УДК 519.2+6
ББК 22.17, 22.19
К 85

Крянев А. В., Лукин Г. В. **Математические методы обработки неопределенных данных.** — 2-е изд., испр. — М.: ФИЗМАТЛИТ, 2006. — 216 с. — ISBN 5-9221-0724-0.

В первых главах монографии изложены основные понятия параметрической и непараметрической статистики, включая понятия оценки и свойств, предъявляемых к оценкам с точки зрения их вычисления при обработке данных на компьютере. В 7–13 главах монографии изложены методы и алгоритмы восстановления регрессионных зависимостей, включая методы прогнозирования и решения задач планирования оптимальных экспериментов.

Предполагается, что читатель предварительно освоил курс теории вероятностей и математической статистики на базе, например, книги В.С. Пугачева «Теория вероятностей и математическая статистика».

В монографии представлены некоторые новые методы робастного оценивания и учета априорной информации, включая алгоритмы их численной реализации.

Основная цель монографии — ознакомить читателя с наиболее эффективными и апробированными классическими и новыми статистическими методами оценки и восстановления, научить использовать эти методы при решении конкретных задач обработки неопределенных данных.

Монография предназначена научным работникам, аспирантам, студентам старших курсов различных специальностей.

Допущено Министерством образования Российской Федерации в качестве учебного пособия для студентов высших учебных заведений, обучающихся по направлению и специальности «Прикладная математика и информатика».

ISBN 5-9221-0724-0

ОГЛАВЛЕНИЕ

Предисловие 5
Введение 10
Основные обозначения 13

Глава 1. **Элементы математической статистики** 15
1.1. Параметрическая статистика. Свойства оценок 16
1.2. Метод моментов 19
1.3. Неравенства Фишера–Крамера–Рао 24
1.4. Метод максимального правдоподобия (ММП) 29

Глава 2. **Методы учета априорной информации в рамках параметрической статистики** 35
2.1. Метод Байеса 35
2.2. Минимаксный метод учета априорной информации 40
2.3. Обобщенный метод максимального правдоподобия учета априорной детерминированной информации 41
2.4. Обобщенный метод максимального правдоподобия учета априорной стохастической информации 44

Глава 3. **Устойчивые методы оценивания параметра положения** 48
3.1. Минимаксный метод Хьюбера 49
3.2. Робастные М-оценки 55

Глава 4. **Методы непараметрической статистики** 60
4.1. Восстановление функции распределения 60
4.2. Восстановление плотности распределения методом гистограмм . 62
4.3. Восстановление плотности распределения методом Розенблатта–Парзена 63
4.4. Восстановление плотности распределения проекционными методами 66
4.5. Восстановление плотности распределения регуляризованным методом гистограмм 69
4.6. Метод корневой оценки плотности распределения 71

Глава 5. **Проверка гипотез о законе распределения** 74
5.1. Проверка гипотезы о законе распределения в рамках непараметрической статистики 74
5.2. Проверка гипотезы о законе распределения в рамках параметрической статистики 76

Глава 6. **Численные методы статистического моделирования** 79
6.1. Способы моделирования случайных величин 79
6.2. Применение метода статистического моделирования для решения некоторых прикладных задач 84

Глава 7. **Метод наименьших квадратов для линейных моделей с неопределенными данными** 90
7.1. Примеры линейных моделей с неопределенными данными 90

7.2. Классическая схема МНК 93
7.3. Обобщения классической схемы МНК 103
7.4. Прогнозирование с помощью линейных регрессионных моделей . 109

Глава 8. **Робастные методы для линейных моделей с неопределенными данными** 111
8.1. Робастные М-оценки параметров линейных моделей 111
8.2. Численные методы нахождения робастных оценок параметров линейных моделей 113

Глава 9. **Учет априорной информации в линейных моделях с неопределенными данными** 117
9.1. Метод Байеса в рамках линейных моделей 118
9.2. Минимаксный метод учета детерминированной априорной информации 121
9.3. Обобщенный метод максимального правдоподобия учета априорной стохастической информации в рамках линейных моделей . . 124
9.4. Обобщенный метод максимального правдоподобия учета априорной детерминированной информации в рамках линейных моделей 137
9.5. Регуляризованный метод наименьших квадратов 140

Глава 10. **Метод наименьших квадратов для нелинейных моделей с неопределенными данными** 146
10.1. МНК-оценки параметров нелинейных моделей и их свойства . . . 146
10.2. Численные методы нахождения МНК-оценок параметров нелинейных моделей 149

Глава 11. **Методы выделения детерминированных и хаотических компонент временных рядов** 153
11.1. Метод сглаживающих ортогональных полиномов 153
11.2. Метод сглаживающих линейных сплайнов 158
11.3. Робастные линейные сглаживающие сплайны 163
11.4. Метод сглаживающих кубических сплайнов 167
11.5. Метод вейвлетов 177

Глава 12. **Методы прогнозирования хаотических временных рядов** 182
12.1. Методы прогнозирования с использованием априорной информации и ортогональных полиномов 182
12.2. Методы прогнозирования с использованием априорной информации и линейных сплайнов 185
12.3. Методы прогнозирования с использованием сингулярно-спектрального анализа 187

Глава 13. **Планирование оптимальных измерений при восстановлении функциональных зависимостей** 192
13.1. Постановка задач планирования оптимальных измерений 193
13.2. Методы решения задач планирования оптимальных измерений . 199

Список литературы 205
Предметный указатель 211

Предисловие

Представляемая нами книга написана на основе курса лекций, читаемого одним из авторов на протяжении более чем 20 лет студентам старших курсов МИФИ и слушателям факультетов повышения квалификации и переподготовки.

При подготовке настоящего издания были внесены дополнения, включающие некоторые последние достижения, в том числе авторов книги, в основном по двум направлениям: робастное оценивание и оценивание с учетом дополнительной априорной информации. В книге также представлены некоторые современные методы прогнозирования временных рядов и элементы теории вейвлет-преобразования.

Книга в первую очередь предназначена лицам, использующим математические методы обработки неопределенных данных при решении прикладных задач различного содержания. Кроме того, большинство разделов книги используются студентами МИФИ, выполняющими учебно-исследовательские работы, студентами-дипломниками и аспирантами, тематика исследований которых требует применения математических методов обработки неопределенных данных.

В книге рассматриваются неопределенные данные, неопределенность которых может быть обусловлена двумя причинами: 1) стохастическим характером данных; 2) неполной априорной информацией об искомом решении, задаваемой, например, в виде принадлежности решения детерминированному множеству. Вышеуказанный неопределенный характер данных и определил название книги.

В гл. 1 представлены базовые элементы математической статистики. Анализируются основные свойства оценок и соотношения между ними, в частности эффективность и робастность оценок. Изложен традиционный материал, в том числе интервальное оценивание математического ожидания и дисперсии нормально распределенной случайной величины. Рассматриваются два универсальных метода в рамках параметрической статистики: метод моментов и метод максимального правдоподобия (ММП). Метод моментов изложен в обобщенном виде, позволяющем существенно расширить рамки его применения. С помощью неравенства Фишера–Крамера–Рао анализируется эффективность оценок. Представлены основные свойства ММП-оценок. Изложение теоретических результатов сопровождается примерами, при рассмотрении которых конструируются функции правдоподобия и находятся информация Фишера и информационная матрица Фишера, ММП-оценки и их

характеристики. Особое внимание уделяется многомерному нормальному распределению.

Во 2-й главе рассматриваются методы учета дополнительной априорной информации в рамках параметрической статистики. Представлены четыре метода учета априорной информации в зависимости от ее вида. Если априорная информация об искомых параметрах имеет стохастический характер, то предлагается использовать два метода: метод Байеса и обобщенный метод максимального правдоподобия (ОММП) с заданием априорной выборки. Если же априорная информация об искомых параметрах u задается в виде принадлежности u априорному множеству R_a, то предлагается использовать два метода — минимаксный и ОММП с учетом соотношения $u \in R_a$. Анализируются схемы применения методов учета априорной информации и алгоритмы их численной реализации. Рассматриваются примеры применения методов, в частности случай, когда члены исходной выборки подчиняются нормальному закону.

В 3-й главе представлены робастные методы оценивания параметра положения в условиях наличия больших выбросов. Подчеркивается, что схема робастного оценивания параметра положения может быть без существенных качественных изменений обобщена на многомерный вариант оценивания параметров регрессионной модели. За основу робастного оценивания взят минимаксный подход Хьюбера. Предлагается итерационный метод численного нахождения робастной оценки параметра положения, основанный на геометрическом положении робастной оценки искомого параметра. Рассматривается общая схема получения робастных М-оценок, основанная на использовании функции влияния.

В гл. 4 представлены некоторые часто используемые для решения прикладных задач методы непараметрической статистики восстановления функции и плотности распределения. Подчеркивается, что задача восстановления плотности распределения по выборке, в отличие от аналогичной задачи для функции распределения, принадлежит к классу некорректно поставленных задач и поэтому может быть эффективно решена только при использовании дополнительной априорной информации об искомой плотности. Форма и объем априорной информации могут быть различными, и в зависимости от этого используются различные методы восстановления плотности распределения. В главе описаны 5 методов восстановления плотности распределения, использующие различные виды и уровни априорной информации. В методах гистограмм, Розенблатта–Парзена и корневой оценки плотности априорная информация используется для определения подходящих значений коэффициентов, аналогичных параметру регуляризации. В проекционных методах априорная информация может быть использована в виде задания априорной реперной плотности, в регуляризованном методе гистограмм априорная информация об искомой плотности задается в виде априорного класса плотностей.

В гл. 5 дается краткое изложение схем проверки гипотез о восстанавливаемом законе распределения в рамках параметрической и непараметрической статистик. Представлены критерии согласия Колмогорова, ω^2, χ^2.

Глава 6 посвящена численным методам статистического моделирования. В этой главе представлены способы моделирования случайных величин, в частности датчики равномерно распределенной нормированной случайной величины γ, основанные на методах Лемера и Неймана. Дано применение метода статистического моделирования для вычисления определенных интегралов и решения интегральных уравнений Фредгольма второго рода.

В гл. 7 изложен метод наименьших квадратов (МНК) для линейных моделей с неопределенными данными. Представлена классическая схема МНК и ее обобщения. Приводятся наиболее значимые свойства МНК-оценок и их обобщений. В конце главы дается линейная прогнозная модель, использующая МНК на этапе ее «обучения».

В гл. 8 представлены робастные схемы для линейных моделей с неопределенными данными. Все робастные схемы оценивания для линейных моделей строятся на основе функций влияния и М-оценивания. Особое внимание уделяется М-оценке Хьюбера. Предлагаются итерационные численные схемы нахождения нелинейных робастных оценок параметров линейных моделей, в частности итеративный МНК и метод вариационно-взвешенных квадратических приближений. Для нахождения робастной оценки Хьюбера предлагается эффективная итерационная процедура, сходящаяся к робастной оценке за конечное число итераций.

В гл. 9 изложены схемы учета дополнительной априорной информации в линейных моделях с неопределенными данными. Отмечается, что в условиях, когда информационная матрица Фишера исходной линейной модели вырождена или близка к вырожденной, задача оценивания параметров линейной модели принадлежит к классу некорректно поставленных задач и без учета дополнительной априорной информации невозможно получить приемлемые по точности оценки искомых параметров. В главе представлены различные схемы учета априорной информации, основанные на методах Байеса, минимаксном и ОММП. Для ОММП даны две схемы оценивания в зависимости от вида априорной информации. При применении первой схемы ОММП можно учитывать априорную информацию стохастического характера, задаваемую в виде априорной выборки. В этом случае предполагается, что совместная плотность вероятностей членов выборки зависит от искомого вектора u линейной модели. При применении второй схемы ОММП учитывается априорная информация детерминированного вида, задаваемая в виде априорного детерминированного множества, которому заведомо принадлежит искомый вектор u. В § 9.5 рассматривается также регуляризованный метод наименьших квадратов, позволяющий учитывать погрешность в элементах матрицы A исходной линейной модели $Au = f - \varepsilon$.

В гл. 10 в сжатой форме изложен метод наименьших квадратов для нелинейных моделей. Приведен итерационный метод Ньютона–Гаусса численного нахождения МНК-оценок решений нелинейных моделей. Даны регуляризованные модификации Левенберга–Марквардта итерационных процессов Ньютона–Гаусса.

Главы 11, 12 посвящены анализу и прогнозированию временных рядов. В гл. 11 представлены методы выделения детерминированных компонент временных рядов. Все представленные в главе методы основаны на представлении детерминированной компоненты в виде разложения по базисной системе функций и оценке коэффициентов разложения с помощью МНК или робастной схемы. В качестве базисной системы функций берутся: 1) система полиномов, ортогональных на множестве фиксированных значений аргумента; 2) линейные сплайны; 3) кубические сплайны; 4) вейвлеты. Предлагаются эффективные численные схемы расчета искомых коэффициентов детерминированных компонент. В гл. 12 представлено несколько сравнительно новых методов прогнозирования временных процессов. Первые два из них основаны на учете априорных экспертных оценок прогнозируемых величин и применении схем выделения детерминированных компонент, изложенных в гл. 11. Третий из представленных в гл. 12 методов прогнозирования базируется на сингулярно-спектральном анализе и выделении главных компонент исследуемого временного ряда. Указывается на соответствие между прогнозированием с помощью метода главных компонент и прогнозированием на основе линейной регрессионной модели.

В завершающей книгу гл. 13 дано краткое введение в планирование оптимальных измерений при восстановлении функциональных зависимостей. Рассматриваемая задача планирования особенно важна при проведении физических экспериментов, когда экспериментатор имеет возможность выбора значений аргумента, при которых производятся измерения восстанавливаемой функции. Вводятся различные типы оптимальности планов экспериментов и изложены методы решения задач оптимизации планов измерений для рассматриваемых типов оптимальности, включая алгоритмы их численного решения.

Ссылки на используемую литературу даны в квадратных скобках, а сам список литературы представлен в алфавитном порядке и включает в себя лишь цитируемые источники. Представленный список не претендует на полноту, а отражает лишь некоторые стороны изложенного в книге материала. Для читателя, желающего углубить свои знания по теории вероятностей и традиционным главам математической статистики, рекомендуем монографии В. С. Пугачева [62], Ж. Закса [33], М. Кендалла и А. Стюарта [38], С. Р. Рао [65], энциклопедию «Вероятность и математическая статистика» [19]. Теория матриц достаточно полно изложена в монографии Ф. Р. Гантмахера [23]. С численными методами, используемыми при обработке неопределенных данных, можно ознакомиться в книгах [20, 36, 51, 56, 70, 77].

Авторы благодарят заведующего кафедрой «Прикладная математика» МИФИ, д.ф.-м.н., профессора Н.А. Кудряшова, д.ф.-м.н., про-

фессора этой же кафедры Т. И. Савёлову и рецензентов за просмотр рукописи и высказанные замечания, учтенные в окончательном варианте книги.

Авторы будут благодарны за все замечания и предложения, которые можно направлять по e-mail: `book@kryanev.ru`. Материалы, использованные в данной монографии, а также информацию о направлениях деятельности авторов можно найти на сайте `www.kryanev.ru`.

Москва, январь 2003 г.

А. В. Крянев
Г. В. Лукин

Введение

Статистические методы все более широко применяют при решении различных прикладных задач. В первую очередь традиционно они используются при обработке данных экспериментов, причем круг задач обработки непрерывно расширяется прежде всего за счет расширения областей, где используются экспериментальные методы, и усложнения самих экспериментов.

В последние годы в связи с запросами практики интенсивно развиваются методы исследования, основанные на измерениях косвенной информации об изучаемых системах и объектах. Часто задачи математической обработки и интерпретации неопределенных данных по косвенным проявлениям изучаемых объектов принадлежат к классу некорректных задач. Разработка статистических методов решения некорректных задач обработки и интерпретации неопределенных данных еще далека от завершения, но и в этой области уже создано много подходов и методов, а на их основе — апробированных эффективных численных алгоритмов решения поставленных задач. Знать эти методы и уметь их использовать — насущная необходимость для любого экспериментатора и специалиста, обрабатывающего неопределенные данные.

Вторая значительная особенность последнего периода развития статистических методов исследования — широкое применение современной вычислительной техники при обработке неопределенных данных. Применение компьютеров при обработке данных дает возможность решать такие задачи, решение которых совсем недавно было невозможно. Например, большинство задач обработки результатов измерений косвенных проявлений изучаемых объектов невозможно решить без применения компьютеров. Кроме того, сокращение времени обработки при применении компьютеров дает возможность создания оперативных систем управления сложными объектами, в которых слежение за поведением объекта осуществляется с помощью экспериментальных методов (например, с помощью систем различных датчиков и приборов измерения). Бурное внедрение компьютеров и вычислительных методов в статистику, и, в частности, в статистическую обработку данных, заставило заново переосмыслить место и значение многих классических методов математической статистики под углом возможности их использования с применением компьютеров и потребовало разработки самих численных реализаций, новых методов и создания эффективных численных алгоритмов обработки данных. Знать основные из этих

методов и алгоритмов, владеть ими не только полезно, но и необходимо обработчику неопределенных данных.

В последние годы в самой математической статистике, в силу ее внутреннего развития и потребности практики, созданы новые подходы и методы, более полно и адекватно учитывающие реальные ситуации, возникающие при получении неопределенных данных. Прежде всего это относится к критическому переосмыслению основных предположений, упрощений и гипотез, которые берутся в качестве основы получения тех или иных статистических подходов и методов. Всякий разумный метод обработки должен обладать свойством устойчивости в том смысле, что любое возможное (физически реализуемое) малое уклонение от базовых предположений приведет лишь к малым изменениям в конечных результатах обработки данных. Такие устойчивые методы обработки в математической статистике получили название *робастных*.

При решении прикладных задач обработку неопределенных данных, как правило, проводят в условиях, когда у исследователя помимо исходных неопределенных данных имеется дополнительная априорная информация об изучаемом объекте. Ясно, что учет этой дополнительной информации позволяет в ряде случаев существенно повысить точность конечных результатов обработки. Более того, для целого класса задач обработки данных без учета дополнительной априорной информации вообще невозможно получить приемлемый конечный результат.

В математической статистике разработаны различные методы учета дополнительной информации об изучаемом объекте и ее «соединения» в процессе обработки с неопределенными данными, ознакомить с такими методами — также одна из целей настоящей книги.

В главах 7–13 книги представлены методы и алгоритмы восстановления регрессионных зависимостей и их применения для прогнозирования.

Задачи восстановления регрессионных зависимостей встречаются во всех областях естествознания при обработке неопределенных данных.

Обширный класс регрессионных зависимостей составляют линейные задачи. К линейным регрессионным задачам относятся, например, многие задачи восстановления функциональной зависимости, задачи исследования изучаемых объектов и систем по измерениям косвенной информации об этих объектах и системах (линейные задачи идентификации). Математические модели линейных регрессионных задач — системы линейных уравнений, правая часть которых имеет случайный характер. В некоторых задачах идентификации не только правая часть исходной системы уравнений, но и оператор левой части имеет неопределенный характер и часто задается со случайными погрешностями. Из-за наличия погрешностей исходная система уравнений, как правило, является противоречивой, т. е. не существует ни одного точного решения исходной системы при заданной правой части с погрешностями. Поэтому попытки решения таких систем «в лоб» с помощью традиционных методов, ориентированных на получение точного решения,

часто приводят к неудовлетворительному результату. В то же время, учитывая случайный характер погрешностей, для получения приемлемого приближенного решения этих систем естественно применять статистические методы.

Применение статистических подходов и методов для решения задач восстановления регрессионных зависимостей имеет ряд особенностей и специфических моментов, требующих качественного анализа и разработки эффективных численных алгоритмов получения приближенных решений. Ознакомить читателя с основными, наиболее эффективными и апробированными методами и численными алгоритмами решения задач восстановления регрессионных зависимостей — одна из целей настоящей книги.

В книге рассматриваются также нелинейные задачи идентификации, численное решение которых требует разработки специальных методов и алгоритмов.

Часто исследователь является не только пассивным обработчиком неопределенных данных, но и может повлиять на процесс их получения. В таких условиях естественно добиваться такого набора неопределенных данных (в рамках тех ограничений и возможностей, в которых проводится набор), при котором возможно получение наилучшего конечного результата по идентификации изучаемого объекта или системы. В книге изложены некоторые основные методы и алгоритмы решения задач планирования оптимального набора при восстановлении функциональных зависимостей.

