v^2 \geqslant 0.$$

ЧА-задача:

$$T = c = \min! \quad (15.43)$$

$$ec + ea_0 + Xa \geqslant \tilde{Y}_1, \quad ec + ea_0 - Xa \geqslant \tilde{Y}_2, \quad (15.44)$$
$$ec - ea_0 - Xa \geqslant -\tilde{Y}_1, \quad ec - ea_0 + Xa \geqslant -\tilde{Y}_2.$$

При этом предполагается, что матрица

$$X = (x_{ij}) = (x_j^i), \qquad i = 1, 2, \ldots, n, \quad (15.45)$$

невырожденная; кроме того, введены обозначения

$$\tilde{Y}_1 = (\tilde{y}^1, \ldots, \tilde{y}^n)', \quad \tilde{Y}_2 = (\tilde{y}^{n+1}, \ldots, \tilde{y}^{2n})', \quad (15.46)$$
$$v^1 = (v_1, \ldots, v_n)', \quad v^2 = (v_{n+1}, \ldots, v_{2n})'. \quad (15.47)$$

В случае когда план симметричный объема $N = 2n + 1$

$$D: x^0 = 0, \ x^i = -x^{i+n}, \ i = 1, 2, \ldots, n$$

целевая функция (15.41) записывается следующим образом:

$$A = (2n + 1)a_0 + 2v_0 + 2\sum_{i=1}^{n} v_i + v_{i+n}. \quad (15.41')$$

При этом к ограничениям (15.42) добавляется условие

$$a_0 + v_0 \geqslant \tilde{y}^0, \ v_0 \geqslant 0, \quad (15.42')$$

а к ограничениям (15.44) — условия

$$c + a_0 \geqslant \tilde{y}^0, \ c - a_0 \geqslant -\tilde{y}^0. \quad (15.44')$$

Если в качестве плана эксперимента используются специальные планы D_0 или $\tilde{D}_0$, в которых

$$x^l = e^l \quad (e^l — l\text{-й единичный вектор})^{1)}, \tag{15.48}$$

а для параметров модели вместо (a_0, a) вводится обозначение (b_0, b), то получаем следующие специальные оптимизационные задачи:
СМНАО-задача:

$$A = \begin{cases} 2nb_0 + 2\sum_{l=1}^{n}(v_l + v_{l+n}) = \min! \text{ (для } D_0), \\ 2(n+1)b_0 + 2v_0 + \sum_{l=1}^{n}(v_l + v_{l+n}) = \min! \text{ (для } \tilde{D}_0), \end{cases} \tag{15.49}$$

$$\left.\begin{array}{ll} b_0 + b_l + v_l \geqslant \tilde{y}^l, & v_l \geqslant 0, \\ b_0 - b_l + v_{l+n} \geqslant \tilde{y}^{l+n}, & v_{l+n} \geqslant 0, \end{array}\right\} l = 1, 2, \ldots, n, \tag{15.50}$$

$$b_0 + v_0 \geqslant \tilde{y}^0, \; v_0 \geqslant 0 \text{ (для } \tilde{D}_0). \tag{15.51}$$

СЧА-задача:

$$T = c = \min!, \tag{15.52}$$

$$\left.\begin{array}{ll} c + b_0 + b_l \geqslant \tilde{y}^l, & c + b_0 - b_l \geqslant \tilde{y}^{l+n}, \\ c - b_0 - b_l \geqslant -\tilde{y}^l, & c - b_0 + b_l \geqslant -\tilde{y}^{l+n}, \end{array}\right\} i = 1, 2, \ldots, n \tag{15.53}$$

$$c + b_0 \geqslant \tilde{y}^0, \; c - b_0 \geqslant -\tilde{y}^0 \text{ (для } \tilde{D}_0). \tag{15.54}$$

Справедливо следующее утверждение:

Утверждение 1

Пусть M_0 — множество всех решений (b_0, b) специальной задачи метода наименьших абсолютных отклонений (СМНАО) или специальной задачи чебышевской аппроксимации (СЧА). Тогда множество всех решений (a_0, a) МНАО- или соответственно ЧА-задачи определяется выражениями

$$a_0 = b_0, \; a = X^{-1}b, \; (b_0, b) \in M_0. \tag{15.55}$$

В силу этого ниже рассматриваются решения только специальных задач аппроксимации.

---

1) Здесь $e^l = (e_1^l, e_2^l, \ldots, e_l^l, \ldots, e_n^l)^*$;

$$e_l^l = \begin{cases} 0, & l \neq j, \\ 1, & l = j. \end{cases} — \textit{Прим. ред.}$$

15.8. Решение специальных задач

Путем одновременного рассмотрения прямых задач (15.49)—(15.54) и соответствующих им двойственных задач удалось для СМНАО- и СЧА-задач получить полные и явные решения.

Обозначим

$$\tilde{\tilde{y}}^i = \tilde{y}^i + \tilde{y}^{i+n}, \qquad i = 1, 2, \ldots, n. \tag{15.56}$$

Упорядочим величины $\tilde{\tilde{y}}^i$ следующим образом:

$$\tilde{\tilde{y}}^{i_1} \geqslant \tilde{\tilde{y}}^{i_2} \geqslant \ldots \geqslant \tilde{\tilde{y}}^{i_n}. \tag{15.57}$$

В дальнейшем будем различать следующие случаи:

МНАО-D_0-1:	$n = 2k+1$,
МНАО-D_0-2:	$n = 2k$,
МНАО-$\tilde{D}_0$-1:	$n = 2k+1$,
МНАО-$\tilde{D}_0$-2:	$n = 2k$,
МНАО-$\tilde{D}_0$-2.1:	$2\tilde{y}^0 \geqslant \tilde{\tilde{y}}^{i_k}$,
МНАО-$\tilde{D}_0$-2.2:	$\tilde{\tilde{y}}^{i_k} \geqslant 2\tilde{y}^0 \geqslant \tilde{\tilde{y}}^{i_{k+1}}$,
МНАО-$\tilde{D}_0$-2.3:	$\tilde{\tilde{y}}^{i_{k+1}} \geqslant 2\tilde{y}^0$,
ЧА-D_0:	нет особых условий,
ЧА-$\tilde{D}_0$-1:	$2\tilde{y}^0 \geqslant \tilde{\tilde{y}}^{i_1}$,
ЧА-$\tilde{D}_0$-2:	$\tilde{\tilde{y}}^{i_1} \geqslant 2\tilde{y}^0 \geqslant \tilde{\tilde{y}}^{i_n}$,
ЧА-$\tilde{D}_0$-3:	$\tilde{\tilde{y}}^{i_n} \geqslant 2\tilde{y}^0$.

Определим индексы i_k, i_l и i_m с помощью соотношений

$$2\tilde{\tilde{y}}^{i_k} \geqslant (\tilde{\tilde{y}}^{i_1} + \tilde{\tilde{y}}^{i_n}) \geqslant 2\tilde{\tilde{y}}^{i_{k+1}} \quad (\text{ЧА-}D_0,\ \text{ЧА-}\tilde{D}_0\text{-2}), \tag{15.58}$$

$$2\tilde{\tilde{y}}^{i_l} \geqslant 2\tilde{y}^0 + \tilde{\tilde{y}}^{i_n} \geqslant 2\tilde{\tilde{y}}^{i_{l+1}} \quad (\text{ЧА-}\tilde{D}_0\text{-1}), \tag{15.59}$$

$$2\tilde{\tilde{y}}^{i_m} \geqslant 2\tilde{y}^0 + \tilde{\tilde{y}}^{i_1} \geqslant 2\tilde{\tilde{y}}^{i_{m+1}} \quad (\text{ЧА-}\tilde{D}_0\text{-3}). \tag{15.60}$$

Множество решений СМНАО- и соответственно СЧА-задач определяется для указанных выше случаев неравенствами:
Для МНАО во всех случаях

$$b_0 - \tilde{y}^{i+n} \leqslant b_i \leqslant \tilde{y}^i - b_0, \quad i = i_1, \ldots, i_k, \tag{15.61}$$

$$\tilde{y}^i - b_0 \leqslant b_i \leqslant b_0 - \tilde{y}^{i+n}, \quad i = i_{k+1}, \ldots, i_n, \tag{15.62}$$

где

$$b_0 = \frac{1}{2}\tilde{\tilde{y}}^{l_{k+1}} \text{ (МНАО-D}_0\text{-1, МНАО-}\tilde{\text{D}}_0\text{-1, МНАО-}\tilde{\text{D}}_0\text{-2,3)}, \quad (15.63)$$

$$\frac{1}{2}\tilde{\tilde{y}}^{l_{k+1}} \leqslant b_0 \leqslant \frac{1}{2}\tilde{\tilde{y}}^{l_k} \text{ (МНАО-D}_0\text{-2)}, \quad (15.64)$$

$$b_0 = \frac{1}{2}\tilde{\tilde{y}}^{l_k} \text{ (МНАО-}\tilde{\text{D}}_0\text{-2.1)}, \quad (15.65)$$

$$b_0 = \tilde{y}^0 \text{ (МНАО-}\tilde{\text{D}}_0\text{-2.2)}. \quad (15.66)$$

Для задач ЧА для различных случаев получаем следующие результаты:

ЧА-D$_0$/ЧА-$\tilde{\text{D}}_0$-2:

$$b_0 = \frac{1}{4}(\tilde{\tilde{y}}^{l_1} + y^{l_m}), \quad b_l = \frac{1}{2}(\tilde{y}^l - \tilde{y}^{l+n}), \qquad l = i_1, \ldots, i_n, \quad (15.67)$$

$$\tilde{y}^l - \frac{1}{2}\tilde{\tilde{y}}^{l_1} \leqslant b_l \leqslant \frac{1}{2}\tilde{\tilde{y}}^{l_1} - \tilde{y}^{l+n}, \qquad l = i_2, \ldots, i_k, \quad (15.68)$$

$$\frac{1}{2}\tilde{\tilde{y}}^{l_n} - \tilde{y}^{l+n} \leqslant b_l \leqslant \tilde{y}^l - \frac{1}{2}\tilde{\tilde{y}}^{l_n}, \qquad l = i_{k+1}, \ldots, i_{n+1}. \quad (15.69)$$