Основные обозначения

$X_1, \ldots, X_n$ — элементы выборки
$p(x)$ — плотность вероятностей
$F(x)$ — функция распределения вероятностей
$p_\xi(x)$ — плотность вероятностей случайной величины ξ
$p(x;\ u)$ — плотность вероятностей, зависящая от параметра u
$p(x \mid u)$ — условная плотность вероятностей при фиксированном значении u
$\widehat{p}(x)$ — оценка плотности вероятностей $p(x)$
$\widehat{F}(x)$ — оценка функции распределения $F(x)$
$\widehat{u}$ — оценка параметра u
$F_n(x)$ — эмпирическая функция распределения
$\mathbf{M}\widehat{u}$ — математическое ожидание случайной величины $\widehat{u}$
$P_{\text{дов}}$ — доверительная вероятность
α_k — начальный момент k-го порядка
μ_k — центральный момент k-го порядка
$\sigma^2(\xi), \mathbf{D}(\xi)$ — дисперсия случайной величины ξ
s^2 — эмпирическая дисперсия
$\Phi(t) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^t e^{-\frac{s^2}{2}}\, ds$ — интеграл вероятностей
$\Gamma(t)$ — гамма-функция
χ^2_n — случайная величина, распределенная по закону хи-квадрат с n степенями свободы
$I_n, I_n(u)$ — информация Фишера или информационная матрица Фишера
$L_n, L_n(u)$ — функция правдоподобия
$l_n, l_n(u)$ — логарифмическая функция правдоподобия
K_ξ — ковариационная матрица случайного вектора ξ
$\overline{Y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} Y_i$ — средняя арифметическая
$\mathcal{N}(m, \sigma^2)$ — нормальный закон распределения с математическим ожиданием m и дисперсией σ^2
$\widehat{u}_{\text{Б}}$ — оценка Байеса параметра u
K_a — ковариационная матрица априорного распределения
u_a — центр априорного распределения параметра u
$p_{\text{апр}}(u)$ — плотность априорного распределения

U_a, R_a — априорное множество

$\widehat{u}_{\text{ММП}}$ — оценка метода максимального правдоподобия параметра u (ММП-оценка)

$\widehat{u}_{\text{ОММП}}$ — оценка обобщенного метода максимального правдоподобия параметра u (ОММП-оценка)

Δu — смещение оценки u

$\arg\min_u(max) f(u)$ — значение u, для которого целевая функция $f(u)$ достигает минимума (максимума)

K — параметр Хьюбера

$\rho(s)$ — функция влияния M-робастной оценки параметра положения

I_0, I_+, I_- — множество индексов, соответствующих средней части, правому и левому «хвостам» распределения Хьюбера

$K(u)$ — функция распределения Колмогорова

$K_n(u)$ — функция распределения Масси

$\delta(x)$ — обобщенная функция Дирака

$H_k(x)$ — полином степени k Чебышева–Эрмита

$L_k^{\alpha}(x)$ — полином степени k Чебышева–Лагерра

$L(c)$ — функция Лагранжа

γ — случайная величина, равномерно распределенная на промежутке $[0,1]$

ε — доля засорения в модели смеси Хьюбера или вектор погрешностей правой части регрессионной модели

$\widehat{u}_{\text{МНК}}$ — оценка метода наименьших квадратов параметра u

$\lambda_{\min}(C)(\lambda_{\max}(C))$ — минимальное (максимальное) собственное значение квадратной матрицы C

$\operatorname{cov}(X,Y)$ — ковариация случайных величин X и Y

$\operatorname{grad}_u f(u)$ — градиент функции $f(u)$

A^T — транспонированная матрица матрицы A

C^{-1} — обратная матрица невырожденной квадратной матрицы C

$\widehat{u}_{\text{ММ}}$ — оценка минимаксного метода параметра u

(u,v) — скалярное произведение векторов u и v

$\|u\|$ — евклидова норма вектора u

$S_n(x)$ — кубический интерполяционный сплайн

$S_{n\alpha}(x)$ — кубический сглаживающий сплайн

$S_{\alpha}(t)$ — линейный сглаживающий сплайн

$\psi(t)$ — базисный вейвлет

$\mathcal{E}(N)$ — план эксперимента

$\varepsilon(N)$ — нормированный план эксперимента

$M(\varepsilon)$ — информационная матрица, соответствующая плану эксперимента $\varepsilon(N)$

$\mathcal{D}(\varepsilon) = M^{-1}(\varepsilon)$ — обратная матрица информационной матрицы $M(\varepsilon)$

Глава 1
ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Математическая статистика, и, в частности, математическая обработка экспериментальных данных как ее составная часть, базируются на многих разделах математики и прежде всего на теории вероятностей. Предполагается, что читатель знаком с такими основными понятиями теории вероятностей, как случайные события и величины, вероятность события, операции над случайными событиями, случайными величинами и вероятностями, законы распределения случайных величин, предельные теоремы теории вероятностей. В качестве базовой книги, где изложены теория вероятностей и математическая статистика, рекомендуем использовать монографию [62]. Некоторые дополнительные сведения из теории вероятностей, используемые в математической статистике, будут вводиться в соответствующих местах монографии по мере необходимости.

Одной из основных, базовых задач математической статистики является задача восстановления неизвестного закона распределения случайной величины по конечному числу ее реализаций. Более подробно, рассматривается случайная величина ξ, для которой известна (из эксперимента) совокупность ее случайных реализаций

$$X_1, \ldots, X_n. \tag{1.1}$$

Требуется с максимально возможной точностью определить закон распределения ξ, например плотность вероятностей $p_\xi(x)$.

Совокупность случайных реализаций (1.1) называется *выборкой*, число реализаций n — *объемом выборки*. Если случайные реализации, рассматриваемые как случайные величины, независимы, то выборка называется *повторной*, в противном случае — *бесповторной*. Для повторной выборки, и только для нее, совместная плотность вероятностей совокупности случайных величин (1.1) (если она существует) имеет вид $\prod_{i=1}^{n} p_\xi(x_i)$. В дальнейшем, если специально не оговорено, предполагается, что (1.1) — повторная выборка. В математической статистике любую измеримую функцию (скалярную или векторную) от членов выборки (1.1) принято называть *статистикой*.

Иногда исследователя интересует не сам закон распределения случайной величины ξ, а лишь некоторые его характеристики, например *математическое ожидание* и *дисперсия* (соответственно, вектор математических ожиданий и ковариационная матрица, если ξ — векторная

случайная величина) или некоторые моменты случайной величины более высокого порядка.

При восстановлении плотности вероятности по выборке различают два подхода — параметрический и непараметрический.

Параметрический подход основан на гипотезе о принадлежности искомой плотности вероятностей $p_\xi(x;\ u)$ классу плотностей, определяемых совокупностью параметров u (если u фиксирован, то плотность $p_\xi(x;\ u)$ определена однозначно). Например, часто вводится гипотеза о нормальности распределения скалярной случайной величины ξ, т. е.

$$p_\xi(x;\ m, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-m)^2/(2\sigma^2)}$$

с неизвестными параметрами m, σ^2.

В рамках параметрического подхода задача восстановления плотности вероятностей сводится к задаче восстановления значений тех неизвестных параметров u, от которых зависит плотность вероятностей. При решении прикладных задач методами математической статистики исследователя чаще всего интересуют именно значения восстанавливаемых параметров, а не сама плотность вероятностей, поэтому методам параметрической статистики будет уделено наибольшее внимание.

Непараметрические методы восстановления законов распределения наблюдаемых случайных величин не предполагают такой жесткой фиксации класса законов распределения при помощи конечного числа параметров, хотя довольно часто в рамках непараметрического подхода используют метод «погружения» в параметрическое семейство с потенциальной возможностью неограниченного увеличения числа искомых параметров.

1.1. Параметрическая статистика. Свойства оценок

Интуитивно ясно, что по выборке конечного объема принципиально невозможно точное восстановление закона распределения в рамках как параметрического, так и непараметрического подходов. В результате восстановления мы получаем приближенные значения закона распределения (например, плотности вероятностей) или приближенные значения параметров (если использован параметрический подход). В математической статистике приближенные значения параметров плотности вероятностей или самой плотности вероятностей, полученные на основе выборки (1.1), принято называть *оценками*.

Оценки параметра u и плотности p_ξ будем обозначать символами $\widehat{u}$ и $\widehat{p}_\xi$ соответственно.

Любая оценка $\widehat{u}$ зависит от членов выборки (1.1) и поэтому является случайной величиной.

Оценка $\widehat{u}$ называется *несмещенной*, если $\mathbf{M}\,\widehat{u} = u_T$, где u_T — искомое точное значение параметра u. Если $\mathbf{M}\,\widehat{u} \neq u_T$, то оценка $\widehat{u}$ *смещена*.

Поскольку оценка $\widehat{u} = \widehat{u}(X_1, \ldots, X_n)$ зависит от членов выборки, то она, вообще говоря, будет изменяться при увеличении объема выборки

(добавлении новых членов выборки). Интуитивно ясно, что при $n \to \infty$ информация о законе распределения в выборке неограниченно увеличивается, и поэтому достаточно разумная оценка параметра должна сближаться (в том или ином смысле) с его точным значением.

Оценка $\widehat{u} = \widehat{u}(X_1, \ldots, X_n)$ называется *состоятельной по вероятности* (в среднем квадратичном), если $\widehat{u}(X_1, \ldots, X_n)$ сходится по вероятности (в среднем квадратичном) к u_T при $n \to \infty$. Если тип сходимости безразличен или ясен из контекста, то будем говорить о *состоятельности* оценки без указания типа сходимости.

Конечно, состоятельность является хорошим свойством оценки, но все же скорее успокоительного, а не конструктивного характера, поскольку на практике объем выборки всегда конечен. Для практики наиболее важным является вопрос о близости (в том или ином смысле) оценки к искомому значению параметра при фиксированном объеме выборки. Меру близости можно вводить по-разному. Например, если u — скаляр и оценка $\widehat{u}$ несмещена, то в качестве меры близости чаще всего выбирают дисперсию оценки (если $\widehat{u}$ — несмещенная оценка вектора, то мерой близости может служить ковариационная матрица оценки). Оценки, наиболее близкие к истинным значениям, называют *эффективными*. Одна из основных задач математической статистики — нахождение эффективных оценок.

Как уже указывалось выше, основой параметрического подхода является жесткое постулирование класса распределений. К сожалению, на практике реальные законы распределения случайных величин, образующих выборку (1.1), как правило, отличны от постулированного. Хотя такое отличие может быть и малым, оно всегда имеется. Если оценка малочувствительна к малым изменениям (в том или ином смысле) исходного постулированного закона распределения, то такая оценка называется *робастной*.

Любая оценка $\widehat{u}$ n-мерного вектора u представима в виде

$$\widehat{u} = S(X_1, \ldots, X_n), \tag{1.2}$$

где S — оператор из конечномерного пространства элементов $(X_1, \ldots, X_n)$ в m-мерное пространство E^m.

Если S — линейный оператор, то оценка (1.2) называется *линейной*, в противном случае $\widehat{u}$ — *нелинейная оценка*. В частности, если X_i — векторы размерности k, то линейная оценка представима в виде $\widehat{u} = SX + u_0$, где S — матрица размерности $(m \times n\,k)$, X — совокупность компонент векторов выборки (1.1), u_0 — фиксированный детерминированный вектор.

Класс линейных оценок играет важную роль в математической статистике, поскольку в этом классе часто удается найти «хорошие» оценки, а линейность оценки позволяет в большинстве случаев вычислять ее с меньшими вычислительными затратами по сравнению с нелинейными оценками.

Часто встречаются оценки, которые вместо свойств несмещенности обладают более общим свойством *асимптотической несмещенности*:

$\mathbf{M}\,\widehat{u}(X_1,\ldots,X_n) \to u_T$ при $n \to \infty$. В частности, состоятельные в среднем квадратичном оценки асимптотически несмещены.

Выше были рассмотрены некоторые основные свойства оценок. Конкретные оценки могут обладать одними из этих свойств и не обладать другими. Спрашивается, каким из перечисленных свойств оценок следует отдавать предпочтение при выборе подходящей оценки? Универсального ответа на этот вопрос не существует. Конечно, всегда желательно, чтобы оценка была эффективной. Но, к сожалению, часто эффективные оценки неустойчивы и тогда, в условиях опасности нарушения основных предположений, следует отдавать предпочтение робастным оценкам, эффективность которых максимальна. С другой стороны, и эффективные (или почти эффективные), и робастные оценки, как правило, даже для простых параметрических моделей могут быть вычислены только с помощью компьютеров. Линейные оценки наиболее легки для вычислений, но в большинстве случаев не обладают свойством робастности.

До сих пор мы рассматривали точечное оценивание параметров, в результате которого вместо неизвестного точного значения u_T получаем одно его приближенное значение — оценку $\widehat{u}$.

В математической статистике существует еще один способ оценивания — *интервальный*. При интервальном оценивании находят две оценки скалярного параметра u $u_{\text{н}}, u_{\text{в}}$ такие, что определена вероятность $\mathbf{P}\{u \in (u_{\text{н}}, u_{\text{в}})\} = p(u)$. Величину $p_{\text{дов}} = inf_u p(u)$ называют доверительным уровнем, а $1 - p_{\text{дов}}$ уровнем риска. В приложениях часто доверительный уровень и уровень риска выражают в процентах, а границы $u_{\text{н}}, u_{\text{в}}$ подбирают так, чтобы доверительный уровень равнялся одному из стандартных значений: 0,9; 0,95; 0,99; 0,999. Например, если $p_{\text{дов}} = 0{,}95$, то говорят о 95 % доверительном уровне и 5 % уровне риска.

Интервальные оценки более информативны, чем точечные, но их получение требует больших вычислительных затрат. Наиболее часто интервальные оценки получают, используя подходящую точечную оценку $\widehat{u}$, для которой удается построить плотность вероятностей $p_{\widehat{u}}(u)$. Тогда доверительный интервал $[\widehat{u} - \delta_1, \widehat{u} + \delta_2]$ находят из равенств

$$\int\limits_{\widehat{u}-\delta_1}^{\widehat{u}} p_{\widehat{u}}(u)\,du = \frac{p_{\text{дов}}}{2}, \qquad \int\limits_{\widehat{u}}^{\widehat{u}+\delta_2} p_{\widehat{u}}(u)\,du = \frac{p_{\text{дов}}}{2}.$$

В частности, если $p_{\widehat{u}}(u)$ симметрична относительно $\widehat{u}$, доверительный интервал $[\widehat{u} - \delta, \widehat{u} + \delta]$ симметричен относительно $\widehat{u}$, а $\delta > 0$ находят из равенства

$$\int\limits_{\widehat{u}-\delta}^{\widehat{u}+\delta} p_{\widehat{u}}(u)\,du = p_{\text{дов}}.$$

С методами математической статистики можно ознакомиться в книгах [2–5, 33, 38, 40, 48, 65].

1.2. Метод моментов

Пусть (1.1) — повторная выборка скалярной случайной величины ξ, у которой существуют моменты достаточно большого порядка.

В качестве оценки $\widehat{\alpha}_k$ начального момента k-го порядка $\mathbf{M}\,\xi^k$ возьмем статистику

$$\widehat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k. \tag{1.3}$$

В частности, оценка математического ожидания $\widehat{m}$ дается равенством

$$\widehat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{1.4}$$

В качестве оценки $\widehat{\mu}_k$ центрального момента k-го порядка $\mathbf{M}\,(\xi - \mathbf{M}\,\xi)^k$ возьмем статистику

$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{m})^k; \tag{1.5}$$

в частности, оценка дисперсии $\widehat{\sigma}^2$ дается равенством

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{m})^2. \tag{1.6}$$

Оценки (1.3) несмещены, так как $\mathbf{M}\,\widehat{\alpha}_k = \dfrac{1}{n}\sum\limits_{i=1}^{n}\mathbf{M}\,X_i^k = \mathbf{M}\,\xi^k$, состоятельны в среднем квадратичном в силу закона больших чисел (теорема Чебышева).

Оценки (1.5) также состоятельны в среднем квадратичном, но смещены, и, следовательно, обладают свойством асимптотической несмещенности. Например, для оценки (1.6)

$$\mathbf{M}\,\widehat{\sigma}^2 = \frac{n-1}{n}\,\sigma^2, \quad \mathbf{D}\,(\widehat{\sigma}^2) = \frac{\mu_4 - \sigma^4}{n} - \frac{2(\mu_4 - 2\sigma^4)}{n^2} + \frac{\mu_4 - 3\sigma^4}{n^3},$$

где μ_4 — центральный момент 4-го порядка случайной величины ξ.

Часто вместо асимптотически несмещенной оценки (1.6) используют несмещенную оценку дисперсии:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{m})^2, \tag{1.7}$$

которую называют *эмпирической дисперсией*.

Дисперсии несмещенных оценок (1.4), (1.7) равны $\mathbf{D}\,(\widehat{m}) = \sigma^2/n$, $\mathbf{D}\,s^2 = n^2(n-1)^{-2}\mathbf{D}\,(\widehat{\sigma}^2)$. В частности, если ξ распределена по нормальному закону, то $\mu_4 = 3\sigma^4$ и $\mathbf{D}\,(s^2) = 2\sigma^4/(n-1)$.

На оценках моментов (1.3) случайной величины ξ основан метод моментов оценки параметров u и искомой плотности вероятностей $p_\xi(x;\ u)$.

Пусть задано m линейно независимых функций $a_1(x), \ldots, a_m(x)$ (m — размерность вектора u). Имеем

$$\mathbf{M}\, a_j(\xi) = \int\limits_{-\infty}^{\infty} p_\xi(x;\ u) a_j(x)\, dx = b_j(u),$$

где $b_j(u)$ — известные функции искомых параметров u. В качестве состоятельных оценок $\mathbf{M}\, a_j(\xi)$ возьмем статистики

$$\widehat{a}_j(X) = \frac{1}{n} \sum_{i=1}^{n} a_j(X_i).$$

Тогда согласно методу моментов в качестве оценок параметров берется решение системы уравнений

$$b_j(\widehat{u}) = \widehat{a}_j(X), \qquad j = 1, \ldots, m. \tag{1.8}$$

В частности, если $a_j(X) = X^j$, $j = 1, \ldots, m$, то

$$\widehat{a}_j(x) = \widehat{\alpha}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j.$$

Пример 1.1. $p_\xi(x;\ u) = \begin{cases} u e^{-ux}, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$ где $u > 0$.

Возьмем $a_1(x) = x$. Имеем: $\mathbf{M}\, a_1(x) = \mathbf{M}\, \xi = 1/u$. Поэтому система (1.8) сводится к одному уравнению $\dfrac{1}{\widehat{u}} = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$, откуда $\widehat{u} = n \Big/ \sum\limits_{i=1}^{n} X_i$.

Замечание 1.1. Если интегралы

$$\int\limits_{-\infty}^{\infty} p_\xi(x;\ u) a_j(x)\, dx$$

аналитически не «берутся», то при решении нелинейной системы (1.8) (обычно одним из методов итераций) левые части равенств (1.8) подсчитываем, вычисляя эти интегралы численно.

Если ξ — векторная случайная величина размерности r, то каждый член $X_i = (X_{i1}, \ldots, X_{ir})^T$ выборки (1.1) — r-мерный вектор.

В качестве несмещенной и состоятельной оценки $\widehat{m}$ вектора математических ожиданий векторной случайной величины ξ можно взять

$$\widehat{m} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \Big(\sum_{i=1}^{n} X_{i1}, \ldots, \sum_{i=1}^{n} X_{ir} \Big)^T = (\overline{X}_1, \ldots, \overline{X}_r)^T.$$

Статистики

$$\widehat{h}_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ki} - \overline{X}_i)(X_{kj} - \overline{X}_j),$$

$$\widehat{r}_{ij} = \widehat{h}_{ij}(n-1)\left(\sum_{k=1}^{n}(X_{ki} - \overline{X}_i)^2 \sum_{k=1}^{n}(X_{kj} - \overline{X}_j)^2\right)^{-1/2}$$

являются состоятельными оценками ковариации $\operatorname{cov}(\xi_i, \xi_j)$ и коэффициента корреляции $r(\xi_i, \xi_j)$ компонент ξ_i и ξ_j векторной случайной величины ξ. Оценка $\widehat{h}_{ij}$ несмещена, оценка $\widehat{r}_{ij}$ смещена и потому является асимптотически несмещенной.

Следовательно, состоятельной и несмещенной оценкой ковариационной матрицы K_ξ является симметричная положительно определенная $(r \times r)$-матрица $\widehat{K}_\xi$, (i,j)-й элемент которой равен $\widehat{k}_{ij}$. Матрицу $\widehat{K}_\xi$ можно представить в виде

$$\widehat{K}_\xi = \frac{1}{n-1}X^T X, \quad X = \begin{pmatrix} X_{11} - \overline{X}_1 & X_{12} - \overline{X}_2 & \dots & X_{1r} - \overline{X}_r \\ X_{21} - \overline{X}_1 & X_{22} - \overline{X}_2 & \dots & X_{2r} - \overline{X}_r \\ \dots & \dots & \dots & \dots \\ X_{n1} - \overline{X}_1 & X_{n2} - \overline{X}_2 & \dots & X_{nr} - \overline{X}_r \end{pmatrix},$$

где X — $(n \times r)$-матрица.

Пусть ξ — скалярная нормально распределенная случайная величина и $p_\xi(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-m)^2/(2\sigma^2)}$.

Рассмотрим сначала случай, когда m неизвестен, σ^2 известен. В качестве оценки $\widehat{m}$ возьмем статистику (1.4). Случайная величина $\widehat{m}$ распределена по нормальному закону $\mathcal{N}(m, \sigma/\sqrt{n})$. Поэтому нормированная случайная величина $(\widehat{m} - m)/(\sigma\sqrt{n})$ распределена по нормированному нормальному закону $\mathcal{N}(0,1)$. Имеем

$$\mathbf{P}\left\{\frac{|\widehat{m} - m|}{\sigma/\sqrt{n}} \leqslant t\right\} = p_{\text{дов}} = 2\Phi(t),$$

где $\Phi(t) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^t e^{-s^2/2}\,ds$ — интеграл вероятностей. В таблице значений интеграла вероятностей находим $t = t(p_{\text{дов}})$, для которого $2\Phi(t) = p_{\text{дов}}$. Тогда доверительный интервал для параметра m с уровнем доверия $100p_{\text{дов}}\%$ имеет вид

$$\widehat{m} - t(p_{\text{дов}})\frac{\sigma}{\sqrt{n}} \leqslant m \leqslant \widehat{m} + t(p_{\text{дов}})\frac{\sigma}{\sqrt{n}}. \tag{1.9}$$

Для 90 %, 95 %, 99 %, 99,9 % уровней доверия коэффициент $t(p_{\text{дов}})$ равен 1,64; 1,96; 2,58; 3,29 соответственно. При увеличении уровня доверия длина доверительного интервала возрастает, например, длина

доверительного интервала при уровне риска в 0,1 % почти в два раза больше длины доверительного интервала для уровня риска в 10 %.

Рассмотрим теперь случай, когда m и σ^2 оба неизвестны. Тогда в качестве точечных оценок этих параметров возьмем статистики (1.4) и (1.7). При построении доверительных интервалов для m и σ^2 используется следующее утверждение.

Теорема 1.1. *Пусть* (1.1) — *повторная выборка случайной величины, распределенной по нормальному закону* $\mathcal{N}(m, \sigma)$. *Тогда* $\widehat{m}$ *и* s^2 *независимы и* $s^2(n-1)/\sigma^2$ *распределена по закону* χ^2_{n-1}.

Доказательство. Совместная плотность вероятностей совокупности независимых центрированных случайных величин $Y_i = X_i - m$, $i = 1, \ldots, n$, дается равенством

$$p(y) = p(y_1, \ldots, y_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\right) \sum_{i=1}^{n} y_i^2.$$

В n-мерном пространстве $(y_1, \ldots, y_n)^T$ совершим поворот вокруг начала координат так, чтобы в новой системе координат $z = (z_1, \ldots, z_n)^T$ выполнялось равенство $z_n = \dfrac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} y_i$. Имеем: $z = Uy$, $Z = UY$, где U — ортогональная матрица поворота. Случайные величины $Z_1, \ldots, Z_n$ независимы. Кроме того, $\sum\limits_{i=1}^{n} Z_i^2 = \sum\limits_{i=1}^{n} Y_i^2$, поскольку при повороте расстояния не изменяются. Обозначим $m_Y = \dfrac{1}{n} \sum\limits_{i=1}^{n} Y_i$. Имеем

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} (Y_i - m_Y)^2 + n\, m_Y^2 = \sum_{i=1}^{n-1} Z_i^2 + n\, m_Y^2,$$

откуда

$$s^2(n-1) = \sum_{i=1}^{n} (X_i - \widehat{m})^2 = \sum_{i=1}^{n} (Y_i - m_Y)^2 = \sum_{i=1}^{n-1} Z_i^2.$$

Из равенства $s^2(n-1) = \sum\limits_{i=1}^{n-1} Z_i^2$ следует, что величина $\widehat{m} = m + Z_n/\sqrt{n}$ не зависит от s^2, а s^2 распределена по закону χ^2_{n-1}. Теорема доказана.

Нормированная случайная величина $U_n = (\widehat{m} - m)/(\sigma\sqrt{n}\,)$ распределена по закону $\mathcal{N}(0,1)$ и не зависит от случайной величины $V_n = s^2(n-1)/\sigma^2$, распределенной по закону χ^2_{n-1}. Поэтому отношение Стьюдента $T_{n-1} = U_n/\sqrt{V_n/(n-1)} \equiv (\widehat{m} - m)/(s\sqrt{n}\,)$ распределено

по закону Стьюдента с плотностью вероятностей

$$p_{T_{n-1}}(t) = \frac{1}{\sqrt{\pi(n-1)}}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\left(1+\frac{t^2}{n-1}\right)^{-n/2}, \qquad t\in(-\infty,\infty).$$

Имеем: $\mathbf{P}\left\{\dfrac{|\widehat{m}-m|}{s/\sqrt{n}} < t\right\} = p_{\text{дов}}$, где $t = t(n-1, p_{\text{дов}})$ находится из равенства $2\displaystyle\int_0^t p_{T_{n-1}}(s)\,ds = p_{\text{дов}}$.

Итак, доверительный интервал для параметра m имеет вид

$$\widehat{m} - t(n-1, p_{\text{дов}})\frac{s}{\sqrt{n}} \leqslant m \leqslant \widehat{m} + t(n-1, p_{\text{дов}})\frac{s}{\sqrt{n}}. \tag{1.10}$$

При фиксированной $p_{\text{дов}}$ коэффициент $t(n-1, p_{\text{дов}})$ убывает с ростом n и $t(n-1, p_{\text{дов}}) \to t(p_{\text{дов}})$ — из формулы (1.9) при $n\to\infty$. При фиксированном n коэффициент $t(n-1, p_{\text{дов}})$ возрастает с ростом $p_{\text{дов}}$. Значения $t(n-1, p_{\text{дов}})$ для различных n и стандартных значений $p_{\text{дов}}$ табулированы. Например, для $n=5$ коэффициент Стьюдента равен 2,13; 2,78; 4,6; 8,61 для 90 %, 95 %, 99 % и 99,9 % уровней доверия. Так как при фиксированном уровне доверия $t(n-1, p_{\text{дов}}) > t(p_{\text{дов}})$, доверительный интервал (1.10) шире доверительного интервала (1.9). Увеличение доверительного интервала для m в случае неизвестного σ^2 является своеобразной «платой» за незнание σ^2.

Построение доверительного интервала для σ^2 также основано на результате теоремы 1.1. Действительно,

$$p_{\text{дов}} = \mathbf{P}\left\{u_2 < \frac{s^2(n-1)}{\sigma^2} < u_1\right\} = \int_{u_1}^{u_2} p_{\chi^2_{n-1}}(u)\,du,$$

где

$$p_{\chi^2_{n-1}}(u) = \frac{2^{-n/2}}{\Gamma((n-1)/2)}\,u^{(n-3)/2}e^{-u/2},$$

$u>0$ — плотность вероятностей случайной величины, распределенной по закону χ^2_{n-1}.

Границы $u_1(n-1, p_{\text{дов}})$, $u_2(n-1, p_{\text{дов}})$ обычно выбирают согласно правилу

$$\int_0^{u_2} p_{\chi^2_{n-1}}(u)\,du = \frac{1-p_{\text{дов}}}{2}, \qquad \int_{u_1}^{+\infty} p_{\chi^2_{n-1}}(u)\,du = \frac{1-p_{\text{дов}}}{2}.$$

Таким образом, доверительный интервал для σ^2 имеет вид

$$z_1(n-1, p_{\text{дов}})s^2 < \sigma^2 < z_2(n-1, p_{\text{дов}})s^2, \tag{1.11}$$

где $z_k(n-1, p_{\text{дов}}) = (n-1)u_k(n-1, p_{\text{дов}})$, $k=1,2$.

Коэффициент $z_1 \in (0{,}1)$ обладает следующими свойствами: при фиксированном n z_1 — монотонно убывающая функция $p_{\text{дов}}$, причем $z_1 \to 0$ при $p_{\text{дов}} \to 1$; $z_1 \to 1$ при $p_{\text{дов}} \to 0$; z_1 при фиксированной $p_{\text{дов}}$ монотонно возрастает с ростом n и $z_1 \to 1$ при $n \to \infty$. Коэффициент $z_2 > 1$, $z_2 \to 1$, монотонно убывая при $n \to \infty$ и фиксированной $p_{\text{дов}}$. Коэффициенты z_1, z_2 для различных значений n и $p_{\text{дов}}$ табулированы.

1.3. Неравенства Фишера–Крамера–Рао

Обозначим $p(x_1, \ldots, x_n;\ u)$ совместную плотность вероятностей совокупности случайных величин, образующих выборку (1.1). Функция членов выборки $p(X_1, \ldots, X_n;\ u)$ называется *функцией правдоподобия* и обозначается $L_n(X_1, \ldots, X_n;\ u)$. Если (1.1) — повторная выборка, то

$$L_n(X_1, \ldots, X_n;\ u) = \prod_{i=1}^{n} p_\xi(X_i;\ u).$$

Часто вместо функции правдоподобия используют *логарифмическую функцию правдоподобия*

$$l_n(X_1, \ldots, X_n;\ u) = \ln\ L_n(X_1, \ldots, X_n;\ u).$$

Для повторной выборки

$$l_n(X_1, \ldots, X_n;\ u) = \sum_{i=1}^{n} l_1(X_i;\ u),$$

где $l_1(X_i;\ u) = \ln\ p_\xi(X_i;\ u)$.

Как мы уже знаем, существует много несмещенных и состоятельных оценок параметра u. Спрашивается: существует ли верхняя граница точности таких оценок, построенных на основе выборки (1.1)?

Рассмотрим сначала случай, когда u — скаляр. Тогда степень близости любой несмещенной оценки $\widehat{u}$ к истинному значению u_T дается дисперсией оценки $\mathbf{D}\,\widehat{u} = \mathbf{M}\,(\widehat{u} - u_T)^2$.

Теорема 1.2 (Фишер–Крамер–Рао). *Пусть $\widehat{u} = \widehat{u}(X_1, \ldots, X_n)$ — любая несмещенная оценка u. Тогда*

$$\mathbf{D}\,\widehat{u} \geqslant \frac{1}{I_n}, \tag{1.12}$$

где

$$I_n = -\mathbf{M}\frac{\partial^2 l_n(X_1, \ldots, X_n;\ u)}{\partial u^2}.$$

Доказательство. Имеем

$$\int_{-\infty}^{\infty} \ldots \int L_n(X_1, \ldots, X_n;\ u)\, dX_1 \ldots dX_n = 1.$$

Откуда

$$\int\limits_{-\infty}^{\infty} \dots \int \frac{\partial L_n}{\partial u}\, dX_1 \dots dX_n = 0. \tag{1.13}$$

Но $\mathbf{M}\,\widehat{u} = u$ или

$$\int\limits_{-\infty}^{\infty} \dots \int \widehat{u}(X_1, \dots, X_n) L_n(X_1, \dots, X_n;\, u)\, dX_1 \dots dX_n = u.$$

Дифференцируя последнее равенство, получаем

$$\int\limits_{-\infty}^{\infty} \dots \int \widehat{u}\, \frac{\partial L_n}{\partial u}\, dX_1 \dots dX_n = 1.$$

Умножим равенство (1.13) на u_T и вычтем из последнего равенства:

$$\int\limits_{-\infty}^{\infty} \dots \int (\widehat{u} - u_T) \frac{\partial L_n}{\partial u}\, dX_1 \dots dX_n = 1.$$

Левую часть последнего равенства можно записать в виде

$$\int\limits_{-\infty}^{\infty} \dots \int (\widehat{u} - u_T) \frac{1}{L_n} \frac{\partial L_n}{\partial u}\, L_n\, dX_1 \dots dX_n \equiv \mathbf{M}\,(\widehat{u} - u_T) \frac{\partial l_n}{\partial u}.$$

Итак, получено равенство

$$\mathbf{M}\,(\widehat{u} - u_T) \frac{\partial l_n}{\partial u} = 1.$$

Обозначим $Y = \widehat{u} - u_T$, $Z = \dfrac{\partial l_n}{\partial u}$. Из равенства (1.13) следует, что $\mathbf{M}\,Z = 0$. Поэтому левая часть равенства $\mathbf{M}\,YZ = 1$ — ковариация центрированных случайных величин Y, Z. Но для любых двух случайных величин Y, Z справедливо неравенство Коши–Буняковского $\mathbf{M}^2 YZ \leqslant \leqslant \mathbf{M}\,Y^2\, \mathbf{M}\,Z^2$. Поэтому

$$1 \leqslant \mathbf{M}\,(\widehat{u} - u_T)^2\, \mathbf{M} \left(\frac{\partial l_n}{\partial u}\right)^2.$$

Имеем: $\dfrac{\partial l_n}{\partial u} = \dfrac{1}{L_n} \dfrac{\partial L_n}{\partial u}$, откуда

$$\frac{\partial^2 l_n}{\partial u^2} = \frac{1}{L_n} \frac{\partial^2 L_n}{\partial u^2} - \frac{1}{L_n^2} \left(\frac{\partial L_n}{\partial u}\right)^2 \equiv \frac{1}{L_n} \frac{\partial^2 L_n}{\partial u^2} - \left(\frac{\partial l_n}{\partial u}\right)^2.$$

Поэтому

$$\mathbf{M} \left(\frac{\partial l_n}{\partial u}\right)^2 = -\mathbf{M}\, \frac{\partial^2 l_n}{\partial u^2},$$

так как

$$\mathbf{M}\frac{1}{L_n}\frac{\partial L_n^2}{\partial u^2} \equiv \int\limits_{-\infty}^{\infty}\ldots\int \frac{\partial^2 L_n}{\partial u^2}\,dX_1\ldots dX_n = 0.$$

Теорема доказана.

Замечание 1.2. В тексте теоремы 1.2 не зафиксированы условия, налагаемые на функцию правдоподобия L_n, при выполнении которых справедливы выкладки доказательства теоремы. Эти условия будут приведены ниже при формулировке неравенства Фишера–Крамера–Рао для многомерного случая.

Величина I_n не зависит от оценки $\widehat{u}$, а определяется только законом распределения членов выборки и характером зависимости этого закона от искомого параметра u. Из доказательства теоремы 1.2 следует, что I_n положительна. Если выборка повторная, то

$$I_n = -\sum_{i=1}^{n}\mathbf{M}\,\frac{\partial^2 l_1}{\partial u^2} = n\,I_1, \quad \text{где} \quad I_1 = -\mathbf{M}\,\frac{\partial^2 l_1}{\partial u^2}.$$

Таким образом, I_n характеризует знание о параметре u, «заложенное» в выборке (1.1), и не зависит от конкретной реализации выборки, поэтому I_n называют *информацией* (*Фишера*). Из вышесказанного следует, что если выборка повторная, то увеличение выборки в k раз увеличивает информацию Фишера также в k раз.

Из неравенства Фишера–Крамера–Рао следует, что величина, обратная к информации Фишера, ограничивает снизу дисперсию любой несмещенной оценки параметра u, построенной на основе выборки (1.1). Оценки, для которых в неравенстве Фишера–Крамера–Рао достигается равенство, называются *эффективными*. Отношение $e_n = 1/(I_n\,\mathbf{D}\,\widehat{u})$, $0 \leqslant e_n \leqslant 1$, называется *эффективностью оценки* $\widehat{u}$. Для эффективных оценок (если они существуют) и только для них $e_n = 1$.

Пример 1.2. Пусть

$$p_\xi(x;\,u) = \frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-(x-u)^2/(2\sigma^2)}$$

и выборка (1.1) повторная. Тогда

$$L_n(X_1,\ldots,X_n;\,u) = \prod_{i=1}^{n} p_\xi(X_i;\,u) = \frac{\exp\left[-\dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i-u)^2\right]}{(2\pi)^{n/2}\sigma^n},$$

$$l_n = -\frac{n}{2}\,\ln\,2\pi - n\,\ln\,\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-u)^2, \quad I_1 = -\mathbf{M}\,\frac{\partial^2 l_1}{\partial u^2} = \frac{1}{\sigma^2}.$$

Поэтому $I_n = n/\sigma^2$. Для несмещенной оценки

$$\widehat{u} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \mathbf{D}\,\widehat{u} = \frac{\sigma^2}{n},$$

и, следовательно, оценка $\widehat{u}$ эффективна.

Пример 1.3. Пусть $p_\xi(x;\ u) = ue^{-ux}$, $x > 0$, и выборка (1.1) повторная. Тогда

$$L_n = u^n e^{-u\sum\limits_{i=1}^{n} X_i}, \quad X_i > 0, \quad i = 1, \ldots, n,$$

$$l_n = n \ln u - u\sum_{i=1}^{n} X_i, \quad I_1 = \frac{1}{u^2}, \quad I_n = \frac{n}{u^2}.$$

Итак, для неизвестного параметра экспоненциального закона информация Фишера зависит от значения самого искомого параметра. Для оценки $\widehat{u} = n\Big/\sum\limits_{i=1}^{n} X_i$ $e_n < 1$.

Пусть теперь u — вектор размерности m. Матрица $I_n = (I_{ij})$, где

$$I_{ij} = -\mathbf{M}\,\frac{\partial^2 l_n}{\partial u_i \partial u_j}, \quad i, j = 1, \ldots, m,$$

называется *информационной матрицей Фишера*. Из определения информационной матрицы следует ее симметричность. Имеем

$$\frac{\partial l_n}{\partial u_j} = \frac{1}{L_n}\frac{\partial L_n}{\partial u_j}, \quad \frac{\partial^2 l_n}{\partial u_i \partial u_j} = \frac{1}{L_n}\frac{\partial^2 L_n}{\partial u_i \partial u_j} - \frac{1}{L_n^2}\frac{\partial L_n}{\partial u_i}\frac{\partial L_n}{\partial u_j}.$$

Но

$$\mathbf{M}\,(\frac{1}{L_n}\frac{\partial^2 L_n}{\partial u_i \partial u_j}) = \int\limits_{-\infty}^{\infty}\ldots\int \frac{\partial^2 L_n}{\partial u_i \partial u_j}\, dX_1 \ldots dX_n = 0,$$

$$\mathbf{M}\,(\frac{1}{L_n}\frac{\partial L_n}{\partial u_i}) = \int\limits_{-\infty}^{\infty}\ldots\int \frac{\partial L_n}{\partial u_i}\, dX_1 \ldots dX_n = 0.$$

Поэтому

$$-\mathbf{M}\,\frac{\partial^2 l_n}{\partial u_i \partial u_j} = \mathbf{M}\,(\frac{1}{L_n}\frac{\partial L_n}{\partial u_i}\cdot\frac{1}{L_n}\frac{\partial L_n}{\partial u_j}) = \operatorname{cov}(\frac{1}{L_n}\frac{\partial L_n}{\partial u_i}, \frac{1}{L_n}\frac{\partial L_n}{\partial u_j}).$$

Таким образом, информационная матрица Фишера является ковариационной матрицей совокупности центрированных случайных величин $\left(\dfrac{1}{L_n}\dfrac{\partial L_n}{\partial u_1}, \ldots, \dfrac{1}{L_n}\dfrac{\partial L_n}{\partial u_m}\right)$ и поэтому неотрицательна.

Предположим, что I_n невырождена и

$$\int_{-\infty}^{\infty} \dots \int \left|\frac{\partial l_n}{\partial u_i}\right| dX_1 \dots dX_n < \infty$$

для любого u и $i = 1, \dots, m$.

Пусть $\widehat{u} = (\widehat{u}_1, \dots, \widehat{u}_m)^T$ — несмещенная оценка искомого вектора u_T, построенная по выборке (1.1).

Теорема 1.3. *Ковариационная матрица K_u любой несмещенной оценки $\widehat{u}$ удовлетворяет неравенству*

$$K_u \geqslant I_n^{-1}. \tag{1.14}$$

Неравенство (1.14) — обобщение неравенства Фишера–Крамера–Рао (1.12) на случай векторного искомого параметра u. Оно означает, что симметричная матрица $K_u - I_n^{-1}$ неотрицательна. Элементы главной диагонали матрицы K_u являются дисперсиями $\sigma^2_{u_i}$ компонент $\widehat{u}_i$ вектора оценки $\widehat{u}$. Поэтому из неравенства Фишера–Крамера–Рао (1.14), в частности, следуют неравенства $\sigma^2_{u_i} \geqslant I^{-1}_{ii}$, $i = 1, \dots, m$, где I^{-1}_{ii} — i-й элемент главной диагонали матрицы, обратной к информационной матрице Фишера.

Неравенство Фишера–Крамера–Рао обобщается и на случай смещенных оценок. Заметим, что неравенство (1.14) эквивалентно выполнению неравенства $(K_u v, v) \geqslant (I_n^{-1} v, v)$ для любого $v \in E^m$. Обозначим $b_i(u)$ смещение i-й компоненты вектора оценки $\widehat{u} = (\widehat{u}, \dots, \widehat{u}_m)^T$, т. е. $\mathbf{M}\,\widehat{u} = u_i + b_i(u)$.

Теорема 1.4. *Пусть* 1° *выполнены условия теоремы* 1.3 (*условия регулярности*); 2° $\dfrac{\partial b_i(u)}{\partial u_j}$ *существуют для всех u, где $i, j = 1, \dots, m$. Тогда для любого вектора $v \in E^m$ справедливо неравенство*

$$(K_u v, v) \geqslant (I_n^{-1} v, v)\left(1 + \frac{(\mathbf{D}\,(u) I_n^{-1} v, v)}{(I_n^{-1} v, v)}\right)^2, \tag{1.15}$$

где (i, j)-й элемент матрицы $\mathbf{D}\,(u)$ дается равенством

$$D_{ij}(u) = \frac{\partial b_i(u)}{\partial u_j}.$$

В частности, для одномерного случая ($m = 1$) неравенство (1.15) эквивалентно неравенству

$$\mathbf{M}\,(\widehat{u} - u)^2 \geqslant \frac{(1 + b'(u))^2}{I_n}, \tag{1.16}$$

где $b(u)$ — смещение оценки $\widehat{u}$.