ЧА-$\tilde{\text{D}}_0$-1:

$$b_0 = \frac{1}{2}\tilde{y}^0 + \frac{1}{4}\tilde{\tilde{y}}^{l_n}, \; b_{l_n} = \frac{1}{2}(\tilde{y}^{l_n} - \tilde{y}^{l_n+n}), \quad (15.70)$$

$$\tilde{y}^l - \tilde{y}^0 \leqslant b_l \leqslant \tilde{y}^0 - \tilde{y}^{l+n} \quad l = i_1, \ldots, i_t, \quad (15.71)$$

$$\frac{1}{2}\tilde{\tilde{y}}^{l_n} - \tilde{y}^{l+n} \leqslant b_l \leqslant \tilde{\tilde{y}}^l - \frac{1}{2}\tilde{\tilde{y}}^{l_n}, \quad l = i_{t+1}, \ldots, i_{n-1}. \quad (15.72)$$

ЧА-$\tilde{\text{D}}_0$-3:

$$b_0 = \frac{1}{2}\tilde{y}^0 + \frac{1}{4}\tilde{\tilde{y}}^{l_1}, \quad b_{l_1} = \frac{1}{2}(\tilde{y}^{l_1} - \tilde{y}^{l_1+n}), \quad (15.73)$$

$$\tilde{y}^l - \frac{1}{2}y^{l_1} \leqslant b_l \leqslant \frac{1}{2}\tilde{\tilde{y}}^{l_1} - \tilde{y}^{l+n}, \; l = i_2, \ldots, i_m, \quad (15.74)$$

$$\tilde{y}^0 - \tilde{y}^{l+n} \leqslant b_l \leqslant \tilde{y}^l - \tilde{y}^0, \quad l = i_{m+1}, \ldots, i_n. \quad (15.75)$$

Решения, при которых b_l выбираются на концах интервалов, будем называть граничными решениями (Ecklösungen). Решения, при которых b_l выбираются в середине интервалов, будем называть центральными (Zentrumlösungen). Для центральных решений

$$b_l = \frac{1}{2}(\tilde{y}^l - \tilde{y}^{l+n}), \quad l = 1, 2, \ldots, n. \quad (15.76)$$

Исключение составляет величина b_0, которая в рассмотренных случаях выбирается различным образом, согласно выражениям (15.63), (15.65)—(15.67), (15.70), (15.73), а также по формуле

$$b_0 = \frac{1}{4}(\tilde{\tilde{y}}^{l_k} + \tilde{\tilde{y}}^{l_{k+1}}) \quad \text{(для МНАО-D}_0\text{-2)}. \quad (15.77)$$

15.9. Влияние ошибки в наблюдениях

Рассмотрим случай, когда все наблюдения $\tilde{y}^t$ содержат ошибку одной и той же величины, т. е.

$$F(\tilde{y}^t) = F, \qquad t = 0, 1, \ldots, 2n. \tag{15.78}$$

При этом легко найти ошибку в коэффициентах b_l для случая, когда используются граничные решения задачи. Имеем

$$F(b_l) = F \tag{15.79}$$

для

$t = 0$: во всех случаях,
$t = i_{k+1}$: (МНАО-D_0-1/D_0-2/$\tilde{D}_0$-1/$\tilde{D}_0$-2.3),
$t = i_k$: (МНАО-D_0-2/$\tilde{D}_0$-2.1),
$t = i_1$: (ЧА-D_0/$\tilde{D}_0$-2/$\tilde{D}_0$-3),
$t = i_n$: (ЧА-D_0/$\tilde{D}_0$-1/$\tilde{D}_0$-2).

В остальных случаях

$$F(b_l) = 2F. \tag{15.80}$$

Можно провести следующее высказывание, устанавливающее оптимальные свойства планов D_0 и $\tilde{D}_0$ при наличии ошибки в наблюдениях.

Утверждение 2

Среди всех симметричных планов D и соответственно $\tilde{D}$ с одинаковыми величинами

$$\sum_{i=1}^{n} (x_j^i)^2 = 1, \qquad j = 1, 2, \ldots, n, \tag{15.81}$$

только для планов D_0 и соответственно $\tilde{D}_0$ выполняется следующее условие оптимальности:

$$\frac{F(a_l)}{F} = \min l, \qquad i = 0, 1, \ldots, n. \tag{15.82}$$

15.10. Несмещенность и эффективность

Положим, что наблюдения $\tilde{y}^t$ являются случайными величинами. В качестве оценок $\hat{b}_l$ параметров модели b_l будем рассматривать решения, приведенные в разд. 15.8. Обсудим вопрос о несмещенности и эффективности этих оценок.

Сделаем следующие допущения:

1) модель (15.38) адекватна;

2) все $\tilde{y}^t$ независимы и распределены симметрично относительно

своих средних $E\{\tilde{y}^l\} = m_l$, причем дисперсии всех y^l одинаковы, т. е. $D\{\tilde{y}^l\} = \sigma^2$.

В центре $x^0 = 0$ плана $\tilde{D}_0$ произведем два независимых наблюдения $\tilde{y}^{01}$ и $\tilde{y}^{02}$ и положим

$$2\tilde{y}^0 = \tilde{\tilde{y}}^0 = \tilde{y}^{01} + \tilde{y}^{02}. \tag{15.83}$$

Тогда все $\tilde{\tilde{y}}^l$ будут иметь одну и ту же функцию плотности распределения $f(x)$ и функцию распределения $F(x)$. В формулах (15.57)—(15.60) для оценок $\hat{b}_l$ входят упорядоченные величины $\tilde{\tilde{y}}^l$. Статистический анализ решения можно провести, основываясь на следующем утверждении:

Утверждение 3

Пусть X^l, $i = 1, 2, \ldots, n$, суть независимые, одинаково распределенные случайные величины с плотностью распределения $f(x)$ и функцией распределения $F(x)$. Тогда плотность распределения $f_k(x)$ для величины X^{l_k}, имеющей k-й ранг в упорядоченной последовательности

$$X^{l_1} > X^{l_2} > \ldots > X^{l_n}, \tag{15.84}$$

имеет вид

$$f_k(x) = \frac{n!}{(k-1)!\,(n-k)!} F(x)^{n-k} (1 - F(x))^{k-1} f(x). \tag{15.85}$$

Исходя из этого утверждения, можно получить следующие результаты:

Утверждение 4

В случае МНАО все оценки являются асимптотически несмещенными. В методе ЧА оценки этим свойством не обладают.

Утверждение 5

При использовании МНАО оценки получаются несмещенными, если в случае МНАО-D_0-2 в качестве $\hat{b}_0$ выбирается середина интервала, т. е.

$$\hat{b}_0 = \frac{1}{4} (\tilde{\tilde{y}}^{l_k} + \tilde{\tilde{y}}^{l_{k+1}}). \tag{15.86}$$

Утверждение 6

При использовании ЧА центральные решения дают несмещенные оценки.

Утверждение 7

Во всех случаях центральные решения дают эффективные оценки, причем

$$D(\hat{b}_0) = \begin{cases} \sigma^2/4 \text{ для ЧА и МНАО-}D_0\text{-2}, \\ \sigma^2/2 \text{ в остальных случаях}, \end{cases} \tag{15.87}$$

$$D(\hat{b}_i = \sigma^2/2, \quad i = 1, 2, \ldots, n. \tag{15.88}$$

ЛИТЕРАТУРА

1. Смирнов В. И., Курс высшей математики, т. III, ч. 1, изд. 6, Гостехиздат, М., 1954.
2. Dietrich G., Stahl H., Matrizen und Determinanten und ihre Anwendung in der Technik und Ökonomie, 3. Aufl. Fachbuchverlag, Leipzig, 1970.
3. Kochandörffer R., Determinanten und Matrizen, Teubner, Leipzig, 1957.
4. Смирнов В. И., Курс высшей математики, т. I, изд. 9, ГОНТИ, Л.—М., 1948.
5. Müller P. N. (Hrsg.), Wahrscheinlichkeitsrechnung und matematische Statistik, Lexikon, Akademie-Verlag, Berlin, 1970.
6. Fisz M., Rachunek prawdopodobeństwa i statystyka matematyczna, Biblioteka Matematyczna, t. 18, Państwowe Wydawnictwo Naukowe, Warszawa, 1958.
7. Гнеденко Н. В., Курс теории вероятностей, Физматгиз, М., 1961.
8. Смирнов Н. В., Дунин-Барковский Н. В., Курс теории вероятностей и математической статистики (для технических приложений), изд-во «Наука», М., 1969.
9. Свешников А. А., Прикладные методы теории случайных функций, изд-во «Наука», М., 1968.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Гантмахер Ф. Р., Теория матриц, изд-во «Наука», М., 1967.
2. Ефимов Н. В., Квадратичные формы и матрицы, изд-во «Наука», М., 1967.
3. Корн Г., Корн Т., Справочник по математике для научных работников и инженеров, изд-во «Наука», М., 1968.
4. Большев Л. Н., Смирнов Н. В., Таблицы математической статистики, изд-во «Наука», М., 1965.
5. Фадеев Д. К., Фадеева В. Н., Вычислительные методы линейной алгебры, Физматгиз., М., 1960.
6. Вентцель Е. С., Теория вероятностей, Физматгиз, М., 1962.
7. Хальд А., Математическая статистика с техническими приложениями, ИЛ, М., 1956.
8. Езекиэл М., Фокс К., Методы анализа корреляций и регрессий линейных и криволинейных, изд-во «Статистика», М., 1966.