Так же как и для одномерного случая, несмещенная оценка $\widehat{u} = (\widehat{u}_1, \dots, \widehat{u}_m)^T$ называется *эффективной* (*совместно эффективной*),

если в (1.14) достигается равенство. Оказывается, что оценка $\widehat{u}$ является эффективной тогда и только тогда, когда закон распределения оценки $\widehat{u}$ нормальный (теорема Хогга–Крейга) [33].

1.4. Метод максимального правдоподобия (ММП)

ММП позволяет получать оценки, которые в определенном смысле близки к достижению границы Фишера–Крамера–Рао, а если эффективные оценки существуют, то совпадают с ними.

Согласно ММП за оценку параметра u принимается решение экстремальной задачи

$$\widehat{u}_{\text{ММП}} = \arg\max_u L_n(X_1, \ldots, X_n;\ u). \tag{1.17}$$

Если функция правдоподобия дифференцируема по u, то ММП-оценка (1.17) является решением системы уравнений правдоподобия

$$\operatorname{grad}_u\ L_n(X_1, \ldots, X_n;\ \widehat{u}) = 0. \tag{1.18}$$

Часто вместо экстремальной задачи (1.17) и системы (1.18) удобно рассматривать эквивалентные им задачи

$$\widehat{u}_{\text{ММП}} = \arg\max_u l_n(X_1, \ldots, X_n;\ u), \tag{1.19}$$

$$\operatorname{grad}_u\ l_n(X_1, \ldots, X_n;\ \widehat{u}) = 0. \tag{1.20}$$

Прежде чем перечислять общие свойства ММП-оценок, найдем ММП-оценки параметров для некоторых стандартных распределений.

Пример 1.4. Пусть

$$p_\xi(x;\ u) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-m_0)^2/(2\sigma^2)},$$

где $u = (m_0, \sigma^2)^T$ и выборка (1.1) повторная. Имеем

$$l_n(X_1, \ldots, X_n;\ m_0, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - m_0)^2.$$

Получаем систему уравнений правдоподобия

$$\frac{\partial l_n}{\partial m_0} = \frac{1}{\widehat{\sigma}^2}\sum_{i=1}^n (X_i - \widehat{m}_0) = 0, \quad \frac{\partial l_n}{\partial \sigma^2} = -\frac{n}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}\sum_{i=1}^n (X_i - \widehat{m}_0)^2 = 0,$$

откуда

$$\widehat{m}_0 = \frac{1}{n}\sum_{i=1}^n X_i, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \widehat{m}_0)^2.$$

Из сказанного выше следует, что ковариационная матрица вектора $\widehat{u} = (\widehat{m}_0, \widehat{\sigma}^2)^T$ равна $K_u = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2(n-1)\sigma^4/n^2 \end{pmatrix}$. Информационная

матрица Фишера $I_n(m_0, \sigma^2) = \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/(2\sigma^4) \end{pmatrix}$. Следовательно, при $n \to \infty$ ковариационная матрица ММП-оценки достигает нижней границы Фишера–Крамера–Рао. Такое свойство оценок называют *асимптотической эффективностью*. Заметим, что ММП-оценка $\widehat{m}_0$ несмещена, а ММП-оценка $\widehat{\sigma}^2$ асимптотически несмещена.

Пример 1.5. Пусть $p_\xi(x;\ u) = u e^{-ux}$, $x > 0$. Имеем

$$l_n(X_1, \ldots, X_n;\ u) = n \ln u - u \sum_{i=1}^{n} X_i, \quad \frac{\partial l_n}{\partial u} = \frac{n}{\widehat{u}} - \sum_{i=1}^{n} X_i = 0,$$

откуда $\widehat{u}_{\text{ММП}} = n \Big/ \sum_{i=1}^{n} X_i$.

Пример 1.6. Пусть

$$p_\xi(x;\ u) = \frac{1}{b-a}, \quad a \leqslant x \leqslant b, \quad u = (a, b)^T;$$

$$L_n(X_1, \ldots, X_n;\ a, b) = \frac{1}{(b-a)^n}, \quad a \leqslant X_{\min}, \quad b \geqslant X_{\max},$$

где

$$X_{\min} = \min\{X_1, \ldots, X_n\}, \quad X_{\max} = \max\{X_1, \ldots, X_n\}.$$

Следовательно, максимум L_n достигается при $\widehat{b}_{\text{ММП}} = X_{\max}$, $\widehat{a}_{\text{ММП}} = X_{\min}$.

Пример 1.7. Пусть совместная плотность вероятностей членов выборки дается равенством

$$p(x_1, \ldots, x_n;\ u) = \frac{1}{(2\pi)^{n/2} |K_x|^{1/2}} \exp\{-\frac{1}{2} (K_x^{-1}(x - u1), x - u1)\},$$

где $x = (x_1, \ldots, x_n)^T$, $1 = (1, \ldots, 1)^T$, т. е. все члены выборки имеют одинаковое математическое ожидание u, распределены нормально и между ними существуют корреляционные связи. Имеем

$$l_n(X_1, \ldots, X_n;\ u) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |K_x| - \frac{1}{2} (K_x^{-1}(X - u1), X - u1),$$

где $X = (X_1, \ldots, X_n)^T$. Отсюда

$$\frac{\partial l_n}{\partial u} = -u(K_x^{-1}1, 1) + (K_x^{-1}1, X), \quad \widehat{u}_{\text{ММП}} = \frac{(K_x^{-1}1, X)}{(K_x^{-1}1, 1)}.$$

В частности, если $K_x = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, то $(K_x^{-1}1, 1) = \sum\limits_{i=1}^{n} \dfrac{1}{\sigma_i^2}$, $(K_x^{-1}1, X) = \sum\limits_{i=1}^{n} \dfrac{X_i}{\sigma_i^2}$ и $\widehat{u}_{\text{ММП}} = \sum\limits_{i=1}^{n} \dfrac{X_i}{\sigma_i^2} \Big/ \sum\limits_{i=1}^{n} \dfrac{1}{\sigma_i^2}$ — средневзвешенная оценка.

ММП-оценки обладают свойством инвариантности.

Теорема 1.5. *Пусть совокупность функций* $g_1(u), \ldots, g_k(u)$, $1 \leqslant k \leqslant m$, *отображает область* $U \subset E^m$ *изменения параметра* u *в* k*-мерный параллелепипед* $\Pi_k \in E^k$. *Тогда* ММП*-оценка параметра* $g_j(u)$ *дается равенством* $\widehat{u}_{j\,ММП} = g_j(\widehat{u}_{ММП})$.

Пример 1.8. Пусть (1.1) — повторная выборка логнормального распределения, т. е. $\ln X_i \sim \mathcal{N}(m_0, \sigma^2)$. Известно, что математическое ожидание μ и дисперсия d^2 логнормального распределения связаны с m_0 и σ^2 соотношениями

$$\mu = e^{m_0 + \sigma^2/2}, \quad d^2 = \mu^2(e^{\sigma^2} - 1).$$

Обозначим

$$Y_i = \ln X_i, \quad i = 1, \ldots, n, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (Y_i - \overline{Y})^2.$$

Тогда согласно теореме 1.5 ММП-оценки параметров μ, d^2 логнормального распределения даются формулами

$$\widehat{\mu}_{\text{ММП}} = \exp\{\overline{Y} + \widehat{\sigma}^2/2\}, \quad \widehat{d}^2_{\text{ММП}} = \widehat{\mu}^2_{\text{ММП}}(e^{\widehat{\sigma}^2} - 1).$$

Следующее свойство ММП-оценок связано с понятием достаточной статистики.

Пусть $T(X_1, \ldots, X_n)$ — такая статистика, что условное распределение любой другой статистики $Z(X_1, \ldots, X_n)$, не связанной функциональной зависимостью с T, при заданной $T(X_1, \ldots, X_n)$ не зависит от искомого параметра u. Тогда $T(X_1, \ldots, X_n)$ называют *достаточной статистикой*. Если значение достаточной статистики $T(X_1, \ldots, X_n)$ на выборке (1.1) известно, то знание значения любой другой статистики на выборке (1.1) не добавляет никакой новой информации об искомом параметре u. Не всякое семейство совместной плотности вероятностей членов выборки (1.1) допускает существование достаточной статистики. С другой стороны, если достаточная статистика существует, то всегда желательно ее найти, поскольку тогда вместо всей выборки (1.1) можно сохранять лишь значение достаточной статистики, не теряя при этом информации об искомых параметрах.

Ниже приведен простой критерий существования и определения достаточной статистики.

Теорема 1.6 (факторизации Фишера–Неймана). *Для того чтобы статистика* $T(X_1, \ldots, X_n)$ *была достаточной, необходимо и доста-*

точно выполнение представления

$$L_n(X_1, \ldots, X_n;\ u) = f(T(X_1, \ldots, X_n);\ u)\,\varphi(X_1, \ldots, X_n),$$

где сомножитель φ не зависит от искомого параметра u.

Пример 1.9. Пусть $p_\xi(x;\ u) \sim \mathcal{N}(m_0, \sigma^2)$ и (1.1) — повторная выборка. Имеем

$$L_n(X_1, \ldots, X_n;\ u) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - m_0)^2\right] =$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - m_0)^2\right]\right\}.$$

Поэтому набор двух скалярных статистик $\left\{\overline{X}, \sum\limits_{i=1}^{n}(X_i - \overline{X})^2\right\}$ — достаточная (векторная) статистика, если неизвестны m_0 и σ^2. Ясно, что набор $\left\{\sum\limits_{i=1}^{n} X_i, \sum\limits_{i=1}^{n} X_i^2\right\}$ также является достаточной статистикой.

Пример 1.10. Пусть $p_\xi(x;\ u) = 1/u$, $0 \leqslant x \leqslant u$, и выборка (1.1) повторная. Тогда $L_n(X_1, \ldots, X_n;\ u) = 1/u^n$, $X_i < u$. Функцию L_n можно представить в виде $L_n = 1/u^n$, $X_{\max} < u$. Поэтому $X_{\max}$ — достаточная статистика, если u неизвестно.

Если для закона распределения выборки (1.1) имеется достаточная статистика $T(X_1, \ldots, X_n)$, то ММП-оценки выражаются только через $T(X_1, \ldots, X_n)$ и тем самым не теряют информации о параметре u. Действительно, пусть

$$L_n(X_1, \ldots, X_n;\ u) = f(T(X_1, \ldots, X_n);\ u)\,\varphi(X_1, \ldots, X_n).$$

Тогда уравнение правдоподобия имеет вид

$$\operatorname{grad}_u \ln f(T(X_1, \ldots, X_n);\ u) = 0,$$

и тем самым $\widehat{u}_{\text{ММП}}$ зависит только от достаточной статистики $T(X_1, \ldots, X_n)$.

Пример 1.11. Пусть

$$p_\xi(x;\ u) = \frac{1}{b-a}, \quad a \leqslant x \leqslant b,$$

и выборка (1.1) повторная. Тогда $\widehat{a}_{\text{ММП}} = X_{\min}$, $\widehat{b}_{\text{ММП}} = X_{\max}$, причем $\{X_{\min}, X_{\max}\}$ — достаточная статистика при неизвестных a и b. В частности, ММП-оценкой математического ожидания $\mathbf{M}\,\xi$ будет $(X_{\min} + X_{\max})/2$.

Пусть

$$p_\xi(x;\ u) = b(x)\exp\left[\sum_{j=1}^{m} u_j T_j(x) + \psi(u)\right].$$

Тогда $p_\xi(x;\ u)$ называют *экспоненциальной плотностью.*

Теорема 1.7. *Пусть* 1° *выборка* (1.1) *повторная и* $p_\xi(x;\ u)$ — *экспоненциальная плотность;* 2° $(m \times m)$*-симметричная матрица* R *с элементами* $R_{ij} = \dfrac{\partial^2 \psi}{\partial u_i \partial u_j}$ *положительно определена для всех* $u \in U \subset E^m$. *Тогда* ММП*-оценка* $\widehat{u}_{ММП}$ *единственна, является достаточной статистикой и корнем системы уравнений*

$$-\frac{1}{n}\sum_{i=1}^{n} T_j(X_i) = \frac{\partial \psi(\widehat{u}_{ММП})}{\partial u_j}, \qquad j = 1, \ldots, m. \tag{1.21}$$

Доказательство. Имеем

$$L_n(X_1, \ldots, X_n;\ u) = \prod_{i=1}^{n} b(X_i)\exp\left\{\sum_{j=1}^{m} u_j \sum_{i=1}^{n} T_j(X_i) + n\psi(u)\right\}.$$

Поэтому набор $\left\{\sum\limits_{i=1}^{n} T_1(X_i), \ldots, \sum\limits_{i=1}^{n} T_m(X_i)\right\}$ является достаточной статистикой. Приравнивая к нулю частные производные $\dfrac{\partial L_n}{\partial u_j}$, получаем систему (1.21), которая в силу положительной определенности матрицы R имеет единственное решение, доставляющее максимум L_n. Теорема доказана.

При выполнении некоторых условий регулярности ММП-оценки состоятельны, а именно: если (1.1) — повторная выборка, то $\widehat{u}_{\text{ММП}} \to u_T$ по вероятности при $n \to \infty$. Более того, для любого $\varepsilon > 0$ существуют такие K, $\alpha > 0$, что $\mathbf{P}\{|\widehat{u}_{\text{ММП}} - u_T| \geqslant \varepsilon\} \leqslant Ke^{-\alpha n}$ для всех $n \geqslant 1$.

Ранее при рассмотрении примеров было отмечено, что ММП-оценки являются либо эффективными, либо близки к нижней границе Фишера–Крамера–Рао при больших n. Оказывается, что и в общем случае при выполнении условий регулярности ММП-оценки являются асимптотически эффективными и асимптотически несмещенными [33].

Еще одно замечательное свойство ММП-оценок связано с их асимптотической нормальностью.

Последовательность случайных величин a_n называют *асимптотически нормальной*, если найдутся такие детерминированные невырожденные матрицы A_n и векторы μ_n, что закон распределения величин $A_n(a_n - \mu_n)$ стремится к нормальному $\mathcal{N}(0, K_a)$ при $n \to \infty$, где K_a — положительно определенная $(m \times m)$-матрица.

При выполнении условий регулярности закон распределения последовательности $\sqrt{n}\,(\widehat{u}_{\text{ММП}} - u_T)$ стремится к нормальному

$\mathcal{N}(0, I^{-1}(u_T))$ при $n \to \infty$, где $I^{-1}(u_T)$ — матрица, обратная к информационной матрице Фишера. Используя свойство асимптотической нормальности ММП-оценок, после нахождения точечных ММП-оценок можно строить доверительные интервалы для искомых параметров. Действительно, при достаточно большом n закон распределения $\widehat{u}_{\text{ММП}} - u_T$ близок к нормальному,

$$\mathcal{N}(0, \frac{1}{\sqrt{n}}\, I^{-1}(\widehat{u}_{\text{ММП}})).$$

В частности, закон распределения $\widehat{u}_{j\text{ММП}} - u_{jT}$ близок к одномерному нормальному,

$$\mathcal{N}(0, \frac{1}{\sqrt{n}}\, I_{jj}^{-1}(\widehat{u}_{\text{ММП}})),$$

где $\widehat{u}_{j\text{ММП}}$ — j-я компонента вектора $\widehat{u}_{\text{ММП}}$, $I_{jj}^{-1}(\widehat{u}_{\text{ММП}})$ — j-й элемент главной диагонали матрицы $I^{-1}(\widehat{u}_{\text{ММП}})$. Поэтому доверительные интервалы для компонент ММП-оценки $\widehat{u}_{\text{ММП}}$ имеют вид

$$\widehat{u}_{j\text{ММП}} - t(p_{\text{дов}})\sqrt{\frac{I_{jj}^{-1}(\widehat{u}_{\text{ММП}})}{n}} \leqslant u_{jT} \leqslant u_{j\text{ММП}} + t(p_{\text{дов}})\sqrt{\frac{I_{jj}^{-1}(\widehat{u}_{\text{ММП}})}{n}},$$

где $t(p_{\text{дов}})$ при заданной доверительной вероятности $p_{\text{дов}}$ находится по таблице интеграла вероятностей согласно равенству

$$\frac{1}{\sqrt{2\pi}} \int\limits_0^{t(p_{\text{дов}})} e^{-s^2/2}\, ds = \frac{p_{\text{дов}}}{2}.$$

Теоретическое обоснование ММП дано в монографии [33].

Глава 2
МЕТОДЫ УЧЕТА АПРИОРНОЙ ИНФОРМАЦИИ В РАМКАХ ПАРАМЕТРИЧЕСКОЙ СТАТИСТИКИ

При восстановлении значений искомых параметров исследователь располагает, как правило, дополнительной (априорной) информацией о параметрах, помимо информации, «заложенной» в выборке (1.1). До сих пор оценки параметров строились только на основе выборки (1.1) и не учитывали дополнительную информацию о параметрах. Ясно, что учет дополнительной информации позволит улучшить надежность и точность оценивания.

В настоящее время используют два основных метода учета дополнительной информации в зависимости от ее характера — байесовский и минимаксный. Ниже, кроме этих двух методов учета априорной информации, изложен еще один — обобщенный метод максимального правдоподобия.

2.1. Метод Байеса

Применение *байесовского подхода* предполагает трактовку искомого параметра u как случайной величины. Тогда априорная информация задается в виде априорного распределения $p_{\text{апр}}(u)$ случайной величины u.

В большинстве прикладных задач искомый параметр u является детерминированной величиной, и поэтому искусственно навязанная байесовским подходом трактовка детерминированных величин u как величин случайных вызывает многочисленную критику в адрес метода Байеса. Следует однако отметить, что метод Байеса во многих случаях позволяет успешно учесть априорную информацию и получить разумные оценки.

Итак, при байесовском подходе u — случайная величина. Обозначим $p(x_1, \ldots, x_n, u)$ совместную плотность вероятностей совокупности случайных величин $X_1, \ldots, X_n, u$. Имеем

$$p(x_1, \ldots, x_n, u) = p(x_1, \ldots, x_n \mid u) p_{\text{апр}}(u) = \\ = p(u \mid x_1, \ldots, x_n)\, p(x_1, \ldots, x_n),$$

где $p(u \mid x_1, \ldots, x_n)$ — условная плотность вероятностей случайной величины u при фиксированных значениях $x_1, \ldots, x_n$, $p(x_1, \ldots, x_n)$ — маргинальная совместная плотность вероятностей совокупности слу-

чайных величин $X_1, \ldots, X_n$, которую можно найти по формуле

$$p(x_1, \ldots, x_n) = \int\limits_U p(x_1, \ldots, x_n, u)\, du = \int\limits_U p(x_1, \ldots, x_n \mid u) p_{\text{апр}}(u)\, du.$$

Из вышесказанного следует формула

$$p(u \mid x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n \mid u) p_{\text{апр}}(u)}{\int\limits_U p(x_1, \ldots, x_n \mid u) p_{\text{апр}}(u)\, du}.$$

Поскольку имеется одна реализация совокупности случайных величин $X_1, \ldots, X_n$, даваемая выборкой (1.1), то окончательно получаем формулу Байеса

$$p(u \mid X_1, \ldots, X_n) = L_n(X_1, \ldots, X_n;\ u) p_{\text{апр}}(u)/C,$$

где нормировочный множитель

$$C = \int\limits_U L_n(X_1, \ldots, X_n;\ u) p_{\text{апр}}(u)\, du$$

не зависит от u.

Функция $p(u \mid x_1, \ldots, x_n)$ называется *апостериорной плотностью вероятностей*.

В качестве точечной байесовской оценки (Б-оценки) $\widehat{u}_{\text{Б}}$ выбирают одну из характеристик апостериорной плотности вероятностей, например моду, математическое ожидание или медиану апостериорного распределения. Ясно, что в общем случае все три оценки будут различными.

Существует и другой способ получения точечной байесовской оценки, при котором задается положительная функция потерь $W(\|\widehat{u} - u\|)$, где $\widehat{u}$ — оценка параметра u. На основе выбранной функции потерь $W(t)$ и полученной апостериорной плотности $p(u;\ x_1, \ldots, x_n)$ строится функционал

$$R(\widehat{u};\ X_1, \ldots, X_n) = \int\limits_U W(\|\widehat{u} - u\|) p(u \mid X_1, \ldots, X_n)\, du,$$

который принято называть *апостериорным риском* оценки $\widehat{u}$ при заданной выборке (1.1). Тогда Б-оценка определяется как решение экстремальной задачи:

$$\widehat{u}_{\text{Б}} = \arg \min_{\widehat{u}} R(\widehat{u};\ X_1, \ldots, X_n), \tag{2.1}$$

т. е. Б-оценка минимизирует апостериорный риск.

При определении Б-оценки согласно формуле (2.1) также остается неоднозначность байесовской точечной оценки из-за возможности выбрать различные функции потерь $W(t)$. Отметим, что при первом способе обычно в качестве Б-оценки берут моду апостериорного распределения (наиболее вероятное значение апостериорной случайной

величины u); при втором способе чаще всего выбирают квадратичную функцию потерь $W(t) = t^2$. Де Гроот и Рао изучали вид Б-оценок для различных функций потерь. Было установлено, что для $W(t) = |t|$, $W(t) = t^2$ и $W(t) = \eta(t+\delta) - \eta(t-\delta)$, $\delta > 0$ (прямоугольное окно), Б-оценка совпадает соответственно с медианой, математическим ожиданием и модой (для унимодальной и симметричной относительно моды апостериорной плотности) апостериорного распределения. Если апостериорное распределение унимодально и симметрично относительно моды, то все Б-оценки для любой симметричной выпуклой функции потерь совпадают.