Таблица 15.1

Нормированное нормальное распределение

$$\varphi(x),\ x \geqslant 0;\ \varphi(-x) = \varphi(x);\ \overline{\Phi}(x) = \int_0^x \varphi(x)\,dx;\ x \geqslant 0,$$

$$\Phi(x) = 0{,}5 + \overline{\Phi}(x);\ \Phi(-x) = 0{,}5 - \overline{\Phi}(x).$$

x	$\varphi(x)$	$\bar{\Phi}(x)$	x	$\varphi(x)$	$\bar{\Phi}(x)$
0,00	0,3989	0,000000	1,50	0,1295	0,433193
0,05	0,3984	0,019939	1,55	0,1200	0,439429
0,10	0,3970	0,039828	1,60	0,1109	0,445201
0,15	0,3945	0,059618	1,65	0,1023	0,450528
0,20	0,3910	0,079260	1,70	0,0940	0,455434
0,25	0,3867	0,098706	1,75	0,0863	0,459941
0,30	0,3814	0,117911	1,80	0,0790	0,464070
0,35	0,3752	0,136831	1,85	0,0721	0,467843
0,40	0,3683	0,155422	1,90	0,0656	0,471283
0,45	0,3605	0,173645	1,95	0,0596	0,474412
0,50	0,3521	0,191463	2,00	0,0540	0,477250
0,55	0,3429	0,208840	2,05	0,0488	0,479818
0,60	0,3332	0,225747	2,10	0,0440	0,482136
0,65	0,3230	0,242154	2,15	0,0396	0,484222
0,70	0,3123	0,258035	2,20	0,0355	0,486097
0,75	0,3011	0,273373	2,25	0,0317	0,487776
0,80	0,2897	0,288145	2,30	0,0283	0,489276
0,85	0,2780	0,302338	2,35	0,0252	0,490613
0,90	0,2661	0,315940	2,40	0,0224	0,491802
0,95	0,2541	0,328944	2,45	0,0198	0,492857
1,00	0,2420	0,341345	2,50	0,0175	0,493790
1,05	0,2299	0,353141	2,55	0,0154	0,494614
1,10	0,2179	0,364334	2,60	0,0136	0,495339
1,15	0,2059	0,374928	2,65	0,0119	0,495975
1,20	0,1942	0,384930	2,70	00,104	0,496533
1,25	0,1826	0,394350	2,75	0,0091	0,497020
1,30	0,1714	0,403200	2,80	0,0079	0,497445
1,35	0,1604	0,411492	2,85	0,0069	0,497814
1,40	0,1497	0,419243	2,90	0,0060	0,498134
1,45	0,1394	0,426471	2,95	0,0051	0,498411

Таблица 15.2

χ^2-распределение

X — случайная величина, распределенная по закону χ^2 с $\varphi = n$ степенями свободы. Таблица содержит значения e, полученные из условия $P(X < e) = 1 - \alpha$

n \ $1-\alpha$	0,99	0,95	0,90	0,80	0,50	0,20
1	6,35	3,841	2,706	1,642	0,455	0,064
2	9,210	5,991	4,005	3,219	1,386	0,446
3	11,345	7,815	6,251	4,642	2,366	1,005
4	13,277	9,488	7,779	5,989	3,357	1,649
5	15,086	11,070	9,236	7,289	4,351	2,343
6	16,812	12,592	10,645	8,558	5,348	3,070
7	18,475	14,067	12,017	9,803	6,346	3,822
8	20,090	15,507	13,362	11,030	7,344	4,594
9	21,666	16,919	14,684	12,242	8,343	5,380
10	23,209	18,307	15,987	13,442	9,342	6,179
11	24,725	19,675	17,275	14,631	10,341	6,989
12	26,217	21,026	18,549	15,812	11,340	7,807
13	27,688	22,362	19,812	16,985	12,340	8,634
14	29,141	23,685	21,064	18,151	13,339	9,467
15	30,578	24,996	22,307	19,311	14,339	10,307
16	32,000	26,296	23,542	20,465	15,338	11,152
17	33,409	27,587	24,769	21,615	16,338	12,002
18	348,05	28,869	25,989	22,760	17,338	12,857
19	36,191	30,144	27,204	23,900	18,338	13,716
20	37,566	31,410	28,412	25,038	19,337	14,578
21	38,932	32,671	29,615	26,171	20,337	15,445
22	40,289	33,924	30,813	27,301	21,337	16,314
23	41,638	35,172	32,007	28,429	22,337	17,187
24	42,980	36,415	33,196	29,553	23,337	18,062
25	44,314	37,652	34,382	30,675	24,337	18,940
26	45,642	38,885	35,563	31,795	25,336	19,820
27	46,963	40,113	36,741	32,912	26,336	20,703
28	48,278	41,337	37,916	34,027	27,336	21,588
29	49,588	42,557	39,087	35,139	28,336	22,475
30	50,892	43,773	40,256	36,250	29,336	23,364

Таблица 15.3

t-распределение

T — случайная величина, распределенная по закону Стьюдента (t-распределение) с числом степеней свободы $\varphi = n$. Таблица содержит значения ε, получаемые из условия $P(|T| < \varepsilon) = 1 - \alpha$

n \ $1-\alpha$	0,99	0,95	0,90	0,80	0,50	0,20
1	63,657	12,706	6,314	3,078	0,727	0,325
2	9,925	4,303	2,920	1,886	9,617	0,289
3	5,841	3,182	2,353	1,638	0,584	0,277
4	4,604	2,776	2,132	1,533	0,569	0,271
5	4,032	2,571	2,015	1,476	0,559	0,267
6	3,707	2,447	1,943	1,440	0,553	0,265
7	3,499	2,365	1,895	1,415	0,549	0,263
8	3,355	2,306	1,860	1,397	0,546	0,262
9	3,250	2,262	1,833	1,383	0,543	0,261
10	3,169	2,228	1,812	1,372	0,542	0,260
11	3,106	2,201	1,796	1,363	0,540	0,260
12	3,055	2,179	1,782	1,356	0,539	0,259
13	3,012	2,160	1,771	1,350	0,538	0,259
14	2,977	2,145	1,761	1,345	0,537	0,258
15	2,947	2,131	1,753	1,341	0,536	0,258
16	2,921	2,120	1,746	1,337	0,535	0,258
17	2,898	2,110	1,740	1,333	0,534	0,257
18	2,878	2,101	1,734	1,330	0,534	0,257
19	2,861	2,093	1,729	1,328	0,533	0,257
20	2,845	2,086	1,725	1,325	0,533	0,257
21	2,831	2,080	1,721	1,323	0,532	0,257
22	2,819	2,074	1,717	1,321	0,532	0,256
23	2,807	2,069	1,714	1,319	0,532	0,256
24	2,797	2,064	1,711	1,318	0,531	0,256
25	2,787	2,060	1,708	1,316	0,531	0,256
26	2,779	2,056	1,706	1,315	0,531	0,256
27	2,771	2,052	1,703	1,314	0,531	0,256
28	2,763	2,048	1,701	1,313	0,530	0,256
29	2,756	2,045	1,699	1,311	0,530	0,256
30	2,750	2,042	1,697	1,310	0,530	0,256
40	2,704	2,021	1,684	1,303	0,529	0,255
60	2,660	2,000	1,671	1,296	0,527	0,254
120	2,617	1,980	1,658	1,289	0,526	0,254
	2,576	1,960	1,645	1,282	0,524	0,253

95%-ное F-распределение

F — случайная величина, подчиненная F-распределению с числом степеней свободы $\varphi_1 = n_1$ для числителя и $\varphi_2 = n_2$ для знаменателя. Таблица содержит значения e, получаемые из условия $P(F < e) = 0{,}95$

n_2 \ n_1	1	2	3	4	5	6	7	8
1	161	200	216	225	230	234	237	239
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37
3	10,13	9,55	9,28	9,12	9,01	8,94	8,49	8,85
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34

n_2 \ n_1	1	2	3	4	5	6	7	8
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15
48	4,04	3,19	2,80	2,57	2,41	2,30	2,21	2,14
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13
55	4,02	3,16	2,78	2,54	2,38	2,27	2,18	2,11
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98
400	3,86	3,02	2,62	2,39	2,23	2,12	2,03	1,96
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94