Пусть $p_{\text{апр}}(u)$ принадлежат некоторому семейству плотностей P_a. Для нахождения апостериорного распределения важно ответить на вопрос: какому семейству плотностей P_c должна принадлежать совместная плотность вероятностей выборки (1.1), чтобы апостериорная плотность также принадлежала семейству P_a? Семейство P_c называют *сопряженным* семейству P_a. Знание сопряженного семейства значительно облегчает решение задачи нахождения апостериорного распределения, поскольку в большинстве случаев сводит эту задачу к пересчету параметров, определяющих семейство P_a. Ниже приводятся утверждения о виде сопряженных семейств P_c, если основное семейство P_a связано с нормальными распределениями.

Рассмотрим сначала случай одномерного нормального закона распределения.

Теорема 2.1. *Пусть*

$$p_\xi(x;\, u) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-u)^2/(2\sigma^2)},$$

выборка (1.1) *повторная*, *и неизвестно*, *мера точности* $r = 1/\sigma^2$ *задана. Допустим*, *что* $p_{\text{апр}}(u) \sim \mathcal{N}(\mu, 1/\tau)$, *где* $\tau = 1/\sigma_a^2$ — *мера точности. Тогда апостериорное распределение является нормальным* $\mathcal{N}(\mu', 1/(\tau + nr))$ *со средним* $\mu' = \dfrac{\tau\mu + nr\overline{X}}{\tau + nr}$ *и мерой точности* $\tau' = \tau + nr$, *где* $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.

Из теоремы 2.1 следует, что Б-оценки неизвестного параметра u для любой симметричной выпуклой функции потерь совпадают и даются формулой

$$\widehat{u}_{\text{Б}} = \left(\frac{\mu}{\sigma_a^2} + \frac{n\overline{X}}{\sigma^2}\right) \Big/ \left(\frac{1}{\sigma_a^2} + \frac{n}{\sigma^2}\right). \tag{2.2}$$

Из (2.2), в частности, следует, что $\widehat{u}_{\text{Б}} \sim \overline{X}$ при больших n. Последнее означает, что при увеличении объема выборки (1.1) «цена» априорной информации о параметре u падает и Б-оценка, суммирующая информацию о параметре u, «заложенную» в выборке (1.1), и априорную информацию, приближается к ММП-оценке $\widehat{u}_{\text{ММП}}$, построенной только на основе выборки (1.1). Наоборот, если n фиксировано, а $\sigma_a^2 \to 0$, то

$\widehat{u}_{\text{Б}} \to \mu$, т. е. при $\sigma_a^2 \to 0$ «цена» априорной информации неограниченно увеличивается.

Рассмотрим теперь многомерный случай. Пусть все члены повторной выборки (1.1) — m-мерные векторы, распределенные по нормальному закону $X_i \sim \mathcal{N}(u, \sigma^2 K_x)$, где K_x — положительно определенная $(m \times m)$-матрица.

Требуется оценить компоненты вектора $u = (u_1, \ldots, u_m)^T$.

Пусть априорное распределение искомого вектора U — нормальное $\mathcal{N}(u_a, \sigma_a^2 K_a)$, где K_a — положительно определенная $(m \times m)$-матрица.

Тогда апостериорная плотность определяется равенством

$$p(u \mid x_1, \ldots, x_n) = C \exp\Big\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (K_x^{-1}(x_i - u), x_i - u) - \\ - \frac{1}{2\sigma_a^2} (K_a^{-1}(u - u_a), u - u_a)\Big\},$$

где $C = [(2\pi)^{m(n+1)/2} \sigma^{mn} \sigma_a^m |K_x|^{n/2} |K_a|^{1/2}]^{-1}$ — нормировочный множитель.

В качестве байесовской оценки $u_{\text{Б}}$ возьмем моду апостериорного распределения. Тогда получаем равенство

$$u_{\text{Б}} = \Big(\frac{n}{\sigma^2} K_x^{-1} + \frac{1}{\sigma_a^2} K_a^{-1}\Big)^{-1} \Big(\frac{n}{\sigma^2} K_x^{-1} \overline{X} + \frac{1}{\sigma_a^2} K_a^{-1} u_a\Big), \qquad (2.3)$$

где $\overline{X} = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$.

Рассмотрим теперь более сложный случай, когда неизвестны и u, и $r = 1/\sigma^2$. Тогда, согласно методу Байеса, необходимо задавать априорное совместное распределение двух случайных величин (U, R).

Теорема 2.2. *Пусть априорное совместное распределение (U, R) таково: условное распределение U при фиксированном r — нормальное со средним μ и мерой точности τ, а маргинальное распределение R — гамма-распределение с параметрами α $\beta > 0$ $(p(r) = \beta^\alpha r^{\alpha-1} e^{-\beta r}/\Gamma(\alpha))$. Тогда апостериорное совместное распределение (U, R) таково: условное распределение U при фиксированном r — нормальное со средним*

$$\mu' = \frac{\tau\mu + n\overline{X}}{\tau + n},$$

а маргинальное распределение R — гамма-распределение с параметрами

$$\alpha' = \alpha + \frac{n}{2}, \quad \beta' = \beta + \frac{1}{2} \sum_{i=1}^{n} (X_i - \overline{X})^2 + \frac{\tau n (\overline{X} - \mu)^2}{2(\tau + n)}.$$

Если в качестве Б-оценок взять математическое ожидание апостериорного распределения, то с помощью теоремы 2.2 получаем

$$\widehat{u}_{\text{Б}} = \frac{\tau\mu + n\overline{X}}{\tau + n}, \quad \widehat{\sigma}^2_{\text{Б}} = \frac{\beta'}{\alpha'}. \tag{2.4}$$

Из (2.4) следует, что $\widehat{u}_{\text{Б}} \sim \overline{X}$, $\widehat{\sigma}_{\text{Б}} = \widehat{\sigma}^2_{\text{ММП}}$ при $n \to \infty$ и $\widehat{u}_{\text{Б}} \to \mu$ при $\tau \to \infty$.

Пусть $p_\xi(x;\ u)$ — m-мерное нормальное распределение с неизвестным вектором средних u и ковариационной матрицей $\sigma^2 K_x$, где мера точности $r = 1/\sigma^2$ неизвестна, K_x — задана. Матрица, обратная к ковариационной матрице rK_x^{-1}, называется *матрицей точности.*

Теорема 2.3. *Пусть априорное совместное распределение (U, R) таково: условное распределение U при фиксированном r — m-мерное нормальное с вектором средних μ и матрицей точности rK_a^{-1}, а маргинальное распределение R — гамма-распределение с параметрами $\alpha, \beta > 0$. Тогда апостериорное совместное распределение таково: условное распределение U при фиксированном r — m-мерное нормальное с вектором средних $\mu' = (K_a^{-1} + nK_x^{-1})^{-1}\,(K_a^{-1}\mu + nK_x^{-1}\overline{X})$, а маргинальное распределение R — гамма-распределение с параметрами*

$$\alpha' = \alpha + \frac{nm}{2},$$

$$\beta' = \beta + \frac{1}{2}\sum_{i=1}^{n}(X_i - \overline{X})^T K_x^{-1}(X_i - \overline{X}) + \frac{1}{2}\,(\mu' - \mu)^T K_a^{-1}(\overline{X} - \mu).$$

Если, как и прежде, в качестве Б-оценки взять моду апостериорного распределения, то из теоремы 2.3 получаем

$$\widehat{u}_{\text{Б}} = (K_a^{-1} + nK_x^{-1})^{-1}\,(K_a^{-1}\mu + nK_x^{-1}\overline{X}), \quad \widehat{\sigma}^2_{\text{Б}} = \frac{\beta'}{\alpha'}. \tag{2.5}$$

Из (2.5) следует, что

$$\widehat{u}_{\text{Б}} \sim \overline{X}, \quad \widehat{\sigma}^2_{\text{Б}} \sim \frac{1}{nm}\sum_{i=1}^{n}(X_i - \overline{X})^T K_x^{-1}(X_i - \overline{X})$$

при $n \to \infty$.

Замечательным свойством байесовского подхода является возможность получения на основе апостериорного распределения не только точечных оценок параметров, но и доверительных интервалов для оцениваемых параметров. Действительно, например, для оценки (2.2) получаем доверительный интервал

$$\widehat{u}_{\text{Б}} - t(p_{\text{дов}})\frac{1}{\sqrt{\tau'}} \leqslant u_T \leqslant \widehat{u}_{\text{Б}} + t(p_{\text{дов}})\frac{1}{\sqrt{\tau'}}, \tag{2.6}$$

где

$$\tau' = \frac{1}{\sigma_a^2} + \frac{n}{\sigma^2}, \quad \frac{1}{\sqrt{2\pi}} \int\limits_0^{t(p_{\text{дов}})} e^{-s^2/2}\, ds = \frac{p_{\text{дов}}}{2}.$$

При отсутствии априорной информации вместо (2.6) получаем доверительный интервал, длина которого $2t(p_{\text{дов}})\sigma/\sqrt{n}$ больше длины интервала (2.6), равной $2t(p_{\text{дов}})/\tau'$. Таким образом, использование дополнительной априорной информации в рамках баейсовского подхода приводит к уменьшению длины доверительного интервала.

Если в качестве Б-оценки взять моду апостериорного распределения, то

$$\widehat{u}_{\text{Б}} = \arg \max_u \{\ln\, p(X_1, \ldots, X_n \mid u) + \ln\, p_{\text{апр}}(u)\}. \tag{2.7}$$

Пусть априорная информация о параметре u отсутствует. Тогда, согласно постулату Байеса, положим $p_{\text{апр}}(u) = \text{const}$ и Б-оценка (2.7) примет вид $\widehat{u}_{\text{Б}} = \arg \max_u\, l_n(X_1, \ldots, X_n;\, u)$, т. е. Б-оценка будет совпадать с ММП-оценкой, основанной на выборке (1.1).

Теоретические основы метода Байеса представлены в монографиях [26, 33].

2.2. Минимаксный метод учета априорной информации

Часто априорная информация об искомом параметре u задается детерминированно в виде $u \in U_a$, где $U_a \subset E^m$ — априорное множество. Интуитивно ясно, что чем «уже» априорное множество U_a, тем больше имеется априорной информации о параметре. При таком виде априорной информации иногда используют *минимаксный метод*, согласно которому минимаксная оценка (ММ-оценка) $\widehat{u}_{\text{ММ}}$ определяется равенством

$$\widehat{u}_{\text{ММ}} = \arg \min_{\widehat{u} \in U_a} V(\widehat{u}), \tag{2.8}$$

где $V(\widehat{u}) = \max_{u \in U_a} \mathbf{M}\, R(\widehat{u}, u)$, $R(\widehat{u}, u) = W(\|\widehat{u}(X_1, \ldots, X_n) - u\|)$, $\mathbf{M}\, R(\widehat{u}, u)$ — функция риска.

Таким образом, минимаксный метод выбирает такую оценку, которая предохраняет от возможно большего риска при изменении u на априорном множестве U_a.

Справедливо неравенство

$$\max_{u \in U_a} \min_{\widehat{u}} \mathbf{M}\, R(\widehat{u}, u) \leqslant \min_{\widehat{u}} \max_{u \in U_a} \mathbf{M}\, R(\widehat{u}, u),$$

причем ММ-оценка существует, если последнее неравенство обращается в равенство.

Как и Б-оценки, ММ-оценки определены неоднозначно из-за возможности выбирать различные функции потерь $W(t)$. Чаще всего берут квадратичную функцию потерь $W(t) = t^2$.

Иногда при определении ММ-оценок ограничивают класс возможных оценок $\widehat{U}$ (например, выбирают класс линейных оценок). Тогда ММ-оценка определяется согласно равенству

$$\widehat{u}_{\text{ММ}} = \arg \min_{\widehat{u} \in \widehat{U} \cap U_a} V(\widehat{u}). \tag{2.9}$$

Нахождение ММ-оценки для произвольного априорного множества U_a и большой размерности m — трудоемкая вычислительная задача. Если функция потерь квадратичная, то для некоторых частных видов априорного множества и закона распределения выборки (1.1) удается найти явный аналитический вид ММ-оценки.

Пусть $p_\xi(x;\ u)$ — m-мерное нормальное распределение с неизвестным вектором математического ожидания u и известной ковариационной матрицей $\sigma^2 K_\xi$. Пусть априорное множество U_a — m-мерный эллипсоид (с центром в u_0 и «радиуса» ρ, т. е. $(u-u_0)^T B(u-u_0) \leqslant \leqslant \rho^2$, B — симметричная положительно определенная $(m \times m)$-матрица). Выберем квадратичную функцию потерь. Тогда ММ-оценка (2.8) дается равенством

$$\widehat{u}_{\text{ММ}} = u_0 + \frac{1}{\sigma^2}\left(\frac{1}{\rho^2} B' + \frac{n}{\sigma^2} K_\xi^{-1}\right)^{-1} K_\xi^{-1}\left(\sum_{i=1}^{n} X_i - n u_0\right), \tag{2.10}$$

где B' - модифицированная матрица B [79].

При $n \to \infty$ (или при $\sigma^2 \to 0$) $\widehat{u}_{\text{ММ}} \sim \overline{X}$, т. е. ММ-оценка сближается с ММП-оценкой, построенной по выборке (1.1), а «цена» априорной информации уменьшается.

Если $\rho \to \infty$, то ММ-оценка (2.10) также сближается с ММП-оценкой $\widehat{u}_{\text{ММП}} = \overline{X}$ и тем самым «цена» априорной информации уменьшается. Если $\rho \to 0$, т. е. априорное множество становится все у́же, стягиваясь к центру u_0, то $\widehat{u}_{\text{ММ}} \to u_0$, а «цена» выборки (1.1) падает.

Теоретические основы ММ-оценивания представлены в монографии [79].

2.3. Обобщенный метод максимального правдоподобия учета априорной детерминированной информации

Можно предложить еще один подход учета априорной информации детерминированного характера $u \in U_a$, согласно которому оценка искомого параметра определяется равенством

$$\widehat{u}_{\text{ОММП}} = \arg \max_{u \in U_a} L_n(X_1, \ldots, X_n;\ u). \tag{2.11}$$

Такой метод оценивания назовем *обобщенным методом максимального правдоподобия* (ОММП), а оценку (2.11) — *оценкой обобщенного максимального правдоподобия* (ОММП-*оценка*).

Часто вместо представления (2.11) ОММП-оценки удобно использовать эквивалентное представление

$$\widehat{u}_{\text{ОММП}} = \arg \min_{u \in U_a} -l_n(X_1, \ldots, X_n;\ u). \tag{2.12}$$

Если функция $-l_n(X_1, \ldots, X_n;\ u)$ сильно выпукла по $u \in U_a$, U_a — выпуклое множество, то ОММП-оценка существует и единственна.

В общем случае ОММП-оценка может быть найдена с помощью численных методов выпуклого программирования, например с помощью метода сопряженных градиентов, методов штрафных функций, метода градиентного спуска [18].

Для некоторых частных случаев ОММП-оценку (2.12) можно найти в явном аналитическом виде.

Предположим, что функция правдоподобия дифференцируема по u, а априорное множество U_a — m-мерный эллипсоид, т. е. $(u - u_0)^T B(u - u_0) \leqslant \rho^2$. Пусть

$$\widehat{u}_{\text{ММП}} = \arg \min_{u \in E^m} -l_n(X_1, \ldots, X_n;\ u) \in U_a.$$

Тогда $\widehat{u}_{\text{ОММП}} = \widehat{u}_{\text{ММП}}$ и априорное множество (из-за своей «обширности») не содержит дополнительной информации об искомом векторе u. Пусть $\widehat{u}_{\text{ММП}} \notin U_a$. Тогда, используя способ Лагранжа, получаем уравнение Эйлера, решением которого является ОММП-оценка

$$-\operatorname{grad}_u\ l_n(X_1, \ldots, X_n;\ u) + \alpha B(u - u_0) = 0, \tag{2.13}$$

где $\alpha > 0$ — множитель Лагранжа. Величина параметра α находится из равенства $(u - u_0)^T B(u - u_0) = \rho^2$.

Пусть, например, $p_\xi(x;\ u)$ — m-мерное нормальное распределение с неизвестным вектором математических ожиданий u и известной ковариационной матрицей $\sigma^2 K_\xi$, выборка (1.1) повторная. Тогда

$$-l_n(X_1, \ldots, X_n;\ u) =$$

$$= \frac{nm}{2}\ \ln(2\pi\sigma^2) + \frac{n}{2}\ \ln|K_\xi| + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - u)^T K_\xi^{-1} (X_i - u),$$

и уравнение (2.13) принимает вид

$$nK_\xi^{-1} u - K_\xi^{-1} \sum_{i=1}^{n} X_i + \alpha\sigma^2 B(u - u_0) = 0,$$

откуда

$$\widehat{u}_{\text{ОММП}} = u_0 + \frac{1}{\sigma^2}\Big(\alpha B + \frac{n}{\sigma^2}\,K_\xi^{-1}\Big)^{-1} K_\xi^{-1}\Big(\sum_{i=1}^{n} X_i - nu_0\Big). \qquad (2.14)$$

ОММП-оценка (2.14) подобна ММ-оценке (2.10).

Из определения ОММП-оценок следует, что если априорная информация отсутствует ($U_a = E^m$), то ОММП-оценка совпадает с ММП-оценкой. Ясно, что в общем случае ОММП-оценка совпадает с ММП-оценкой тогда и только тогда, когда $\widehat{u}_{\text{ММП}} \in U_a$.

ММ-оценки и ОММП-оценки смещены, причем смещение производится в нужную сторону с приближением оценки к искомому точному значению параметра.

Например, смещения Δu оценок (2.10) и (2.14) соответственно равны

$$\Delta u = \frac{1}{\rho^2}\,(\frac{1}{\rho^2}\,B' + \frac{n}{\sigma^2}\,K_\xi^{-1})^{-1} B'(u_0 - u_T)$$

и

$$\Delta u = \alpha(\alpha B + \frac{n}{\sigma^2}\,K_\xi^{-1})^{-1} B(u_0 - u_T).$$

Следовательно, величины смещений стремятся к нулю в трех случаях: 1° $\rho \to \infty$, 2° $n \to \infty$, 3° $\sigma^2 \to 0$.

Поскольку ММ-оценки и ОММП-оценки смещены, их точность можно оценить с помощью матрицы $D(\widehat{u}) = M(\widehat{u} - u_T)(\widehat{u} - u_T)^T$.

Для ММ-оценки (2.10)

$$D(\widehat{u}_{\text{ММ}}) = \Big(\frac{1}{\rho^2}\,B' + \frac{n}{\sigma^2}\,K_\xi^{-1}\Big)^{-1}. \qquad (2.15)$$

Для ОММП-оценки (2.14)

$$D(\widehat{u}_{\text{ОММП}}) = (\alpha B + \frac{n}{\sigma^2}\,K_\xi^{-1})^{-1}. \qquad (2.16)$$

Из формул (2.15), (2.16) следует состоятельность в среднем квадратичном оценок (2.10), (2.14) при: 1° $\rho \to 0$, 2° $n \to \infty$, 3° $\sigma^2 \to 0$.

Оценки (2.10) и (2.14) линейны и для их вычисления необходимо обратить симметричные положительно определенные матрицы

$$\Big(\frac{1}{\rho^2}\,B' + \frac{n}{\sigma^2}\,K_\xi^{-1}\Big) \quad \text{и} \quad \Big(\alpha B + \frac{n}{\sigma^2}\,K_\xi^{-1}\Big)$$

соответственно, предварительно обратив симметричную положительно определенную матрицу K_ξ. Одним из эффективных численных методов обращения симметричной положительно определенной матрицы A является метод квадратного корня (метод Холецкого), основанный на представлении $A = S^T S$, где S — верхняя треугольная матрица.

Пример 2.1. Для одномерного случая ($m = 1$) эллипсоид $(u - u_0)^T B(u - u_0) \leqslant \rho^2$ представляет собой интервал $|u - u_0| \leqslant d$. Можно доказать, что оценки (2.10), (2.14) принимают вид

$$\widehat{u}_{\text{ММ}} = \frac{d^2}{nd^2 + \sigma^2} \sum_{i=1}^{n} X_i + \Big(1 - \frac{nd^2}{nd^2 + \sigma^2}\Big) u_0,$$

$$\widehat{u}_{\text{ОММП}} = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i, & \Big|\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i - u_0\Big| \leqslant d, \\ u_0 + d \operatorname{sign} \Big(\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i - u_0\Big), & \Big|\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i - u_0\Big| > d. \end{cases}$$

Из последнего выражения для ОММП-оценки видно, что она совпадает с ММП-оценкой, если $\widehat{u}_{\text{ММП}} \in [u_0 - d, u_0 + d]$, и является ближайшей точкой интервала $[u_0 - d, u_0 + d]$ к ММП-оценке, если $\widehat{u}_{\text{ММП}} \notin [u_0 - d, u_0 + d]$. Это правило сохраняется и для общего случая выпуклой функции $-l_n(X_1, \ldots, X_n;\ u)$ и выпуклого множества U_a, а именно: если $\widehat{u}_{\text{ММП}} \in U_a$, то $\widehat{u}_{\text{ОММП}} = \widehat{u}_{\text{ММП}}$, если же $\widehat{u}_{\text{ММП}} \notin U_a$, то $\widehat{u}_{\text{ОММП}}$ является ближайшей точкой априорного множества U_a к ММП-оценке вдоль однопараметрической кривой (2.14), $0 \leqslant \alpha < \infty$.