Таблица 15.4 (продолжение)

n_2 \ n_1	9	10	11	12	14	16	20	24
1	241	242	243	244	245	246	248	249
2	19,38	19,39	19,40	19,41	19,42	19,43	19,44	19,45
3	8,81	8,79	8,76	8,74	8,71	8,69	8,66	8,64
4	6,00	5,96	5,94	5,91	5,87	5,84	5,80	5,77
5	4,77	4,74	4,70	4,28	4,64	4,60	4,56	4,53
6	4,10	4,06	4,03	4,00	3,96	3,92	3,87	3,84
7	3,68	3,64	3,60	3,57	3,53	3,49	3,44	3,41
8	3,39	3,35	3,31	3,28	3,24	3,20	3,15	3,12
9	3,18	3,14	3,10	3,07	3,03	2,99	2,93	2,90
10	3,02	2,98	2,94	2,91	2,86	2,83	2,77	2,74
11	2,90	2,85	2,82	2,79	2,74	2,70	2,65	2,61
12	2,80	2,75	2,72	2,69	2,64	2,60	2,54	2,51
13	2,71	2,67	2,63	2,60	2,55	2,51	2,46	2,42
14	2,65	2,60	2,57	2,53	2,48	2,44	2,39	2,35
15	2,59	2,54	2,51	2,48	2,42	2,38	2,33	2,29
16	2,54	2,49	2,46	2,42	2,37	2,33	2,28	2,24
17	2,49	2,45	2,41	2,38	2,33	2,29	2,23	2,19
18	2,46	2,41	2,37	2,34	2,29	2,25	2,19	2,15
19	2,42	2,38	2,34	2,31	2,26	2,21	2,15	2,11
20	2,39	2,35	2,31	2,28	2,22	2,18	2,12	2,08
21	2,37	2,32	2,28	2,25	2,20	2,16	2,10	2,05
22	2,34	2,30	2,26	2,23	2,17	2,13	2,07	2,03
23	2,32	2,27	2,24	2,20	2,15	2,11	2,05	2,00
24	2,30	2,25	2,22	2,18	2,13	2,09	2,03	1,98
25	2,28	2,24	2,20	2,16	2,11	2,07	2,01	1,96

n_2 \ n_1	9	10	11	12	14	16	20	24
26	2,27	2,22	2,18	2,15	2,10	2,05	1,99	1,95
27	2,25	2,20	2,16	2,13	2,08	2,04	1,97	1,93
28	2,24	2,19	2,15	2,12	2,06	2,02	1,96	1,91
29	2,22	2,18	2,14	2,10	2,05	2,01	1,94	1,90
30	2,21	2,16	2,13	2,09	2,04	1,99	1,93	1,89
32	2,19	2,14	2,10	2,07	2,01	1,97	1,91	1,86
34	2,17	2,12	2,08	2,05	1,99	1,95	1,89	1,84
36	2,15	2,11	2,07	2,03	1,98	1,93	1,87	1,82
38	2,14	2,09	2,05	2,02	1,96	1,92	1,85	1,81
40	2,12	2,08	2,04	2,00	1,95	1,90	1,84	1,79
42	2,11	2,06	2,03	1,99	1,93	1,89	1,83	1,78
44	2,10	2,05	2,01	1,98	1,92	1,88	1,81	1,77
46	2,09	2,04	2,00	1,97	1,91	1,87	1,80	1,76
48	2,08	2,03	1,99	1,96	1,90	1,86	1,79	1,75
50	2,07	2,03	1,99	1,95	1,89	1,85	1,78	1,74
55	2,06	2,01	1,97	1,93	1,88	1,83	1,76	1,72
60	2,04	1,99	1,95	1,92	1,86	1,82	1,75	1,70
65	2,03	1,98	1,94	1,90	1,85	1,80	1,73	1,69
70	2,02	1,97	1,93	1,89	1,84	1,79	1,72	1,67
80	2,00	1,95	1,91	1,88	1,82	1,77	1,70	1,65
100	1,97	1,93	1,89	1,85	1,79	1,75	1,68	1,63
125	1,96	1,91	1,87	1,83	1,77	1,72	1,65	1,60
150	1,94	1,89	1,85	1,82	1,76	1,71	1,64	1,59
200	1,93	1,88	1,84	1,80	1,74	1,69	1,62	1,57
400	1,90	1,85	1,81	1,78	1,72	1,67	1,60	1,54
1000	1,89	1,84	1,80	1,76	1,70	1,65	1,58	1,53
∞	1,88	1,83	1,79	1,75	1,69	1,64	1,57	1,52

Таблица 15.4 (продолжение)

n_2 \ n_1	30	40	50	75	100	200	500	∞
1	250	251	252	253	253	254	254	254
2	19,46	19,47	19,48	19,48	19,49	19,49	19,50	19,50
3	8,62	8,59	8,58	8,57	8,55	8,54	8,53	8,53
4	5,75	5,72	5,70	5,68	5,66	5,65	5,64	5,63
5	4,50	4,46	4,44	4,42	4,41	4,39	4,37	4,36
6	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67
7	3,38	3,34	3,32	3,29	3,37	3,25	3,24	3,23
8	3,08	3,05	3,02	3,00	2,97	2,95	2,94	2,93
9	2,86	2,83	2,80	2,77	2,76	2,73	2,72	2,71
10	2,70	2,66	2,64	2,61	2,59	2,56	2,55	2,54
11	2,57	2,53	2,51	2,47	2,46	2,43	2,42	2,40
12	2,47	2,43	2,40	2,36	2,35	2,32	2,31	2,30
13	2,38	2,34	2,31	2,28	2,26	2,23	2,22	2,21
14	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
15	2,25	2,20	2,18	2,15	2,12	2,10	2,08	2,07
16	2,19	2,15	2,12	2,09	2,07	2,04	2,02	2,01
17	2,15	2,10	2,08	2,04	2,02	1,99	1,97	1,96
18	2,11	2,06	2,04	2,00	1,98	1,95	1,93	1,92
19	2,07	2,03	2,00	2,96	1,94	1,91	1,90	1,88
20	2,04	1,99	1,97	1,92	1,91	1,88	1,86	1,84
21	2,01	1,96	1,94	1,89	1,88	1,84	1,82	1,81
22	1,98	1,94	1,91	1,87	1,85	1,81	1,80	1,78
23	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76
24	1,94	1,89	1,86	1,82	1,80	1,77	1,75	1,73
25	1,92	1,87	1,84	1,80	1,78	1,75	1,73	1,71

n_2 \ n_1	30	40	50	75	100	200	500	∞
26	1,90	1,85	1,82	1,78	1,76	1,73	1,70	1,69
27	1,88	1,84	1,81	1,76	1,74	1,71	1,68	1,67
28	1,87	1,82	1,79	1,75	1,73	1,69	1,67	1,65
29	1,85	1,80	1,77	1,73	1,71	1,67	1,65	1,64
30	1,84	1,79	1,76	1,72	1,70	1,66	1,64	1,62
32	1,82	1,77	1,74	1,69	1,67	1,63	1,61	1,59
34	1,80	1,75	1,71	1,67	1,65	1,61	1,59	1,57
36	1,78	1,73	1,69	1,65	1,62	1,59	1,56	1,55
38	1,76	1,71	1,68	1,63	1,61	1,57	1,54	1,53
40	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51
42	1,73	1,68	1,65	1,60	1,57	1,53	1,51	1,49
44	1,72	1,67	1,63	1,58	1,56	1,52	1,49	1,48
46	1,71	1,65	1,62	1,57	1,55	1,51	1,48	1,46
48	1,70	1,64	1,61	1,56	1,54	1,49	1,47	1,45
50	1,69	1,63	1,60	1,55	1,52	1,48	1,46	1,44
55	1,67	1,61	1,58	1,52	1,50	1,46	1,43	1,41
60	1,65	1,59	1,56	1,50	1,48	1,44	1,41	1,39
65	1,63	1,58	1,54	1,49	1,46	1,42	1,39	1,37
70	1,62	1,57	1,53	1,47	1,45	1,40	1,37	1,35
80	1,60	1,54	1,51	1,45	1,43	1,38	1,35	1,32
100	1,57	1,52	1,48	1,42	1,39	1,34	1,31	1,28
125	1,55	1,49	1,45	1,39	1,36	1,31	1,27	1,25
150	1,53	1,48	1,44	1,37	1,34	1,29	1,25	1,22
200	1,52	1,46	1,41	1,35	1,32	1,26	1,22	1,19
400	1,49	1,42	1,38	1,32	1,28	1,22	1,16	1,13
1000	1,47	1,41	1,36	1,30	1,26	1,19	1,13	1,08
∞	1,46	1,39	1,35	1,28	1,24	1,17	1,11	1,00

99%-ное F-распределение

F — случайная величина, подчиненная F-распределению с числом степеней свободы $\varphi_1 = n_1$ для числителя и $\varphi_2 = n_2$ для знаменателя. Таблица содержит значения e, получаемые из условия $P(F < e) = 0{,}99$

n_2 \ n_1	1	2	3	4	5	6	7	8
1	4052	4999	5403	5625	5764	5859	5928	5981
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37
3	34,12	30,82	29,46	28,71	28,24	27,99	27,67	27,49
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30
14	8,86	6,51	5,56	5,04	4,70	4,46	4,28	4,14
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36
25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32

n_2 \ n_1	1	2	3	4	5	6	7	8
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26
28	7,64	5,45	4,57	4,07	3,76	3,53	3,36	3,23
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17
32	7,50	5,34	4,46	3,97	3,65	3,43	3,25	3,13
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93
48	7,20	5,08	4,22	3,74	3,43	3,20	3,04	2,91
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80
70	7,01	4,92	4,08	3,60	3,29	3,07	2,91	2,78
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66
150	6,81	4,75	3,92	3,45	3,14	2,92	2,76	2,63
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60
400	6,70	4,66	3,83	3,36	3,06	2,85	2,69	2,55
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51

Таблица 16.6 (продолжение)

n_2 \ n_1	9	10	11	12	14	16	20	24
1	6022	6056	6083	6106	6143	6169	6209	6235
2	99,39	99,40	99,41	99,42	99,43	99,44	99,45	99,46
3	27,34	27,23	27,13	27,05	26,92	26,83	26,69	26,60
4	14,66	14,55	14,45	14,37	14,25	14,15	14,02	13,93
5	10,16	10,05	9,96	9,89	9,77	9,68	9,55	9,47
6	7,98	7,87	7,79	7,72	7,60	7,52	7,39	7,31
7	6,72	6,62	6,54	6,47	6,36	6,27	6,16	6,07
8	5,91	5,81	5,73	5,67	5,56	5,48	5,36	5,28
9	5,35	5,26	5,18	5,11	5,00	4,92	4,81	4,73
10	4,94	4,85	4,77	4,71	4,60	4,52	4,41	4,33
11	4,63	4,54	4,46	4,40	4,29	4,21	4,10	4,02
12	4,39	4,30	4,22	4,16	4,05	3,97	3,86	3,78
13	4,19	4,10	4,02	3,96	3,86	3,78	3,66	3,59
14	4,03	3,94	3,86	3,80	3,70	3,62	3,51	3,43
15	3,89	3,80	3,73	3,67	3,56	3,49	3,37	3,29
16	3,78	3,69	3,62	3,55	3,45	3,37	3,26	3,18
17	3,68	3,59	3,52	3,46	3,35	3,27	3,16	3,08
18	3,60	3,51	3,43	3,37	3,27	3,19	3,08	3,00
19	3,52	3,43	3,36	3,30	3,19	3,12	3,00	2,92
20	3,46	3,37	3,29	3,23	3,13	3,05	2,94	2,86
21	3,40	3,31	3,24	3,17	3,07	2,99	2,88	2,80
22	3,35	3,26	3,18	3,12	3,02	2,94	2,83	2,75
23	3,30	3,21	3,14	3,07	2,97	2,89	2,78	2,70
24	3,26	3,17	3,09	3,03	2,93	2,85	2,74	2,66
25	3,22	3,13	3,06	2,99	2,89	2,81	2,70	2,62

n_2 \ n_1	9	10	11	12	14	16	20	24
26	3,18	3,09	3,02	2,96	2,86	2,78	2,66	2,58
27	3,15	3,06	2,99	2,93	2,82	2,75	2,63	2,55
28	3,12	3,03	2,96	2,90	2,80	2,71	2,60	2,52
29	3,09	3,00	2,93	2,87	2,77	2,69	2,57	2,49
30	3,07	2,98	2,90	2,84	2,74	2,66	2,55	2,47
32	3,02	2,93	2,86	2,80	2,70	2,62	2,50	2,42
34	2,98	2,89	2,82	2,76	2,66	2,58	2,46	2,38
36	2,95	2,86	2,79	2,72	2,62	2,54	2,43	2,35
38	2,91	2,82	2,75	2,69	2,59	2,51	2,40	2,32
40	2,89	2,80	2,73	2,66	2,56	2,48	2,37	2,29
42	2,86	2,78	2,70	2,64	2,54	2,46	2,34	2,26
44	2,84	2,75	2,68	2,62	2,52	2,44	2,32	2,24
46	2,82	2,73	2,66	2,60	2,50	2,42	2,30	2,22
48	2,80	2,72	2,64	2,58	2,48	2,40	2,28	2,20
50	2,79	2,70	2,63	2,56	2,46	2,38	2,26	2,18
55	2,75	2,66	2,59	2,53	2,43	2,34	2,23	2,15
60	2,72	2,63	2,56	2,50	2,39	2,31	2,20	2,12
65	2,69	2,61	2,53	2,47	2,37	2,29	2,18	2,09
70	2,67	2,59	2,51	2,45	2,35	2,27	2,15	2,07
80	2,64	2,55	2,48	2,42	2,31	2,23	2,12	2,03
100	2,59	2,50	2,43	2,37	2,26	2,19	2,06	1,98
125	2,55	2,50	2,40	2,33	2,23	2,15	2,03	1,94
150	2,53	2,44	2,37	2,31	2,20	2,12	2,00	1,91
200	2,50	2,41	2,34	2,27	2,17	2,09	1,97	1,88
400	2,46	2,37	2,29	2,23	2,12	2,04	1,92	1,84
1000	2,43	2,34	2,27	2,20	2,09	2,02	1,89	1,81
∞	2,41	2,32	2,25	2,18	2,08	2,00	1,88	1,79

Таблица 15.6 (продолжение)

n_2 \ n_1	30	40	50	75	100	200	500	∞
1	6261	6287	6302	6323	6334	6352	6361	6366
2	99,47	99,47	99,48	99,49	99,49	99,49	99,50	99,50
3	26,50	26,41	26,35	26,27	26,23	26,18	26,14	26,12
4	13,84	13,74	13,69	13,61	13,57	13,52	13,48	13,46
5	9,38	9,29	9,24	9,17	9,13	9,08	9,04	9,02
6	7,23	7,14	7,09	7,02	6,99	6,93	6,90	6,88
7	5,99	5,91	5,86	5,78	5,75	5,70	5,67	5,65
8	5,20	5,12	5,07	5,00	4,96	4,91	4,88	4,86
9	4,65	4,57	4,52	4,45	4,42	4,36	4,33	4,31
10	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	3,94	3,86	3,81	3,74	3,71	3,66	3,62	3,60
12	3,70	3,62	3,57	3,49	3,47	3,41	3,38	3,36
13	3,51	3,43	3,38	3,30	3,27	3,22	3,19	3,17
14	3,35	3,27	3,22	3,14	3,11	3,06	3,03	3,00
15	3,21	3,13	3,08	3,00	2,98	2,92	2,89	2,87
16	3,10	3,02	2,97	2,86	2,86	2,81	2,78	2,75
17	3,00	2,92	2,87	2,79	2,76	2,71	2,68	2,65
18	2,92	2,84	2,78	2,71	2,68	2,62	2,59	2,57
19	2,84	2,76	2,71	2,63	2,60	2,55	2,51	2,49
20	2,78	2,69	2,64	2,56	2,54	2,48	2,44	2,42
21	2,72	2,64	2,58	2,51	2,48	2,42	2,38	2,36
22	2,67	2,58	2,53	2,46	2,42	2,36	2,33	2,31
23	2,62	2,54	2,48	2,41	2,37	2,32	2,28	2,26
24	2,58	2,49	2,44	2,36	2,33	2,27	2,24	2,21
25	2,54	2,45	2,40	2,32	2,29	2,23	2,19	2,17

n_2 \ n_1	30	40	50	75	100	200	500	∞
26	2,50	2,42	2,36	2,28	2,25	2,19	2,16	2,13
27	2,47	2,38	2,33	2,25	2,22	2,16	2,12	2,10
28	2,44	2,35	2,30	2,22	2,19	2,13	2,09	2,06
29	2,41	2,33	2,27	2,19	2,16	2,10	2,06	2,03
30	2,38	2,30	2,25	2,16	2,13	2,07	2,03	2,01
32	2,34	2,25	2,20	2,12	2,08	2,02	1,98	1,96
34	2,30	2,21	2,16	2,08	2,04	1,98	1,94	1,91
36	2,26	2,17	2,12	2,04	2,00	1,94	1,90	1,87
38	2,23	2,14	2,09	2,00	1,97	1,90	1,86	1,84
40	2,20	2,11	2,06	1,97	1,94	1,87	1,83	1,80
42	2,18	2,09	2,03	1,94	1,91	1,85	1,80	1,78
44	2,15	2,06	2,01	1,92	1,89	1,82	1,78	1,75
46	2,13	2,04	1,99	1,90	1,86	1,80	1,75	1,73
48	2,12	2,03	1,97	1,88	1,84	1,78	1,73	1,70
50	2,10	2,00	1,95	1,86	1,82	1,76	1,71	1,68
55	2,06	1,96	1,91	1,82	1,78	1,71	1,67	1,64
60	2,03	1,94	1,88	1,79	1,75	1,68	1,63	1,60
65	2,00	1,90	1,85	1,76	1,72	1,65	1,60	1,56
70	1,98	1,88	1,83	1,74	1,70	1,62	1,57	1,53
80	1,94	1,85	1,79	1,70	1,66	1,58	1,53	1,49
100	1,89	1,79	1,73	1,64	1,60	1,52	1,47	1,43
125	1,85	1,75	1,69	1,59	1,55	1,47	1,41	1,37
150	1,83	1,72	1,66	1,56	1,52	1,43	1,38	1,33
200	1,79	1,69	1,63	1,53	1,48	1,39	1,33	1,28
400	1,74	1,64	1,57	1,47	1,42	1,32	1,24	1,19
1000	1,71	1,61	1,54	1,44	1,38	1,28	1,19	1,11
∞	1,70	1,59	1,52	1,41	1,36	1,25	1,15	1,00

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Адаптивный метод построения математической модели в условиях дрейфа 305
Аддитивный дрейф 279
— — планирование эксперимента 291
Адекватности модели проверки 356
— — — при использовании ортогонального центрального композиционного плана 148
— — — при неравномерном дублировании опытов (в случае ротатабельных планов) 156
— — — при равномерном дублировании опытов во всех точках плана 66, 131
Активный корреляционный метод идентификации 24, 495
— — — — изучающий сигнал 500
— — — — исследование точности 496
— — — — общая схема применения 495
— — — — рекомендации к применению 508
Алгоритм вычислительный, рекуррентная форма 376, 377
— рекуррентный вычисления остаточной суммы квадратов 381
— — метода наименьших квадратов 308, 379
Альтернативная гипотеза 47
Амплитудно-фазовая характеристика 485
Анализ гребневый («ридж-анализ») 425
— дискриминантный 463
— дисперсионный 209, 210
— множественный регрессионный 17, 22
Аномальные наблюдения 49
Апостериорная функция распределения 329
Аппроксимация разделяющей поверхности (АРП) 470, 477
— чебышевская 370
Априорная информация 262
Априорное моделирование 260

Балансировки задача 371
Белый шум двоичный 501
Биномиальное распределение вероятностей 517
Бокса теорема 291
Бокса — Уилсона метод 406

Варьируемая переменная (фактор) 116
Вероятность доверительная 35, 36, 37, 39, 61
Весовая(импульсная переходная) функция 484
Взаимодействие факторов 132
— — парное (первого порядка) 132
— — i-го порядка 132
Взаимодействия эффект 219, 225, 439
Взвешенный метод наименьших квадратов 328
Винера — Хопфа уравнение 488
Возмущения контролируемые 278
— неконтролируемые 278
Воспроизводимость эксперимента 344
Время памяти объекта 505
Выбор существенных переменных (в задаче классификации) 457
Выборка 28
— обучающая 456
Выборки реализация 28
— объем 39, 195
Выборочная дисперсия 30
— функция 39
— — симметричная 32
Выборочное среднее 29
Выделение существенных факторов 208
Выходная (зависимая) переменная 20, 25, 114