Поэтому для выпуклой по u логарифмической функции правдоподобия $-l_n(X_1, \ldots, X_n;\ u)$ и выпуклого априорного множества U_a можно предложить следующий алгоритм нахождения ОММП-оценки: 1° найти ММП-оценку; 2° если $\widehat{u}_{\text{ММП}} \in U_a$, то $\widehat{u}_{\text{ОММП}} = \widehat{u}_{\text{ММП}}$; если $\widehat{u}_{\text{ММП}} \notin U_a$, то найти «проекцию» $\widehat{u}_{\text{ОММП}}$ точки $\widehat{u}_{\text{ММП}}$ на априорное множество U_a.

2.4. Обобщенный метод максимального правдоподобия учета априорной стохастической информации

Так же как и в предыдущем разделе, предлагаемый ниже метод учета априорной информации базируется на методе максимального правдоподобия.

Обозначим $L(u) = L_n(X_1, \ldots, X_n;\ u)$ и $l(u) = \ln L(u)$ функцию правдоподобия и логарифмическую функцию правдоподобия исходной выборки (1.1) соответственно.

Пусть наряду с исходной выборкой (1.1) задана априорная выборка

$$Y_1, \ldots, Y_N, \tag{2.17}$$

члены которой независимы от членов выборки (1.1).

Обозначим через $L_a(u) = L_N(Y_1, \ldots, Y_N;\ u)$ и $l_a(u)$ функцию правдоподобия и логарифмическую функцию правдоподобия априорной выборки (2.17), которые зависят от исходных параметров u.

Тогда совместная (обобщенная) функция правдоподобия $L_o(u)$ и совместная логарифмическая функция правдоподобия $l_o(u)$ совокупности членов выборок (1.1), (2.17) определяются с помощью равенств

$$L_o(u) = L(u)\, L_a(u), \tag{2.18}$$

$$l_o(u) = l(u) + l_a(u). \tag{2.19}$$

Следовательно, имеем равенство

$$I_o(u) = I(u) + I_a(u), \tag{2.20}$$

где $I(u)$, $I_a(u)$, $I_o(u)$ — информационные матрицы Фишера исходной выборки (1.1), априорной выборки (2.17) и совместной выборки (1.1), (2.17) соответственно. Из равенства (2.20) следует, что при совместном рассмотрении выборок (1.1) и (2.17) происходит суммирование информации об искомых переменных u, имеющейся в исходной и априорной выборках.

Согласно обобщенному методу максимального правдоподобия (ОММП) при стохастическом виде априорной информации, задаваемой априорной выборкой (2.17), ОММП-оценка искомых параметров u определяется равенством

$$\widehat{u}_{\text{ОММП}} = \arg \max_u L_o(u). \tag{2.21}$$

В частности, если совместная плотность вероятности членов априорной выборки (2.17) не зависит от искомых параметров u, то в априорной выборке отсутствует информация о векторе u, $I_a(u) = 0$ — нулевая матрица и оценка (2.21) совпадает с ММП-оценкой (1.17), определяемой только по исходной выборке (1.1).

Если наряду с исходными параметрами u имеются дополнительные параметры $v = (v_1, \ldots, v_r)^T$, числовые значения которых неизвестны, например некоторые характеристики законов распределения членов выборки (1.1), (2.17), то вместо ОММП-оценки (2.21) (только параметров u) необходимо рассматривать совместную ОММП-оценку всех параметров u, v:

$$(\widehat{u}_{\text{ОММП}}, \widehat{v}_{\text{ОММП}}) = \arg \max_{(u,v)} L_o(u, v). \tag{2.22}$$

Если на u и v не накладывается никаких априорных ограничений, то (2.21), (2.22) — задачи на безусловный экстремум. В частности, если обобщенные функции правдоподобия $L_o(u)$, $L_o(u, v)$ дифференцируемы, то ОММП-оценки (2.21), (2.22) могут быть найдены как решения обобщенных уравнений правдоподобия

$$\operatorname{grad}_u L_o(u) = 0, \tag{2.23}$$

$$\operatorname{grad}_u L_o(u, v) + \operatorname{grad}_v L_o(u, v) = 0, \tag{2.24}$$

соответственно.

Если на u, v накладываются априорные ограничения (см. ниже), то решение экстремальных задач (2.21), (2.22) необходимо искать при выполнении этих ограничений. В этом случае имеем задачи на условный экстремум и необходимо использовать численные методы решения экстремальных задач на условный экстремум [18]. В частности, иногда некоторая часть параметров u (или u, v) по их содержательному смыслу неотрицательна. Тогда вместо задач на безусловный экстремум (2.21), (2.22) будем иметь задачи на условный экстремум:

$$\widehat{u}_{\text{ОММП}} = \arg \max_{u \geqslant 0} L_o(u), \tag{2.25}$$

$$(\widehat{u}_{\text{ОММП}}, \widehat{v}_{\text{ОММП}}) = \arg \max_{u,v \geqslant 0} L_o(u, v), \tag{2.26}$$

соответственно.

Рассмотрим важный частный случай, когда члены выборок (1.1), (2.17) распределены по нормальному закону и независимы, причем $X_i \sim \mathcal{N}(u, \sigma^2 K_x)$, $Y_j \sim \mathcal{N}(u, \sigma_a^2 K_a)$, где K_x, K_a — положительно определенные матрицы.

В этом случае ОММП-оценка (2.21) дается равенством

$$\widehat{u}_{\text{ОММП}} = \left(\frac{n}{\sigma^2} K_x^{-1} + \frac{N}{\sigma_a^2} K_a^{-1}\right)^{-1} \left(\frac{n}{\sigma^2} K_x^{-1} \overline{X} + \frac{N}{\sigma_a^2} K_a^{-1} \overline{Y}\right), \tag{2.27}$$

где $\overline{X} = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$, $\overline{Y} = \dfrac{1}{N} \sum\limits_{i=1}^{N} Y_j$. В частности, если $N = 1$, то оценка (2.27) принимает вид

$$\widehat{u}_{\text{ОММП}} = \left(\frac{n}{\sigma^2} K_x^{-1} + \frac{1}{\sigma_a^2} K_a^{-1}\right)^{-1} \left(\frac{n}{\sigma^2} K_x^{-1} \overline{X} + \frac{1}{\sigma_a^2} K_a^{-1} Y\right). \tag{2.28}$$

Сравнение (2.28) с (2.3) показывает их полную идентичность. Таким образом, предлагаемая в этом параграфе схема ОММП-оценивания позволяет в случае стохастического характера априорной информации избежать искусственно навязанной схемой Байеса трактовки вектора u как случайной величины и получить оптимальную оценку u без потери имеющейся информации об искомых параметрах.

Например, ОММП-оценка (2.27) несмещена и удовлетворяет неравенству Фишера–Крамера–Рао

$$K_{\widehat{u}_{\text{ОММП}}} = I_o^{-1},$$

где обобщенная информационная матрица Фишера $I_o = I + I_a$, $I = (n/\sigma^2) K_x^{-1}$ — информационная матрица Фишера параметров u для исходной выборки (1.1), $I_a = (N/\sigma_a^2) K_a^{-1}$ — информационная матрица Фишера параметров u для априорной выборки (2.17).

Часто в распоряжении исследователя имеется одновременно априорная информация об исходных параметрах u как детерминированного характера, задаваемая в виде принадлежности $u \in R_a$, так и стохасти-

ческого характера, задаваемая в виде априорной выборки (2.17). Тогда ОММП-оценка параметров u определяется равенством

$$\widehat{u}_{\text{ОММП}} = \arg \max_{u \in R_a} L_o(u). \tag{2.29}$$

Для решения задачи (2.29) можно использовать численные методы решения задач на условный экстремум [18]. В частности, если $-l_o(u)$ — выпуклая функция, а R_a — выпуклое множество, то задача (2.29) принадлежит к классу задач выпуклого программирования.

Оценка (2.29) обладает оптимальными свойствами и позволяет наиболее эффективно использовать информацию о параметрах u, «заложенную» в исходной выборке (1.1), априорной выборке (2.17) и в задании априорного множества R_a.

Методы учета априорной информации в рамках параметрической статистики представлены также в работах [42, 43].

Глава 3
УСТОЙЧИВЫЕ МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРА ПОЛОЖЕНИЯ

ММП-оценивание и связанные с ним методы оценивания с учетом априорной информации базируются на точном знании закона распределения выборки (1.1) и дают достаточно эффективные (в различных смыслах для различных методов) оценки параметров. Однако при решении практических задач идеализированная ситуация точного подчинения членов выборки гипотетическому распределению $p(x_1, \ldots, x_n;\ u)$ выполняется редко. Практически всегда наблюдается уклонение от гипотетического распределения и в этих случаях получаемые оценки уже не будут обладать оптимальными свойствами. Спрашивается: не будут ли получаемые оценки слишком далеки от истинных значений, если даже уклонение истинного закона распределения выборки (1.1) от гипотетического в определенном смысле мало? К сожалению, приходится констатировать, что для многих широко используемых стандартных гипотетических распределений даже малое уклонение истинного закона распределения выборки (1.1) от гипотетического может приводить к значительным уклонениям полученных оценок от истинных значений. Например, для нормального закона распределения, наиболее часто принимаемого в качестве гипотетического, наличие даже небольшой части членов выборки, значительно уклоняющихся от среднего значения (большие выбросы), приводит к большому искажению оценок и к катастрофическому падению их эффективности. Можно задать вопрос: а зачем брать гипотетическое распределение? Возьмем истинный закон распределения выборки (1.1) и, применяя, например, ММП-оценивание, получим эффективные оценки. Но в том и заключается трудность, что истинный закон распределения выборки (1.1) исследователю не известен. Известна, возможно, лишь какая-то неполная информация об истинном законе распределения. Например, при наличии больших выбросов часто известна доля членов выборки (1.1), принадлежащих выбросам, и известно, что сами выбросы располагаются симметрично относительно среднего значения. Иногда информация об истинном законе распределения может быть еще более скудной.

Математически задача оценивания в условиях неточного знания истинного закона распределения выборки (1.1) ставится следующим образом. Задается класс плотностей $\{p(x_1, \ldots, x_n;\ u)\}$, которому принадлежит неизвестная истинная плотность выборки (1.1). Требуется в этом классе выбрать такую гипотетическую плотность $p_0(x_1, \ldots, x_n;\ u)$, чтобы получить гарантированную и одновременно

наиболее близкую к истинному значению параметра оценку. Гарантия оценки понимается в следующем смысле: точность оценки, полученной на основе выбранной гипотетической плотности, не меньше точности любой другой оценки вне зависимости от того, какова истинная плотность из класса $\{p(x_1,\dots,x_n;\ u)\}$. Плотность $p_0(x_1,\dots,x_n;\ u)$ называют *устойчивой плотностью относительно класса* $\{p(x_1,\dots,x_n;\ u)\}$, а саму оценку $\widehat{u}$ — *робастной.*

3.1. Минимаксный метод Хьюбера

Рассмотрим случай, когда X_i, u — скаляры, выборка (1.1) повторная, $p_\xi(x;\ u) = p_\xi(x-u)$, $p_\xi(s)$ — четная функция. Таким образом, класс $\{p(x_1,\dots,x_n;\ u\}$ определяется классом $\{p(x)\}$, где $p(x-u)$ симметричны относительно искомого параметра u (параметра сдвига). Дополнительные условия, накладываемые на класс $\{p(x)\}$, будут даны ниже.

Пусть в классе $\{p(x)\}$ выбрана гипотетическая плотность $p_0(x)$. Тогда ММП-оценка определится равенством

$$\widehat{u} = \widehat{u}(X_1,\dots,X_n;\ p_0) = \arg\max_u\ l_n(X_1,\dots,X_n;\ u), \tag{3.1}$$

где

$$l_n(X_1,\dots,X_n;\ u) = \sum_{i=1}^{n} \ln\ p_0(X_i-u).$$

Обозначим $p_u(x)$ истинную плотность из класса $\{p(x)\}$. В качестве меры точности оценки возьмем дисперсию оценки. Дисперсия оценки (3.1) $\mathbf{D}\,\widehat{u}$ зависит от выбранной гипотетической и истинной плотностей, $p_0(x)$ и $p_u(x)$, и определяется равенством

$$\mathbf{D}\,\widehat{u} = D(p_0,p_u) =$$

$$= \int_{-\infty}^{\infty}\dots\int (\widehat{u}(x_1,\dots,x_n;\ p_0)-u)^2 p_u(x_1-u)\dots p_u(x_n-u)\,dx_1\dots dx_n.$$

Построим устойчивый метод оценивания параметра сдвига u на основе минимаксного подхода, согласно которому устойчивая (робастная) оценка (3.1) находится по такой устойчивой относительно класса $\{p(x)\}$ плотности $p_0(x)$, для которой

$$p_0(x) = \arg\min_{p_o\in\{p(x)\}} D(p_0), \tag{3.2}$$

где $D(p_0) = \max_{p_u\in\{p(x)\}} D(p_0,p_u)$.

Из сказанного выше следует, что робастная оценка параметра сдвига является ММП-оценкой при такой устойчивой гипотетической плотности, которая минимизирует максимальное значение дисперсии оценки при любой возможной истинной плотности из класса $\{p(x)\}$. Итак,

выбор устойчивой плотности, согласно минимаксному подходу, приводит к минимальной потере точности получаемой оценки при реализации самой неблагоприятной истинной плотности из класса $\{p(x)\}$.

Обозначим $\varphi(x) = (\ln p_0(x))'$. Робастная ММП-оценка (3.1) удовлетворяет уравнению $\sum\limits_{i=1}^{n} \varphi(X_i - \widehat{u}) = 0$.

В силу состоятельности оценки $\widehat{u}$ для больших n и с учетом симметрии плотностей имеем

$$\sum_{i=1}^{n} \varphi(X_i - \widehat{u}) \approx \sum_{i=1}^{n} \varphi(X_i - u) - (\widehat{u} - u) \sum_{i=1}^{n} \varphi'(X_i - u).$$

Отсюда

$$\widehat{u} - u \approx \frac{\sum\limits_{i=1}^{n} \varphi(X_i - u)}{\sum\limits_{i=1}^{n} \varphi'(X_i - u)} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} \varphi(X_i - u)}{\frac{1}{n}\sum\limits_{i=1}^{n} \varphi'(X_i - u)} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} \varphi(X_i - u)}{\int\limits_{-\infty}^{\infty} \varphi'(x-u) p_u(x-u)\, dx}.$$

Поэтому с учетом четности $p_0(x)$ и нечетности $\varphi(x)$ получаем

$$D(p_0, p_u) = \int\limits_{-\infty}^{\infty} \dots \int (\widehat{u} - u)^2 p_u(x_1 - u) \dots p_u(x_n - u)\, dx_1 \dots dx_n \approx$$

$$\approx \frac{\int\limits_{-\infty}^{\infty} \dots \int \Big(\sum\limits_{i=1}^{n} \varphi(x_i - u) \Big)^2 p_u(x_1 - u) \dots p_u(x_n - u)\, dx_1 \dots dx_n}{n^2 \Big(\int\limits_{-\infty}^{\infty} \varphi'(x-u) p_u(x-u)\, dx \Big)^2} =$$

$$= \frac{\int\limits_{-\infty}^{\infty} \dots \int \sum\limits_{i=1}^{n} \varphi^2(x_i) p_u(x_1) \dots p_u(x_n)\, dx_1 \dots dx_n}{n^2 \Big(\int\limits_{-\infty}^{\infty} \varphi'(x) p_u(x)\, dx \Big)^2} \approx \frac{\int\limits_{-\infty}^{\infty} \varphi^2(x) p_u(x)\, dx}{n \Big(\int\limits_{-\infty}^{\infty} \varphi'(x) p_u(x)\, dx \Big)^2}.$$

Итак,

$$D(p_0, p_u) \cong \frac{1}{n} \int\limits_{-\infty}^{\infty} \Big(\frac{p_0'(x)}{p_0(x)} \Big)^2 p_u(x)\, dx \Big/ \Big(\int\limits_{-\infty}^{\infty} \Big(\frac{p_0'(x)}{p_0(x)} \Big)' p_u(x)\, dx \Big)^2.$$

Но $\int\limits_{-\infty}^{\infty} (\frac{p_0'(x)}{p_0(x)})' p_u(x)\,dx = -\int\limits_{-\infty}^{\infty} \frac{p_0'(x)p_u'(x)}{p_0(x)}\,dx$. Следовательно,

$$D(p_0, p_u) = \frac{1}{n} \int\limits_{-\infty}^{\infty} \Big(\frac{p_0'(x)}{p_0(x)}\Big)^2 p_u(x)\,dx \Big/ \Big(\int\limits_{-\infty}^{\infty} \frac{p_0'(x)p_u'(x)}{p_0(x)}\,dx\Big)^2.$$

Используя неравенство Коши–Буняковского, получаем

$$\Big(\int\limits_{-\infty}^{\infty} \frac{p_0'(x)p_u'(x)}{p_0(x)}\,dx\Big)^2 = \Big(\int\limits_{-\infty}^{\infty} \Big(\frac{p_0'(x)}{p_0(x)}\sqrt{p_u(x)}\Big)\Big(\frac{p_u'(x)}{\sqrt{p_u(x)}}\Big)\,dx\Big)^2 \leqslant$$

$$\leqslant \int\limits_{-\infty}^{\infty} \Big(\frac{p_0'(x)}{p_0(x)}\Big)^2 p_u(x)\,dx \cdot \int\limits_{-\infty}^{\infty} \Big(\frac{p_u'(x)}{p_u(x)}\Big)^2 p_u(x)\,dx.$$

Поэтому

$$D(p_0, p_u) \geqslant \Big(n \int\limits_{-\infty}^{\infty} \Big(\frac{p_u'(x)}{p_u(x)}\Big)^2 p_u(x)\,dx\Big)^{-1} \equiv \frac{1}{nI_1},$$

где

$$I_1 = \mathbf{M}\left(\frac{\partial \ln p_u(X_i - u)}{\partial u}\right)^2$$

— информация Фишера о параметре сдвига u для каждого члена выборки (1.1).

С другой стороны, в силу асимптотической эффективности ММП-оценок

$$D(p_u, p_u) \sim \frac{1}{I_n} = \frac{1}{nI_1}$$

при больших n, т. е. при $p_0(x) = p_u(x)$, достигается минимальное значение $D(p_0)$. Поэтому исходная минимаксная задача сводится к экстремальной задаче $\max_{p(x)\in\{p(x)\}} D(p, p) = D(p_y, p_y)$, которая эквивалентна экстремальной задаче

$$\min_{p(x)\in\{p(x)\}} \int\limits_{-\infty}^{\infty} \Big(\frac{p'(x)}{p(x)}\Big)^2 p(x)\,dx. \tag{3.3}$$

Здесь через $p_y(x)$ обозначена устойчивая (робастная) плотность. Для нахождения устойчивой плотности как решения экстремальной задачи (3.3) необходимо указать класс $\{p(x)\}$.

При решении многих прикладных задач удобным классом $\{p(x)\}$, адекватно описывающим реальную ситуацию, является класс H, образованный смесью $H = \{p(x)\colon p(x) = (1-\varepsilon)g(x) + \varepsilon h(x)\}$, где $g(x)$ —

заданная (реперная) плотность, $\varepsilon \in [0,1]$ — заданное число (доля засорения), $h(x)$ — произвольная неизвестная плотность. Предполагается, как и прежде, что $g(x)$, $h(x)$ — четные функции.