Гамма-функция 516
Генератор плана 126, 144
Гипотез проверка 40
— — адекватности модели 67
— — выявление аномальных наблюдений 49

Гипотез проверка для дисперсий 44
— — независимости двух случайных величин 50
— — о виде закона распределения 48
— — относительно средних 42, 43
— — отсутствия дрейфа 285
— — равенства дисперсий 45, 48
Гипотеза альтернативная 47
Гребневый анализ («ридж-анализ») 425
Греко-латинский квадрат 226

Детерминистские методы классификации 468
— — — интерполяционные методы 468
— — — методы построения разделяющей поверхности 470
Диаграмма рассеяния 245
Динамические модели 482
— характеристики объектов управления 482
Дискриминантная функция 462
Дискриминантный анализ 463
Дисперсионная матрица плана 57, 73, 118, 129, 146, 163, 202
— — — нормированная 169, 173, 178
— — — рекуррентный алгоритм вычисления 380
Дисперсионный анализ 209, 210
— — двухсторонняя (двухфакторная) классификация 215
— — иерархическая классификация 223
— — неполные классификации 224
— — однофакторная классификация 212
— — планы 224, 226, 227
— — трехфакторная классификация в латинском квадрате 225
— — четырехфакторная классификация в греко-латинском квадрате 226
Дисперсия оценок коэффициентов 60, 122, 129, 131
— выборочная 30
— неадекватности 67
— остаточная 79, 148, 381
— оценки зависимой переменной 60, 118, 121, 122, 123
— ошибки единичного наблюдения 137, 180
— ошибок наблюдения 59, 60, 61, 66
— — — оценка 74, 75, 129, 146, 154, 443
Дихотомии последовательной метод 407
Доверительная вероятность (статистическая надежность) 35, 36, 37, 39, 61
Доверительные границы для зависимой переменной 328
Доверительный интервал 35
— — для дисперсии 83
— — — зависимой переменной 328
— — — коэффициентов модели 327, 328
— — — математического ожидания при известной дисперсии 36
— — — — — при неизвестной дисперсии 37
— эллипсоид для коэффициентов модели 393
Дрейф характеристик 279
— — аддитивный 279
— — детерминированный (регулярный) 280
— — неаддитивный 279
— — непрерывный 279
— — проверка условий отсутствия 281
— — случайный 280
Дрейфующие объекты 279
— коэффициенты 280
Дробные факторные планы 126, 174, 231
— — — главные 135
— — — контраст 134
— — — — обобщающий 135
— — — полуреплика 127, 134
— — — разрешающая способность 135
— — — регулярные 234
— — — смешивание оценок коэффициентов 133
Дюамеля интеграл 484

Единичного наблюдения (эксперимента) дисперсия ошибки 180
Единичный импульс 484
— скачок 485

Зависимая (выходная) переменная 20, 52, 114
Задача балансировки 371
— краевая 348
— с начальными условиями 347
Звездные точки композиционного плана 141, 142, 201
Звездных точек плечо 141, 152, 201
Значимости квадратичных эффектов проверка 136, 157
— коэффициентов проверка 76, 139, 140, 152, 157, 159, 238, 252, 267, 384
— уровень 47
Золотое сечение 409

Идентификация объекта управления 486
— — — активный корреляционный метод 495
Иерархическая классификация (в дисперсионном анализе) 223
Изучающий сигнал 495, 500
Импульс единичный 484
Импульсная переходная (весовая) функция 484
Интегрирование когерентное 487
Интервал варьирования фактора 130, 138, 412
— доверительный 35, 36, 37, 83, 327, 328
Интерполяционные методы классификации 468
Информационная матрица плана 117, 119, 129, 137, 145, 154, 163
Информационный профиль ротатабельного плана 118, 152
Истинные значения коэффициентов 54

Каноническое преобразование поверхности отклика 422
Квадратичная модель 53, 123, 140, 198, 201
Кифера планы для квадратичной регрессии на гиперкубе 182
Классификация (дисперсионного анализа) однофакторная 212
— двухсторонняя (двухфакторная) 215
— иерархическая 223
— неполная 224
— трехфакторная 224
— четырехфакторная 226
Классификации (распознавания образов) методы 455
— — выбор существенных переменных 457
— — детерминистские 468
— — основные понятия 456
— — оценка точности классифицирующего правила 474
— — планирование эксперимента 476
— — статистические 461
Классифицирующая функция 468
Классифицирующее правило 457, 464
Ковариационная матрица оценок коэффициентов 59, 61, 114
Ковариация оценок параметров (коэффициентов) 326
Когерентное интегрирование 487
Комплексная частотная характеристика 485
Комплексный коэффициент усиления 485
Композиционные планы второго порядка 123, 141
— — — звездные точки 141, 142, 201
— — — ортогональные центральные композиционные планы 141
— — — ротатабельные центральные композиционные планы 152
— — — ядро 142, 144, 152
Конкордация 266
Коно планы для квадратичной регрессии на гиперкубе 184
Контраст плана 134
— — обобщающий 135
Контролируемые переменные (факторы) 20, 52, 278
— возмущения 278
Корреляционные методы идентификации 488
Коэффициент ранговой корреляции 263
— — — проверка значимости 264
— конкордации (согласия) 266, 268
— — проверка значимости 267, 269
— корреляции 50
Коэффициенты модели истинные 54
— — доверительные интервалы 327, 328
— — изменяющиеся во времени (дрейфующие) 278, 280
— — оценка 57, 74, 86, 88, 103, 129, 146, 155, 300, 379, 509
— — проверка значимости 76, 139, 140, 152, 157, 159, 238, 252, 384
Критерий, мощность 47
— ортогональности плана 121
— оценивания Чебышева 318, 345, 523
— — байесовский 329
— — максимального правдоподобия 34, 328
— — наименьших абсолютных отклонений 523
— — — квадратов 57, 318, 322, 345
— планирования эксперимента 120, 196
— — — А-оптимальности 122, 165
— — — D-оптимальности 122, 162, 336
— — — G-оптимальности 122
— — — в задаче классификации (распознавания образов) 478
— проверки гипотез 40, 42
— ротатабельности 121
Критическая область 40, 42, 46
Критическое значение F-распределения 46, 523, 535 (табл. 15.4, 15.5)

Критическое значение t-распределения 44, 522, 534 (табл. 15.3)
— — χ-распределения 520, 533 (табл. 15.2)
Крутого восхождения (Бокса — Уилсона) метод 149, 406, 410

Лагерра функции 24, 494, 497
— — частотная характеристика 507
Лапласа преобразование 483
— — обратное 484
— функция 61, 266, 519, 532 (табл. 15.1)
Латинский квадрат 224
— куб 227
Линейная модель 53, 122, 124, 129, 210
Линейный эффект 209
Локальные методы классификации 465
— — — Фикса-Ходжеса метод 466

Максимального правдоподобия метод 34, 328
Марквардта метод 323
Масштаб переменных натуральный 117
— — стандартизованный 117
Матрица значений функций независимых переменных 282
— ковариационная оценок коэффициентов 59, 61, 114
— моментов плана 117
— плана 115, 126
— — дисперсионная 57, 73, 118, 120, 146, 163, 202
— — — нормированная 169, 173, 178
— — информационная 117, 119, 129, 137, 145, 154, 163
— — — нормированная 169
— рангов 266, 272
— системы нормальных уравнений 117
— смещения 64
Медиана 245
Множественный регрессионный анализ 17, 22
Модели аналитические (теоретические) 10, 23, 406
— в виде системы обыкновенных дифференциальных уравнений 344
— выбор структуры (вида) 72, 78, 97
— динамические 11, 482
— дрейфующих объектов 279, 280
— квадратичные 53, 123, 140, 182, 198, 201
— кубические 189
— линейные 53, 122, 124, 129, 210
— оценивание параметров 57, 74, 86, 103, 129, 146, 155, 300, 345, 379, 509
— проверка адекватности 66, 131, 148, 156, 356, 362
— статистические (экспериментальные, эмпирические) 10, 406
Мощность критерия 47
Мысленные опыты (эксперименты) 414

Наименьших абсолютных отклонений метод 523
— квадратов метод 53, 322, 345
Насыщенный план 120, 198, 209, 210, 231
Неаддитивный дрейф 279
— — адаптивный метод построения модели 305
Неадекватности дисперсия 67
Независимые (входные) переменные 20, 52
Неконтролируемые возмущения 278
Нелинейные (по параметрам) модели 317, 397
— — оценка параметров 323
— — планирование эксперимента 336
— — — последовательное 327
— — статистический анализ 327, 330
Ненасыщенные планы 210
Неполные классификации (в дисперсионном анализе) 224
Непрерывные планы 166
— — A-оптимальные 171
— — G-оптимальные 171
— — D-оптимальные для полинома на отрезке 175
— — — для кубической регрессии на гиперкубе 189
— — — Кифера для квадратичной регрессии на гиперкубе 182
— — — Коно для квадратичной регрессии на гиперкубе 184
— — — применение при построении точных планов 191
— — — численный метод построения 185
Несмещенные оценки 33, 56
Новлану и Фюрста метод поиска 448
Нормальное распределение вероятностей 517
— — — нормированное 518
— — — плотность 518, 532 (табл. 15.1)
— — — функция Лапласа 61, 519, 532 (табл. 15.1)
Нормированный план 167

Область планирования эксперимента (варьирования переменных) 116, 162, 202
Обобщающий контраст 135
Обучающая выборка 456
— — область решений 477
Обучение распознаванию 456
Определитель дисперсионной матрицы плана, рекуррентный алгоритм вычисления 388
Опрос 261
Оптимизация экспериментальная 405
Ортогонализации методы 82
Ортогональной случайной матрицы формирование 298
Ортогональности критерий 121
Ортогональные к дрейфу планы 291
— полиномы 105, 292
— — нормированные 296
— — процедуры построения 105, 293
— функции 492
— — Лагерра 494, 497
Ортогональный план 117, 119, 137, 292
— — центральный композиционный 141
Основной уровень фактора 138
Остановки эксперимента правило 391
Остаточная дисперсия 79, 148, 381
— сумма квадратов 61, 74, 151
— — — рекуррентный алгоритм вычисления 381
Отбора функций принцип 87, 106
Отклика функция 406
Отношение правдоподобия 462
Оценка дисперсии ошибок единичного эксперимента 285
— — — наблюдения 74, 75, 129, 146, 154, 443
— наилучшая 56
— несмещенная 33, 56
— симметричная 33
— смещенная 134
— состоятельная 33
— точности классифицирующего правила 474
— эффекта 235
— эффективная 33
Ошибка первого рода 46, 47
— второго рода 46
— оценивания 59
— при выборе вида модели 63
— систематическая 369

Параметрические методы классификации 461
Парных сравнений метод 286
Пассивный эксперимент 82, 495
Передаточная функция 483
Переменная входная контролируемая (независимая) 20, 52
— выходная (зависимая, целевая величина) 20, 52, 114
— фиктивная 129, 142
Переходная функция (кривая разгона) 484, 485
Плакетта — Бермана планы 209
Планы экспериментов 80, 115
— — близкие к D-оптимальным 198
— — в задачах классификации 476
— — в производственных условиях 430
— — в условиях аддитивного дрейфа 291
— — дисперсионного анализа 210
— — для квадратичных моделей 141, 152, 182, 198
— — — моделей, описываемых полиномом первого порядка 124
— — — — содержащих линейные члены и взаимодействия различного порядка 132
— — — — нелинейных по параметрам моделей 317, 321, 336, 340, 397
— — D-оптимальные 161
— — дробные факторные 126, 174, 231
— — метода случайного баланса 241
— — насыщенные 210, 231, 234
— — непрерывные 166, 175, 182
— — нормированные 167
— — ортогональные 117, 119, 137, 292
— — — к дрейфу 291
— — — центральные композиционные 141
— — Плакетта — Бермана 234
— — полные факторные 125, 174
— — последовательные 185, 386, 397
— — ротатабельные 118
— — — униформ-планы 152
— — — центральные композиционные 152
— — сверхнасыщенные 209, 210, 240
— — типа B_n 201
— — точные (дискретные) 166, 198
— — центральные 116, 119
— — — композиционные 123, 141
Плечо звездных точек 141, 152, 201
Поверхность разделяющая 470
— — аппроксимация 470, 477
Полиномиальное распределение вероятностей 517
Полный факторный план 125, 174
Порядковая шкала 262

Последовательное планирование 185, 386
— — — для нелинейных моделей 397
Последовательной дихотомии метод 407
Последовательные методы построения математических моделей 376
Последовательный симплексный метод 430
— — — — основные понятия 430
— — — — построение начального симплекса 481
Построения разделяющей поверхности методы 470
Правило остановки эксперимента 391
Принцип отбора функций 87, 106
Проверка адекватности модели 66, 131, 148, 156, 356
— значимости квадратичных эффектов 136, 157
— — — коэффициентов 76, 139, 140, 152, 157, 159, 238, 252, 267, 384
— статистических гипотез 40, 67, 285
— условий отсутствия дрейфа 281
Программа ORFU (выбор вида модели с использованием ортогонализации) 93
— ORPOL 108
— PREDAN (статистический анализ нелинейных моделей) 330, 364
Профиль информационный ротатабельного плана 118, 152
Псевдослучайные периодические последовательности (М-последовательности) 501
— — — период 504
— — — длительность такта 504

Разделяющая поверхность 470
Разрешающая способность плана 135
Рангов матрица 266, 272
Ранговой корреляции коэффициент 263
Рандомизация эксперимента 15, 296
Ранжирование 209, 262
Распознаванию обучение 456
Распознавания образов (классификации) методы 455
Распределение вероятности, функция 28
— — биномиальное 517
— — нормальное 517, 532 (табл. 15.1)
— — полиномиальное 517
— — t-распределение Стьюдента 521, 534 (табл. 15.3)
— — F-распределение 522, 535 (табл. 15.4, 15.5)
— — χ^2-распределение 519, 533 (табл. 15.2)
Реакции фильтров выходные 491
Регрессия квадратичная 182
— кубическая 189
Регуляризации методы 490
Регулярные дробные факторные планы 234
Рекуррентная форма вычислительного алгоритма 377
— — — — для дисперсионной матрицы плана 380
— — — — — определителя дисперсионной матрицы 388
— — — — — остаточной суммы квадратов 381
— — — — — оценок коэффициентов 379
Рекуррентный алгоритм метода наименьших квадратов 308
Ротатабельный план 118
— — информационный профиль 118, 152
— — центральный композиционный 152
— униформ-план 152

Сверхнасыщенный план 209, 210, 240
Симметричная выборочная функция 32
— оценка 33
Симплекс 430
Симплексный метод оптимизации (последовательный симплексный метод) 430
Систематическая ошибка 369
Система функций Лагерра 494, 497
— — — частотные характеристики 507
Случайного баланса метод 209, 240
— поиска метод 446
— — — Нолмау и Фюрета 448
— — — слепой поиск 446
— — — случайных направлений 447
— — — с шагами в обратном направлении 448
Смешанные оценки 134
Смешивания оценок коэффициентов порядок 133
Смещения матрица 64
Собственные числа матрицы 396, 427
Состоятельные оценки 33
Спектр плана 120
Спектральная плотность 489
Статистическая (экспериментальная, эмпирическая) модель 10, 406
Статистические методы классификации 461
— — — локальные 465

Статистические методы классификации локальные Фикса — Ходжеса метод 466
— — — параметрические 461
Статическая модель 11
Стационарная область 410
Степени свободы 31, 37, 44, 45, 61, 62, 63, 210
Сумма квадратов остаточная 61, 74, 151
— — — рекуррентный алгоритм вычисления 381
Схема Хаффмана 502

Теорема Бокса 291
Теоретическая (аналитическая) модель 10, 23, 406
Точные (дискретные) планы 166
— — близкие к D-оптимальным 198
— — D-оптимальные 170
— — типа B_n 201

Уровень значимости 47
— фактора основной 138
Условия отсутствия дрейфа 281

Фактор (влияющая переменная; независимая переменная; варьируемая, управляемая переменная) 20, 116, 406
Факториал 516
Факторные планы дробные 126, 174, 231
— — полные 125, 174
Фибоначчи числа 408
Фикса — Ходжеса метод 466
Фиктивная переменная 123, 142
Функции Лагерра 24, 494, 497
Функция весовая (импульсная переходная) 484
— выборочная 29
— — симметричная 32
— Лапласа 61, 266, 519, 532 (табл. 15.1)
— отклика 406
— передаточная 483
— переходная (кривая разгона) 484, 485

Характеристики динамические 482
Хаффмана схема 502

Целевая величина 20, 21, 406
Центр (центральная точка) плана 115, 141
Центральный план 116, 119
— — композиционный 141
— — — ортогональный 141
— — — ротатабельный 152

Частота среза 506
— точки плана 167
Частотная характеристика 486
— — фильтров Лагерра 507
Чебышева критерий 318, 345, 370, 523
— полиномы 292
Четверть реплики полного факторного плана 130
Числа собственные матрицы 396, 427
— Фибоначчи 408
Число степеней свободы 31, 37, 44, 45, 61, 62, 63, 210

Шаговой регрессии метод 80
Шкала порядковая 262
Шум белый двоичный 501

Эволюционного планирования метод (ЭВОП) 438
— — — оценка дисперсии ошибок наблюдения 443
— — — проверка значимости эффектов 444
— — — рабочие листки 439
— — — циклы и фазы исследования 438
Эквидистантные планы 407
Эксперимент активный 17, 115
— пассивный 82, 495
Экспериментальная оптимизация 405
Экспертные оценки 260
Эллипсоид доверительный для коэффициентов модели 164, 393
— рассеяния 122, 161, 164
— — объем 164
Эмпирическая (статистическая, экспериментальная) модель 10, 406
— функция распределения 28
Эффект взаимодействия 219, 225, 439
— квадратичный 136, 157
— линейный 209
— фактора 439
Эффективная оценка 33

ОГЛАВЛЕНИЕ

Предисловие к русскому изданию 5
Предисловие авторов к немецкому изданию 7

1. Введение 9

2. Экспериментальные методы определения статистических характеристик и проверки гипотез 26

2.1. О задачах математической статистики 26
2.2. Выборки и выборочные функции 27
2.3. Статистическая оценка параметров 32
2.4. Статистическая проверка гипотез 40

3. Построение математической модели с помощью метода наименьших квадратов 52

3.1. Введение 52
3.2. Метод наименьших квадратов 53
3.2.1. Постановка задачи 53
3.2.2. Решение задачи 55
3.2.3. Ошибки оценивания 59
3.2.4. Ошибки при выборе вида модели 63
3.2.5. Проверка адекватности модели 66
3.3. Краткое изложение основных положений метода наименьших квадратов 72
3.3.1. Выбор структуры модели 72
3.3.2. Выбор экспериментальных точек 73
3.3.3. Проведение эксперимента 73
3.3.4. Оценивание параметров модели 74
3.3.5. Вычисление сумм квадратов и оценка дисперсий 74
3.3.5.1. Оценка дисперсии в случае, когда в каждой точке проводится один эксперимент 74
3.3.5.2. Оценка дисперсии в случае, когда в каждой точке приводится ν экспериментов 75
3.3.5.3. Вычисление суммы квадратов, характеризующей неадекватность модели (Defektquadratsumme), и значения F-критерия 75
3.3.6. Статистический анализ точности 75
3.3.6.1. Дисперсия наблюдений σ^2 известна и модель (3.74) адекватна 75
3.3.6.2. Дисперсия ошибки наблюдений σ^2 не известна и в каждой точке плана проводится один эксперимент 76
3.3.6.3. Дисперсия ошибки наблюдений σ^2 не известна и в каждой точке плана проводится несколько экспериментов 76