Теорема 3.1 (Хьюбер). *Пусть* $-\ln g(x)$ *— дважды дифференцируемая выпуклая функция. Тогда в классе* H *существует устойчивая плотность*

$$p_y(x) = \begin{cases} (1-\varepsilon)g(x_0)\exp\{-\widetilde{K}|x-x_0|\}, & x < x_0, \\ (1-\varepsilon)g(x), & x_0 \leqslant x \leqslant x_1, \\ (1-\varepsilon)g(x_1)\exp\{-\widetilde{K}|x-x_1|\}, & x > x_1, \end{cases} \tag{3.4}$$

где x_0, x_1 *— границы интервала, на котором монотонная функция* $g'(x)/g(x)$ *ограничена по модулю константой* $\widetilde{K}$, *определяемой из условия нормировки плотности*

$$1 = (1-\varepsilon)\int\limits_{x_0}^{x_1} g(x)\,dx + \frac{g(x_0)+g(x_1)}{\widetilde{K}}\,(1-\varepsilon).$$

Доказательство теоремы Хьюбера основано на решении неклассической вариационной задачи

$$\min_{p(x)\in H} \int\limits_{-\infty}^{\infty} \left(\frac{p'(x)}{p(x)}\right)^2 p(x)\,dx.$$

Рассмотрим наиболее часто используемый на практике случай, когда реперный закон распределения нормальный, т. е. $g(x) = (2\pi\sigma^2)^{-1/2}\,e^{-x^2/(2\sigma^2)}$. Имеем: $g'(x)/g(x) = -x/\sigma^2$, откуда $|x| \leqslant \widetilde{K}\sigma^2$ и $x_0 = -\widetilde{K}\sigma^2 = -K\sigma$, $x_1 = \widetilde{K}\sigma^2 = K\sigma$, $K = \widetilde{K}\sigma$,

$$p_y(x) = \begin{cases} \dfrac{(1-\varepsilon)}{\sqrt{2\pi}\,\sigma}\,\exp\left\{\dfrac{K}{\sigma}\,(x+K\sigma) - \dfrac{K^2}{2}\right\}, & x < -K\sigma, \\ \dfrac{(1-\varepsilon)}{\sqrt{2\pi}\,\sigma}\,\exp\left\{-\dfrac{x^2}{2\sigma^2}\right\}, & |x| \leqslant K\sigma, \\ \dfrac{(1-\varepsilon)}{\sqrt{2\pi}\,\sigma}\,\exp\left\{-\dfrac{K}{\sigma}\,(x-K\sigma) - \dfrac{K^2}{2}\right\}, & x > K\sigma, \end{cases} \tag{3.5}$$

где K находится из равенства

$$\sqrt{\frac{\pi}{2}}\,\frac{1}{(1-\varepsilon)} = \int\limits_0^K e^{-s^2/2}\,ds + \frac{1}{K}\,e^{-K^2/2}.$$

Параметр Хьюбера K однозначно определен по доле засорения ε, причем $K(\varepsilon) \to \infty$ при $\varepsilon \to +0$, $K(\varepsilon) \to 0$ при $\varepsilon \to 1-0$, $K(\varepsilon)$ — монотонно убывающая функция. Значения функции $K(\varepsilon)$ табулированы. Например, для $\varepsilon = 0{,}01;\ 0{,}05;\ 0{,}1;\ 0{,}25;\ 0{,}5$ значения параметра K равны

1,945; 1,399; 1,14; 0,766; 0,436 соответственно (данные взяты из работы [83]).

Используя теперь устойчивую относительно класса H гипотетическую плотность $p_y(x)$, найдем робастную оценку (3.1).

Имеем

$$l_n(X_1,\dots,X_n;\ u)=\sum_{i=1}^{n}\ln\,p_y(X_i-u)=n\ln\frac{(1-\varepsilon)}{\sqrt{2\pi}\,\sigma}+\frac{n_+ + n_-}{2}K^2-$$

$$-\frac{K}{\sigma}\sum_{i\in I_+}(X_i-u)+\frac{K}{\sigma}\sum_{i\in I_-}(X_i-u)-\frac{1}{2\sigma^2}\sum_{i\in I_0}(X_i-u)^2,$$

где множества индексов I_+, I_-, I_0 определены равенствами

$$I_+=\{i:\ X_i-u>K\sigma\},\quad I_-=\{i:\ X_i-u<-K\sigma\},$$

$$I_0=\{i:\ |X_i-u|\leqslant K\sigma\},$$

n_+, n_- — количество индексов i, принадлежащих множествам I_+, I_- соответственно. Поэтому уравнение правдоподобия $\partial l_n/\partial u=0$ имеет вид $n_+K\sigma-n_-K\sigma+\sum\limits_{i\in I_0}(X_i-\widehat{u})=0$. Последнее равенство перепишем в виде

$$\widehat{u}=\frac{1}{n}\Big[\sum_{i\in\widehat{I}_0}X_i+(\widehat{n}_+ +\widehat{n}_-)\widehat{u}+K\sigma(\widehat{n}_+-\widehat{n}_-)\Big],\tag{3.6}$$

где

$$\widehat{I}_0=\{i:\ |X_i-\widehat{u}|\leqslant K\sigma\},\quad \widehat{I}_+=\{i:\ X_i-\widehat{u}>K\sigma\},$$

$$\widehat{I}_-=\{i:\ X_i-\widehat{u}<-K\sigma\},$$

$\widehat{n}_+$, $\widehat{n}_-$ — количество индексов i, принадлежащих множествам $\widehat{I}_+$, $\widehat{I}_-$ соответственно.

Для решения нелинейного уравнения (3.6) удобно использовать итерационный процесс

$$\widehat{u}_{j+1}=\frac{1}{n}\Big[\sum_{i\in I_{0j}}X_i+(n_{+j}+n_{-j})\widehat{u}_j+K\sigma(n_{+j}-n_{-j})\Big],\tag{3.7}$$

где

$$I_{0j}=\{i:\ |X_i-\widehat{u}_j|\leqslant K\sigma\},\quad I_{+j}=\{i:\ X_i-\widehat{u}_j>K\sigma\},$$

$$I_{-j}=\{i:\ X_i-\widehat{u}_j<-K\sigma\},$$

n_{+j}, n_{-j} — количество индексов i, принадлежащих множествам I_{+j}, I_{-j} соответственно. В качестве нулевого приближения $\widehat{u}_0$ можно брать среднюю арифметическую $\widehat{u}_0=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$.

Итерационный процесс (3.7) легко программируется и быстро сходится за конечное число шагов. Практически сходимость обеспечивается за 1–10 итераций при объеме выборки до 1000 элементов и проценте засорения до 20 % ($\varepsilon = 0{,}2$).

Если доля засорения стремится к 1 и тем самым K стремится к нулю, оценка Хьюбера (3.6) стремится к медиане выборки (1.1). Если засорение стремится к нулю и тем самым K неограниченно увеличивается, оценка Хьюбера (3.6) стремится к средней арифметической $\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.

При решении прикладных задач измерения являются неравноточными. В этих случаях члены X_i выборки (1.1) имеют различные дисперсии σ_i^2. Причиной же засорения (наличия больших выбросов) является ошибочное завышение точности (занижение значений дисперсий) для ряда измерений. Причем номера членов выборок с заниженными дисперсиями неизвестны. Минимаксный метод Хьюбера позволяет снизить вредное влияние измерений с ошибочными дисперсиями и указать эти измерения.

В описанной ситуации для робастной оценки (3.1) логарифмическая функция правдоподобия дается равенством

$$l_n(X_1, \ldots, X_n;\ u) = \sum_{i=1}^{n} \ln\, p_y\left(\frac{X_i - u}{\sigma_i}\right),$$

где $p_y(x)$ — устойчивая плотность (3.5) с $\sigma = 1$. Поэтому

$$l_n(X_1, \ldots, X_n;\ u) = \sum_{i=1}^{n} \ln \frac{(1-\varepsilon)}{\sqrt{2\pi}\,\sigma_i} + \frac{n_+ + n_-}{2} K^2 - K \sum_{i \in I_+} \frac{X_i - u}{\sigma_i} +$$

$$+ K \sum_{i \in I_-} \frac{X_i - u}{\sigma_i} - \frac{1}{2} \sum_{i \in I_0} \frac{(X_i - u)^2}{\sigma_i^2},$$

где

$$I_+ = \{i:\ X_i - u > K\sigma_i\}, \quad I_- = \{i:\ X_i - u < -K\sigma_i\},$$

$$I_0 = \{i:\ |X_i - u| \leqslant K\sigma_i\},$$

и вместо (3.7) получаем итерационный процесс

$$\widehat{u}_{j+1} = \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} \Big[\sum_{i \in I_{0j}} \frac{X_i}{\sigma_i^2} + \sum_{i \in I_{+j}} \frac{K\sigma_i + \widehat{u}_j}{\sigma_i^2} + \sum_{i \in I_{-j}} \frac{-K\sigma_i + \widehat{u}_j}{\sigma_i^2}\Big], \tag{3.8}$$

где

$$I_{0j} = \{i:\ |X_i - \widehat{u}_j| \leqslant K\sigma_i\}, \quad I_{+j} = \{i:\ X_i - \widehat{u}_j > K\sigma_i\},$$

$$I_{-j} = \{i:\ X_i - \widehat{u}_j < -K\sigma_i\}.$$

В качестве нулевого приближения можно брать средневзвешенную оценку $\widehat{u}_0 = \sum\limits_{i=1}^{n} \dfrac{X_i}{\sigma_i^2} \left(\sum\limits_{i=1}^{n} \dfrac{1}{\sigma_i^2} \right)^{-1}$.

Нахождение устойчивой минимаксной оценки Хьюбера относительно класса H предполагает знание доли засорения ε. Часто доля засорения неизвестна. Тогда можно рекомендовать схему устойчивого оценивания параметра сдвига с одновременным подбором подходящего значения доли засорения. Согласно этой схеме подбирается такое значение ε, при котором достигается наилучшее выполнение равенства $\varepsilon = (n_+ + n_-)/n$. Практика численной обработки экспериментальных данных показала, что выбор значения $K = 1{,}8$ дает вполне удовлетворительные результаты при изменении доли засорения в промежутке [0,01; 0,2].

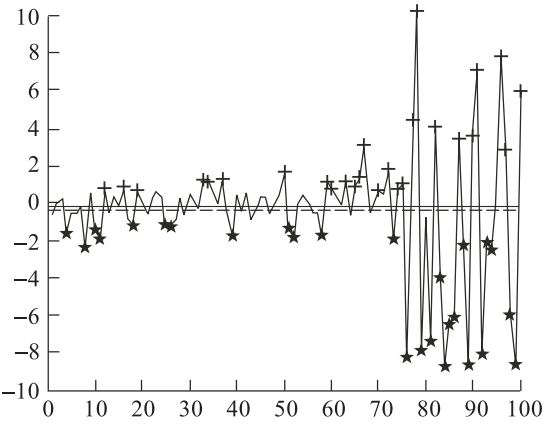

Рис. 3.1

На рис. 3.1 представлены результаты, иллюстрирующие применение итерационного процесса (3.7) на выборке из 100 элементов с коэффициентом засорения $\varepsilon = 0.5$. В качестве реперного закона взят нормальный $N(0,1)$, засорение осуществлено добавлением равномерно распределенных на интервале $[-10, 10]$ величин. Горизонтальная пунктирная линия соответствует нулевому приближению $\widehat{u}_0$, сплошная линия — найденной в результате вычислений оценке. Символами (+, ★) помечены значения с $i \in I_+$ и $i \in I_-$ соответственно.

3.2. Робастные М-оценки

Существуют методы устойчивого оценивания, отличные от минимаксного подхода Хьюбера. Одним из эффективных методов борьбы с большими выбросами и получения устойчивой оценки является оце-

нивание параметра сдвига с помощью усеченной средней

$$u = \overline{X}_{yc} = \frac{1}{n-2m} \sum_{i=m+1}^{n-m} X_{(i)}, \tag{3.9}$$

где $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ — упорядоченные по возрастанию члены выборки (1.1), $m < [n/2]$. Из (3.9) следует, что усеченная средняя получена отбрасыванием m самых левых и m самых правых членов выборки (1.1) и усреднением $n-2m$ оставшихся. Для получения усеченной средней необходимо задать число m. Если известна доля засорения ε, то можно взять $m = [\varepsilon n/2]$.

Модификацией усеченной средней является оценка Винзора

$$\widehat{u} = \overline{X}_{\mathrm{B}} = \frac{1}{n}\Big[\sum_{i=m+1}^{n-m} X_{(i)} + m(X_{(m+1)} + X_{(n-m)})\Big]. \tag{3.10}$$

Из (3.10) следует, что в отличие от усеченной средней в оценке Винзора $2m$ крайних членов выборки не отбрасываются, а m левых крайних проецируются в точку $X_{(m+1)}$, m правых крайних проецируются в точку $X_{(n-m)}$.

Оценка Винзора (3.10) подобна минимаксной оценке Хьюбера (сравните (3.10) с (3.6)). Отличие состоит в том, что проецирование самых левых и правых членов выборки в оценке Хьюбера производится по отношению к самой оценке, а в оценке Винзора проецирование производится до нахождения оценки и априори известно (при заданном m). При увеличении m «степень робастности» усеченной средней и оценки Винзора увеличивается.

Усеченная средняя и оценка Винзора при $m = [n/2]$ совпадают с медианой выборки (1.1).

Обобщением усеченной средней и оценки Винзора на случай неравноточных измерений являются робастные оценки

$$\widehat{u} = \sum_{i=m+1}^{n-m} \frac{X^{(i)}}{\sigma^2_{(i)}} \Big(\sum_{i=m+1}^{n-m} \frac{1}{\sigma^2_{(i)}} \Big)^{-1}, \tag{3.11}$$

$$\widehat{u} = \Big[\sum_{i=m+1}^{n-m} \frac{X^{(i)}}{\sigma^2_{(i)}} + m\Big(\frac{X^{(m+1)}}{\sigma^2_{(m+1)}} + \frac{X^{(n-m)}}{\sigma^2_{(n-m)}} \Big)\Big] \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \Big)^{-1}, \tag{3.12}$$

где $(X^{(1)} - \widehat{u})/\sigma_{(1)} \leqslant (X^{(2)} - \widehat{u})/\sigma_{(2)} \leqslant \ldots \leqslant (X^{(n)} - \widehat{u})/\sigma_{(n)}$ — упорядоченная по возрастанию последовательность чисел $\{(X_i - \widehat{u})/\sigma_i$, $i = 1, \ldots, n\}$.

Средневзвешенная оценка Винзора (3.12) подобна минимаксной оценке из (3.8). Отличие состоит в том, что в оценке (3.12) число $2m$ проецируемых точек и сами точки проекции фиксируются.

Оценки (3.11), (3.12) нелинейны и могут быть найдены с помощью итерационных процессов:

$$\widehat{u}_{j+1} = \sum_{i=m+1}^{n-m} \frac{X_j^{(i)}}{\sigma_{(i)}^2} \Big(\sum_{i=m+1}^{n-m} \frac{1}{\sigma_{(i)}^2} \Big)^{-1},$$

$$\widehat{u}_{j+1} = \Big[\sum_{i=m+1}^{n-m} \frac{X_j^{(i)}}{\sigma_{(i)}^2} + m \Big(\frac{X_j^{(m+1)}}{\sigma_{(m+1)}^2} + \frac{X_j^{(n-m)}}{\sigma_{(n-m)}^2} \Big) \Big] \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \Big)^{-1},$$

где $(X_j^{(1)} - \widehat{u}_j)/\sigma_{(1)} \leqslant (X_j^{(2)} - \widehat{u}_j)/\sigma_{(2)} \leqslant \ldots \leqslant (X_j^{(n)} - \widehat{u}_j)/\sigma_{(n)}$.

В качестве нулевого приближения можно взять средневзвешенную оценку $\widehat{u}_0 = \sum\limits_{i=1}^{n} \frac{X_i}{\sigma_i^2} \Big(\sum\limits_{i=1}^{n} \frac{1}{\sigma_i^2} \Big)^{-1}$.

Широкий класс робастных оценок можно построить с помощью следующего подхода. Обозначим

$$-\ln\, p(X_i;\, u) = \rho\Big(\frac{X_i - u}{\sigma_i} \Big) + b_i,$$

где b_i не зависят от u. Тогда ММП-оценка минимизирует сумму $\sum\limits_{i=1}^{n} \rho\Big(\frac{X_i - u}{\sigma_i} \Big)$. В частности, для минимаксной робастной оценки Хьюбера из (3.8) имеем

$$\rho(s) = \begin{cases} s^2/2, & |s| \leqslant K, \\ K|s| - K^2/2, & |s| > K. \end{cases}$$

Будем выбирать такие функции $\rho(s)$, которые позволят снизить влияние больших выбросов и тем самым определят робастную оценку. Функция Хьюбера $\rho(s)$ обладает этим свойством, поскольку за пределами промежутка $|s| \leqslant K$ квадратичная зависимость заменена на более слабую — линейную.

Пусть задана четная выпуклая положительная функция $\rho(s)$, $\rho(0) = 0$. М-*оценкой* называется значение u, которое минимизирует функционал $\sum\limits_{i=1}^{n} \rho\Big(\frac{X_i - u}{\sigma_i} \Big)$,

$$\widehat{u}_{\mathrm{M}} = \operatorname{Arg}\min_u \sum_{i=1}^{n} \rho\Big(\frac{X_i - u}{\sigma_i} \Big). \tag{3.13}$$

М-оценка однозначно определяется функцией $\rho(s)$.

Чтобы М-оценка была робастной по отношению к большим выбросам, если реперный закон распределения нормальный, необходим более медленный рост функции $\rho(s)$ по сравнению с квадратичной функцией, по крайней мере для больших $|s|$, и (желательно) квадратичная зависимость для малых $|s|$.

Пример 3.1. М-оценки, порожденные функцией $\rho(s) = |s|^p$, где $p \geqslant 1$, называются *L_p-оценками*. Если $p \in [1, 2)$, то L_p-оценки будут робастными, причем «степень робастности» повышается при уменьшении p вплоть до 1. L_p-оценка при $p = 1$ совпадает с медианой выборки (1.1) (в случае равноточных измерений).

Пример 3.2. Робастная оценка, порожденная функцией

$$\rho(s) = \begin{cases} a\left(1 - \cos\dfrac{s}{a}\right), & |s| \leqslant \pi a, \\ 2a, & |s| > \pi a, \end{cases},$$

называется *оценкой Андрюса*. Значение a рекомендуется выбирать так, чтобы точка перегиба $\rho(a)$ равнялась 3 (согласно правилу «трех сигм»). При таком способе выбора $a = 6/\pi \approx 1{,}91$. Для малых $|s|$ $\rho(s) \sim s^2$. При уменьшении a «степень робастности» оценки Андрюса увеличивается.

Пример 3.3. Робастная оценка, порожденная функцией

$$\rho(s) = \frac{1}{\lambda^2}\left[1 - (1 + \lambda|s|)e^{-\lambda|s|}\right],$$

$\lambda > 0$, называется *оценкой Рамсея*. Для малых $|s|$ $\rho(s) \sim s^2$; а $\rho(s) \to$ $\to 1/\lambda^2$ при $|s| \to \infty$. Параметр λ выберем так, чтобы точка перегиба функции $\rho(s)$ равнялась 3. Тогда $\lambda = 1/3$. При увеличении λ «степень робастности» оценки Рамсея увеличивается.

Если функция $\rho(s)$ отлична от квадратичной, то М-оценка нелинейна. Для нахождения М-оценок существует несколько эффективных численных итерационных алгоритмов.

Запишем минимизируемый функционал правой части (3.12) в виде

$$\sum_{i=1}^{n}(X_i - u)^2 W_i, \quad \text{где} \quad W_i = \frac{1}{(X_i - u)^2}\,\rho\left(\frac{X_i - u}{\sigma_i}\right).$$

Если веса W_i фиксированы, то $\widehat{u} = \sum\limits_{i=1}^{n} X_i W_i \cdot \left(\sum\limits_{i=1}^{n} W_i\right)^{-1}$. Определим итерационный процесс

$$\widehat{u}_{j+1} = \sum_{i=1}^{n} X_i W_{ij} \cdot \left(\sum_{i=1}^{n} W_{ij}\right)^{-1}, \quad \text{где} \quad W_{ij} = \frac{1}{(X_i - \widehat{u}_j)^2}\,\rho\left(\frac{X_i - \widehat{u}_j}{\sigma_i}\right).$$

В качестве нулевого приближения можно взять $W_{i0} = 1/\sigma_i^2$, т. е. $\widehat{u}_0$ — средневзвешенная. Этот итерационный процесс называют *итеративным методом наименьших квадратов*.

Если функция $\rho(s)$ дифференцируема, то можно определить еще один численный итерационный метод нахождения М-оценок. М-оценка является решением уравнения

$$\sum_{i=1}^{n} \rho'\left(\frac{X_i - u}{\sigma_i}\right) \cdot \frac{1}{\sigma_i} = 0,$$

которое с учетом четности $\rho(s)$ можно записать в виде

$$\sum_{i=1}^{n}(X_i-u)W_i=0,\quad \text{где}\ \ W_i=\frac{1}{\sigma_i(X_i-u)}\,\rho'\left(\frac{X_i-u}{\sigma_i}\right).$$

Определим итерационный процесс

$$\widehat{u}_{j+1}=\sum_{i=1}^{n}X_iW_{ij}\cdot\left(\sum_{i=1}^{n}W_{ij}\right)^{-1},\quad \text{где}\ \ W_{ij}=\frac{1}{\sigma_i|X_i-\widehat{u}_j|}\,\rho'\left(\frac{|X_i-\widehat{u}_j|}{\sigma_i}\right).$$

В качестве нулевого приближения можно взять

$$\widehat{u}_0=\sum_{i=1}^{n}\frac{X_i}{\sigma_i^2}\cdot\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)^{-1}.$$

Этот итерационный процесс называют *методом вариационно-взвешенных квадратических приближений*. Для L_p-оценок ($\rho(s)=|s|^p$) оба метода совпадают.

Рассмотрим теперь еще один класс плотностей, отличных от класса Хьюбера H. Допустим, что исследователю известно о принадлежности истинной плотности $p_0(x)$ классу симметричных плотностей, сосредоточенных в основном на отрезке $[-d,\,d]$, т. е. $\int\limits_{-d}^{d}p(x)\,dx\geqslant p$, где $p\in(0{,}1)$ задана (класс K_p). С помощью введения класса K_p можно, например, учитывать априорную информацию о степени разброса измеряемых значений относительно среднего значения u.