3.3.7. F-критерий 76
3.3.8. Проверка значимости коэффициентов 76

4. Применение методов ортогонализации при построении математических моделей 78

4.1. Выбор вида модели 78
4.2. Построение модели в случае произвольных функций и произвольного плана 81
4.2.1. Постановка задачи 81
4.2.2. Ортогонализация 82
4.2.3. Оценка параметров модели, представленной с помощью ортогональных функций 86
4.2.4. Принцип отбора функций 87
4.2.5. Оценка коэффициентов модели 88
4.2.6. Указания к программной реализации метода 93
4.2.7. Пример 93
4.3. Построение модели в случае произвольного плана эксперимента 96
4.3.1. Постановка задачи 96
4.3.2. Выбор полиномов 97
4.3.3. Указания к программной реализации метода 99
4.3.4. Пример 100
4.4. Построение полиномиальной модели на прямоугольной решетке 102
4.4.1. Постановка задачи 102
4.4.2. Вычисление коэффициентов 103
4.4.3. Ортогональные полиномы 105
4.4.4. Принцип отбора 106
4.4.5. Модель в нормальной форме 107
4.4.6. Указания к программной реализации метода 108
4.4.7. Пример 111

5. Планирование эксперимента 114

5.1. Вводные замечания 114
5.2. Основные понятия планирования эксперимента 115
5.3. Общие требования к плану эксперимента. О критериях планирования эксперимента 120
5.4. Планы для моделей, описываемых полиномами первого порядка 124
5.4.1. Вид модели 124
5.4.2. Полные факторные планы 125
5.4.3. Дробные факторные планы 126
5.4.4. Формулы для вычислений и свойства полных и дробных факторных планов для линейных моделей 129
5.5. Планы для моделей, содержащих линейные члены и взаимодействия различного порядка 132
5.5.1. Вид модели 132
5.5.2. Применение полных факторных планов для моделей типа (5.37) 132
5.5.3. Применение дробных факторных планов для модели типа (5.37) и порядок смешивания оценок коэффициентов 133
5.5.4. Вычислительные формулы и свойства планов 2^{n-p} . . . 137
5.6. Планы для квадратичных моделей 140
5.6.1. Вводные замечания 140
5.6.2. Ортогональные центральные композиционные планы . . 141
5.6.3. Ротатабельные центральные композиционные планы . . 152

6. D-оптимальные планы 161

6.1. Введение 161

6.2. Критерий D-оптимальности 162
6.3. Точные и непрерывные D-оптимальные планы 166
6.4. Некоторые свойства непрерывных D-оптимальных планов . . . 172
6.5. Непрерывные D-оптимальные планы на отрезке 175
6.6. Непрерывные D-оптимальные планы для квадратичной регрессии на гиперкубе 182
6.7. Численный метод построения непрерывных D-оптимальных планов 185
6.8. Непрерывные D-оптимальные планы для кубической регрессии на гиперкубе 189
6.9. О применении непрерывных D-оптимальных планов 191
6.10. Определение оптимального объема выборки 195
6.11. Точные планы, близкие к D-оптимальным 198

7. Выделение существенных факторов 208

7.1. Введение 208
7.2. Дисперсионный анализ 210
7.2.1. Область применения и вводные замечания 210
7.2.2. Однофакторная классификация 211
7.2.3. Двухсторонняя классификация с однократными наблюдениями 215
7.2.4. Двухсторонняя классификация с многократными наблюдениями 219
7.2.5. Двухступенчатая иерархическая классификация 223
7.2.6. Трехфакторная классификация в латинском квадрате . . 225
7.2.7. Четырехфакторная классификация в греко-латинском квадрате 226
7.2.8. Пример латинского куба 227
7.3. Применение насыщенных регулярных дробных факторных планов для выделения существенных факторов 231
7.4. Выбор существенных факторов с помощью планов Плакетта — Бермана 234
7.5. Метод случайного баланса 240
7.6. Обсуждение и сравнение методов 258

8. Априорное моделирование 260

8.1. Постановка задачи априорного моделирования 260
8.2. Порядковые методы измерения качественной информации . . . 262
8.3. Анализ ранжированных данных. Коэффициент ранговой корреляции 263
8.4. Метод парных сравнений 268
8.5. Применение экспертных оценок при выявлении факторов, существенно влияющих на ход технологического процесса . . . 271
8.6. Сочетание опроса с экспериментом 274

9. Методы построения математических моделей объектов с изменяющимися во времени характеристиками (при наличии неконтролируемого дрейфа) 278

9.1. Введение 278
9.2. Метод проверки условий отсутствия дрейфа характеристик объекта 281
9.3. Построение математических моделей в условиях аддитивного дрейфа, описываемого детерминированной функцией времени . 288
9.4. Планирование эксперимента в условиях аддитивного дрейфа . 291
9.5. Адаптивный метод построения математической модели в условиях неаддитивного случайного дрейфа 305

9.5.1. Постановка задачи 305
9.5.2. Метод решения 307
9.5.3. Описание алгоритма и примеры 310

10. Планирование эксперимента и оценивание параметров нелинейных моделей 317

10.1. Введение и основные понятия 317
10.2. Метод наименьших квадратов 322
10.2.1. Теоретическое введение 322
10.2.2. Программа PREDAN на языке ФОРТРАН 330
10.3. Планирование эксперимента 336
10.3.1. Прямой метод 336
10.3.2. Непрямой метод 340
10.3.3. Правила планирования эксперимента без вычислений 343
10.4. Обработка экспериментальных данных 345
10.4.1. Метод наименьших квадратов 345
10.4.2. Определение структурных параметров 350
10.4.3. Адекватность модели 362
10.4.4. Устранение систематических ошибок 369
10.4.5. Чебышевская аппроксимация 370
10.4.6. Задача балансировки 371

11. Последовательные методы построения математических моделей . . 376

11.1. Последовательные методы в экспериментальных исследованиях 376
11.2. Рекуррентные алгоритмы вычисления 377
11.2.1. Постановка задачи 378
11.2.2. Рекуррентные алгоритмы вычисления оценок коэффициентов и дисперсионной матрицы плана 379
11.2.3. Рекуррентный алгоритм вычисления остаточной суммы квадратов 381
11.2.4. Пример 382
11.3. Последовательное планирование эксперимента 386
11.3.1. Постановка задачи последовательного планирования эксперимента 386
11.3.2. Рекуррентная формула для определителя дисперсионной матрицы 388
11.3.3. Вычислительная процедура последовательного планирования 389
11.4. Правило остановки эксперимента 391
11.4.1. Постановка задачи 392
11.4.2. Метод решения задачи 393
11.4.3. Правило останова 395
11.5. Последовательное планирование эксперимента в задачах построения нелинейных по параметрам моделей 397

12. Методы экспериментальной оптимизации 405

12.1. Постановка задачи 405
12.2. Экспериментальная оптимизация в случае одной переменной 407
12.3. Метод Бокса—Уилсона (метод крутого восхождения) . . . 410
12.4. Планирование эксперимента в производственных условиях . 430
12.4.1. Последовательный симплексный метод планирования эксперимента 430
12.4.2. Метод эволюционного планирования (ЭВОП) 438
12.5. Методы случайного поиска в задачах экспериментальной оптимизации 446

13. Методы классификации 455
13.1. Введение 455
13.2. Основные понятия и определения 456
13.3. Выбор существенных переменных 457
13.4. Статистические методы классификации 461
13.4.1. Параметрические методы 461
13.4.2. Локальные методы 465
13.5. Детерминистские методы решения задачи классификации . . 468
13.5.1. Интерполяционные методы построения классифицирующего правила 468
13.5.2. Методы построения разделяющей поверхности . . . 470
13.6. Оценка точности классифицирующего правила 474
13.7. Планирование эксперимента в задаче классификации . . . 476

14. Экспериментальные методы определения динамических характеристик объектов управления 482
14.1. Динамические характеристики объектов управления 482
14.2. Выбор способа идентификации объекта управления 486
14.3. Общая характеристика корреляционных методов идентификации 488
14.4. Некоторые особенности активного корреляционного метода при использовании функций Лагерра 494
14.4.1. Общие замечания 494
14.4.2. Исследование точности аппроксимации весовой функции объекта с помощью системы функций Лагерра . 496
14.4.3. Изучающий сигнал для активного метода идентификации 500
14.4.4. Основные рекомендации по применению активного метода идентификации с псевдослучайным двоичным сигналом 508

15. Математическое приложение 516
А. Специальные распределения вероятностей 516
15.1. Биномиальное распределение 517
15.2. Нормальное распределение 517
15.3. χ^2-распределение 519
15.4. t-распределение Стьюдента 521
15.5. F-распределение Фишера 522
Б. Метод наименьших абсолютных отклонений и чебышевская аппроксимация 523
15.6. Постановка задачи 523
15.7. Линейные модели и симметричные планы 525
15.8. Решение специальных задач 527
15.9. Влияние ошибки в наблюдениях 529
15.10. Несмещенность и эффективность 529
Таблицы 532
Предметный указатель 541

Уважаемый читатель!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., 2, издательство „Мир".