Теорема 3.2. *Устойчивая относительно класса K_p плотность задается равенством*

$$p_y(x)=\begin{cases}\dfrac{1}{d}\,\dfrac{a}{(1+a)}\cos^2\dfrac{bx}{d}, & |x|\leqslant d,\\[2ex] \dfrac{1}{d}\,\dfrac{a}{(1+a)}\,cos^2\,b\,\exp\left\{-2a\left(\dfrac{|x|}{d}-1\right)\right\}, & |x|>d,\end{cases}\tag{3.14}$$

где a b находятся из системы

$$p=1-\frac{\cos^2 b}{1+b},\quad a=b\operatorname{tg}b,\quad 0<b<\frac{\pi}{2}.$$

Доказательство теоремы 3.2 сводится к решению экстремальной задачи (3.3), где $\{p(x)\}=K_p$.

При уменьшении d и увеличении p «степень робастности» оценки (3.1), основанной на устойчивой плотности (3.14), увеличивается.

Робастные методы оценивания параметра положения представлены в работах [42, 69, 81, 83, 88].

Глава 4
МЕТОДЫ НЕПАРАМЕТРИЧЕСКОЙ СТАТИСТИКИ

Если класс, к которому принадлежит искомое распределение, неизвестен (с точностью до численных значений конечного числа параметров), то для восстановления распределения используют непараметрические методы.

Отметим, что задача восстановления плотности вероятностей $p_\xi(x)$ по повторной выборке (1.1) без учета дополнительной априорной информации о законе распределения является некорректно поставленной (смысл некорректности будет разъяснен ниже).

4.1. Восстановление функции распределения

Достаточно просто по повторной выборке (1.1) дать оценку функции распределения $F(x)$ случайной величины ξ (в дальнейшем, если не сделано специальных оговорок, предполагается, что ξ — скаляр).

В качестве оценки $\widehat{F}(x)$ обычно берут эмпирическую функцию распределения

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \eta(x - X_i), \tag{4.1}$$

где $\eta(x) = \begin{cases} 1, & x > 0, \\ 0, & x < 0, \end{cases}$ — единичная функция.

Теорема 4.1 (Гливенко–Кантелли).

$$\mathbf{P}\left\{ \sup_x \left| F(x) - F_n(x) \right|_{n\to\infty} \to 0 \right\} = 1.$$

Таким образом, согласно теореме Гливенко–Кантелли, эмпирическая функция распределения является состоятельной оценкой искомой функции распределения, причем сходимость $F_n(x)$ к $F(x)$ равномерна по x.

Покажем, что $F_n(x)$ — несмещенная оценка $F(x)$. Действительно,

$$\mathbf{M}\, F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{M}\, \eta(x - X_i) = \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \eta(x - s) p_\xi(s)\, ds =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} p_\xi(s)\, ds = F(x).$$

Докажем состоятельность в среднем квадратичном оценки $F_n(x)$. Действительно,

$$\mathbf{M}\,F_n^2(x) = \mathbf{M}\Big(\frac{1}{n}\sum_{i=1}^{n}\eta(x-X_i)\Big)^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{M}\,\eta(x-X_i)\,\eta(x-X_j) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{M}\,\eta^2(x-X_i) + \frac{1}{n^2}\sum_{i\neq j,\,i,j=1}^{n}\mathbf{M}\,\eta(x-X_i)\,\eta(x-X_j).$$

Имеем

$$\frac{1}{n^2}\sum_{i=1}^{n}\mathbf{M}\,\eta^2(x-X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{M}\,\eta(x-X_i) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\eta(x-s)p(s)\,ds = \frac{1}{n}\,F(x) \to 0$$

при $n\to\infty$. Далее,

$$\frac{1}{n^2}\sum_{i\neq j,\,i,j=1}^{n}\mathbf{M}\,\eta(x-X_i)\,\eta(x-X_j) =$$

$$= \frac{1}{n^2}\sum_{i\neq j,\,i,j=1}^{n}\mathbf{M}\,\eta(x-X_i)\,\mathbf{M}\,\eta(x-X_j) =$$

$$= \frac{1}{n^2}\sum_{i\neq j,\,i,j=1}^{n}\Big(\int_{-\infty}^{\infty}\eta(x-s)p(s)\,ds\Big)^2 = \frac{n(n-1)}{n^2}\,F^2(x) \to F^2(x)$$

при $n\to\infty$. Поэтому дисперсия $\mathbf{D}\,F_n(x) = \mathbf{M}\,F_n^2(x) - F^2(x) \to 0$ при $n\to\infty$.

Теорему Гливенко–Кантелли уточняет неравенство Колмогорова–Смирнова:

$$\mathbf{P}\,\{\sup_x |F(x)-F_n(x)| > \varepsilon\} < 2\exp\{-2\varepsilon^2 n\},$$

справедливое для любого закона распределения при достаточно большом n.

Обозначим $K_n(u)$ функцию распределения случайной величины $U_n = \sqrt{n}\,\max_x |F(x)-F_n(x)|$. Функция $K_n(u)$ не зависит от закона распределения $F(x)$ и выборки (1.1), а определяется только объемом выборки n. Функции $K_n(u)$ для различных значений n и u табулированы [14]. Используя таблицы $K_n(u)$, можно находить доверительные полосы для $F(x)$: $|F(x)-F_n(x)| \leqslant u(p_{\text{дов}})/\sqrt{n}$ для $x\in(-\infty,\infty)$,

где $p_{\text{дов}} = K_n(u(p_{\text{дов}}))$. Существует

$$\lim_{n\to\infty} K_n(u) = K(u) = \sum_{l=-\infty}^{\infty} (-1)^l \exp(-2l^2u^2)$$

(Колмогоров). Функция $K(u)$ также табулирована [14]. Поэтому при большом n (практически, если $n > 100$) для нахождения $u(p_{\text{дов}})$ можно пользоваться равенством $p_{\text{дов}} = K(u(p_{\text{дов}}))$.

Эмпирическая функция распределения $F_n(x)$ является кусочно постоянной. Иногда производят сглаживание $F_n(x)$ с помощью подходящей монотонно возрастающей функции (см. гл. 10). При использовании сглаживания $F_n(x)$ необходимо следить, чтобы оно не приводило к выходу за пределы доверительной полосы.

4.2. Восстановление плотности распределения методом гистограмм

Восстановление плотности вероятностей $p_\xi(x)$ по повторной выборке (1.1) в рамках непараметрического подхода представляет гораздо большие трудности по сравнению с восстановлением функции распределения $F(x)$. Действительно, поскольку $p_\xi(x) = dF(x)/dx$, то формально, получив оценку $\widehat{F}(x)$, можно определить оценку $\widehat{p}_\xi(x) = d\widehat{F}(x)/dx$. К сожалению, задача дифференцирования принадлежит к классу некорректно поставленных задач, поскольку сколь угодно малые изменения в $\widehat{F}(x)$ могут привести к большим изменениям в $\widehat{p}_\xi(x)$. Поэтому любой устойчивый способ оценивания $p_\xi(x)$ должен быть регуляризирующим алгоритмом для задачи дифференцирования (конечно, с учетом того факта, что $p_\xi(x)$ — неотрицательная нормированная функция).

Исторически первым способом оценивания плотности является *метод гистограмм*. Область изменения значений случайной величины разбивается на N интервалов Δx_j (мы сохраняем за интервалом и его длиной одно обозначение). Обозначим n_j количество членов выборки (1.1), попавших в интервал Δx_j, $\sum_{j=1}^{N} n_j = n$, Частоты $\frac{n_j}{n}$ являются оценками вероятности p_j попадания значений ξ в интервал Δx_j, т. е. $p_j = \int\limits_{\Delta x_j} p_\xi(x)\,dx$. Поэтому величины $\frac{n_j}{n\Delta x_j}$ являются оценками среднего значения плотности $p_\xi(x)$ на интервале Δx_j, т. е. $\frac{1}{\Delta x_j}\int\limits_{\Delta x_j} p_\xi(x)\,dx$.

Итак, за оценку плотности принимается

$$\widehat{p}_\Gamma(x) = \frac{n_j}{n\Delta x_j}, \qquad x \in \Delta x_j, \qquad j = 1, \ldots, N. \tag{4.2}$$

Функция $\widehat{p}_\Gamma(x)$ неотрицательна, кусочно постоянна и нормирована, так как

$$\int\limits_{-\infty}^{\infty} \widehat{p}_\Gamma(x)\,dx = \sum_{j=1}^{N} \int\limits_{\Delta x_j} \frac{n_j}{n\Delta x_j}\,dx = \sum_{j=1}^{N} \frac{n_j}{n} = 1.$$

Основной трудностью при построении гистограммы является выбор интервалов разбиения Δx_j. Рекомендуется интервалы Δx_j выбирать так, чтобы в каждый из них попадало одинаковое количество членов выборки (5–10, если n порядка 100). При увеличении n количество членов выборки в каждом интервале необходимо увеличивать. Если обозначить $\Delta x_{\max}$ длину наибольшего интервала, то $\Delta x_{\max}$ играет роль параметра регуляризации при оценивании плотности по гистограмме. При $n \to \infty$ $\Delta x_{\max}$ должен уменьшаться согласованно с ростом n.

Для построения доверительного интервала оценки (4.2) составим статистику

$$\chi^2 = n \sum_{j=1}^{N} \frac{(p_j - n_j/n)^2}{p_j}, \tag{4.3}$$

закон распределения которой при $n \to \infty$ стремится к распределению χ^2_{N-1} (теорема 5.1).

Статистику (4.3) можно представить в эквивалентном виде

$$\chi^2 = n \sum_{j=1}^{N} \Big(\frac{n_j}{n\Delta x_j} - \frac{p_j}{\Delta x_j}\Big)^2 \Big(\frac{p_j}{\Delta x_j}\Big)^{-1} \Delta x_j.$$

Поэтому доверительная полоса для оценки (4.2) имеет вид

$$\sum_{j=1}^{N} \left(\frac{n_j}{n\Delta x_j} - \frac{p_j}{\Delta x_j}\right)^2 \left(\frac{p_j}{\Delta x_j}\right)^{-1} \Delta x_j < \frac{t_N(p_{\text{дов}})}{n},$$

где

$$p_{\text{дов}} = \int\limits_0^{t_N(p_{\text{дов}})} p_{\chi^2_{n-1}}(t)\,dt.$$

Существуют стандартные программы построения гистограмм.

4.3. Восстановление плотности распределения методом Розенблатта–Парзена

Дифференцируя эмпирическую функцию распределения, получаем оценку

$$\widehat{p}_{\text{об}}(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - X_i),$$

где $\delta(x)$ — функция Дирака. Эта оценка плотности неприемлема, поскольку является обобщенной функцией. Модифицируем оценку $\widehat{p}_{\text{об}}$, представив ее в виде классической функции за счет «размазывания» обобщенных функций $\delta(x - X_i)$.

Пусть $\{\psi_n(x)\}$ — последовательность таких симметричных плотностей, что $\psi_n(x) \to \delta(x)$ при $n \to \infty$. Определим оценку плотности равенством

$$\widehat{p}(x) = \frac{1}{n} \sum_{i=1}^{n} \psi_n(x - X_i). \tag{4.4}$$

Теорема 4.2. *Оценка* (4.4) *состоятельна в среднем квадратичном в каждой точке x непрерывности $p_\xi(x)$, если:*

1° $\int\limits_{-\infty}^{\infty} \psi_n(x - s)\, p_\xi(s)\, ds \to p_\xi(x)$ *при* $n \to \infty$;

2° $\frac{1}{n} \max_x \psi_n(x) \to 0$ *при* $n \to \infty$.

Доказательство. Имеем

$$\mathbf{M}\,\widehat{p}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{M}\,\psi_n(x - X_i) = \int\limits_{-\infty}^{\infty} \psi_n(x - s) p_\xi(s)\, ds \to p_\xi(x)$$

при $n \to \infty$. Используя независимость случайных величин, образующих повторную выборку (1.1), получаем

$$\mathbf{M}\,\widehat{p}^2(x) = \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbf{M}\,\psi_n(x - X_i)\psi_n(x - X_j) =$$

$$= \frac{1}{n} \int\limits_{-\infty}^{\infty} \psi_n^2(x - s) p_\xi(s)\, ds + \frac{n-1}{n} \left(\int\limits_{-\infty}^{\infty} \psi_n(x - s) p_\xi(s)\, ds \right)^2.$$

Но

$$\frac{1}{n} \int\limits_{-\infty}^{\infty} \psi_n^2(x - s) p_\xi(s)\, ds \leqslant \frac{1}{n} \max_x \psi_n(x) \int\limits_{-\infty}^{\infty} \psi_n(x - s) p_\xi(s)\, ds \to 0$$

при $n \to \infty$, а

$$\frac{n}{n-1} \left(\int\limits_{-\infty}^{\infty} \psi_n(x - s) p_\xi(s)\, ds \right)^2 \to p_\xi^2(x)$$

при $n \to \infty$. Следовательно, $\mathbf{M}\,\widehat{p}^2(x) \to p_\xi^2(x)$ при $n \to \infty$ в каждой точке непрерывности $p_\xi(x)$. Но тогда $\mathbf{D}\,\widehat{p}(x) = \mathbf{M}\,\widehat{p}^2(x) - \mathbf{M}^2\,\widehat{p}(x) \to 0$ при $n \to \infty$ в каждой точке непрерывности $p_\xi(x)$. Теорема доказана.

Можно построить класс плотностей $\{\psi_n(x)\}$, удовлетворяющих условиям теоремы 4.2, по следующей схеме. Пусть $\psi(x)$ — произвольная непрерывная ограниченная плотность и последовательность положительных чисел $\{h_n\}$ удовлетворяет условиям $h_n \to 0$, $h_n^2 n \to \infty$ при $n \to \infty$. Тогда последовательность функций

$$\psi_n(x) = \frac{1}{h_n}\,\psi\left(\frac{x}{h_n}\right)$$

удовлетворяет условиям теоремы 4.2. Если

$$\psi(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2},$$

то последовательность

$$\psi_n(x) = \frac{1}{\sqrt{2\pi}\,h_n}\,e^{-x^2/(2h_n^2)}.$$

В методе Розенблатта–Парзена число h_n играет роль параметра регуляризации. Конструктивные способы выбора подходящего значения h_n при конечном объеме выборки могут быть предложены только на основании дополнительной информации об искомой плотности.

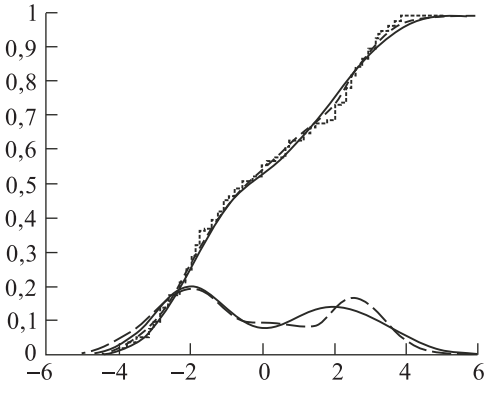

Рис. 4.1

На рис. 4.1 представлен пример восстановления плотности распределения с помощью метода Розенблатта–Парзена по выборке из 100 значений. Для примера использовано двухмодальное распределение из смеси нормальных распределений $\mathcal{N}(-2,1)$ и $\mathcal{N}(2,2)$. Сплошная линия со «ступеньками» соответствует эмпирической функции распределения. Пунктирной линией отмечены восстановленная плотность вероятностей и интеграл от нее. Сплошной линией отмечены аналитические плотность и функция распределения вероятностей.

Метод Розенблатта–Парзена изложен в работах [16, 25, 57, 106, 108].

4.4. Восстановление плотности распределения проекционными методами

Проекционные методы основаны на разложении искомой плотности в ряд по ортогональной системе функций относительно фиксированной реперной плотности.

Пусть известно, что искомая плотность $p_\xi(x)$ достаточно близка к некоторой известной плотности $p_0(x)$, которую будем называть *реперной*. Предположим для определенности, что промежуток изменения возможных значений ξ — вся числовая ось.

Обозначим через $\{q_k(x)\}$ систему полиномов, нормированных на $(-\infty, \infty)$ с весом $p_0(x)$, т. е.

$$\int_{-\infty}^{\infty} q_i(x) q_j(x) p_0(x)\, dx = \delta_{ij},$$

причем $q_0(x) = 1$, $q_k(x)$ — полином степени k. Такая система полиномов определена однозначно и ее можно найти с помощью стандартной процедуры ортогонализации Гильберта–Шмидта.

Обозначим

$$q_k(x) = \sum_{j=0}^{k} a_{kj} x^j$$

и представим искомую плотность в виде

$$p_\xi(x) = p_0(x) \sum_{k=0}^{\infty} C_k q_k(x). \tag{4.5}$$

Имеем

$$C_0 = 1, \quad C_k = \int_{-\infty}^{\infty} q_k(x) p_\xi(x)\, dx.$$

Поэтому

$$C_k = \sum_{j=0}^{k} a_{kj} \int_{-\infty}^{\infty} x^j p_0(x)\, dx = \sum_{j=0}^{k} a_{kj} \alpha_j,$$

где α_j — начальный момент j-го порядка случайной величины ξ.

Если начальные моменты α_j заменить их оценками $\widehat{\alpha}_j$ (например, оценками метода моментов $\widehat{\alpha}_j = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i^j$) и в (4.5) ограничиться ко-

нечным числом слагаемых, то получаем оценку плотности

$$\widehat{p}(x) = p_0(x) \sum_{k=0}^{m} \widehat{C}_k q_k(x), \tag{4.6}$$

где $\widehat{C}_k = \sum_{j=0}^{k} a_{kj} \widehat{\alpha}_j$.

В проекционных методах число членов разложения m играет роль параметра регуляризации. Если случайная величина ξ имеет моменты любых порядков, то принципиально возможно неограниченное увеличение m, которое должно осуществляться согласованно с ростом объема выборки n. При практическом применении оценки (4.6) стараются ограничиться небольшим количествам членов разложения. Ясно, что в этом случае оценка (4.6) будет давать хорошую точность, если только реперная плотность $p_0(x)$ удачно «угадана» и близка к искомой плотности.

Часто проекционную оценку удобно представить через оценки центральных моментов. Тогда вместо случайной величины ξ рассматривают нормированную случайную величину $\xi_0 = (\xi - \alpha_1)/\sigma_\xi$ и проекционная оценка имеет вид

$$\widehat{p}(x) = \frac{1}{\widehat{\sigma}}\, \widehat{p}_0\Big(\frac{x - \widehat{\alpha}_1}{\widehat{\sigma}}\Big), \tag{4.7}$$

где $\widehat{p}_0(y) = p_0(y) \sum_{k=0}^{m} \widehat{C}_{0k} q_k(y)$, $\widehat{C}_{0k} = \sum_{j=0}^{k} a_{kj} \widehat{\mu}_j$, μ_j — нормированный центральный момент j-го порядка, $\widehat{\mu}_1 = 0$, $\widehat{\mu}_2 = 1$, $\widehat{C}_{00} = 1$, $\widehat{C}_{01} = a_{10}$, $\widehat{C}_{02} = a_{20} + a_{22}$.

Пусть, например, в качестве реперной плотности взят нормальный закон распределения, тогда

$$p_0(y) = \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}, \quad q_k(y) = \frac{1}{\sqrt{k!}}\, H_k(y),$$

где $H_k(y)$ — полином Чебышева–Эрмита, и проекционная оценка принимает вид

$$\widehat{p}(x) = \frac{1}{\sqrt{2\pi}\,\widehat{\sigma}}\, e^{-(x-\widehat{\alpha}_1)^2/(2\widehat{\sigma}^2)} \Big[1 + \sum_{k=3}^{m} \frac{1}{\sqrt{k!}} \widehat{C}_k H_k\Big(\frac{x - \widehat{\alpha}_1}{\widehat{\sigma}}\Big)\Big], \tag{4.8}$$

где $\widehat{C}_3 = \widehat{\mu}^3$, $\widehat{C}_4 = \widehat{\mu}^4 - 3$, $\widehat{C}_5 = \widehat{\mu}^5 - 10\widehat{\mu}^3$, $\widehat{C}_6 = \widehat{\mu}^6 - 15\widehat{\mu}^4 + 30$.

Напомним, что $\mu_k = \mu'_k/\sigma^k$, μ'_k — центральный момент случайной величины ξ, и если ξ нормально распределена, то $\mu_{2k} = (2k-1)!!$, $\mu_{2k+1} = 0$.

Оценку (4.8) принято называть оценкой Грама–Шарлье.

Задача 4.1. Пусть промежуток изменения случайной величины ξ — полуось $x \geqslant 0$ и в качестве реперного закона распределения взят экс-

поненциальный закон $p_0(x) = \alpha e^{-\alpha x}$, $\alpha > 0$. Показать, что $\{q_k(x)\}$ — нормированная система полиномов Чебышева–Лагерра $L_n^{\alpha}(x)$. Найти оценку Грама–Шарлье для этого случая.

Проекционные оценки можно рассмотреть для случая, когда ортонормированная система разложения отлична от системы полиномов. Действительно, пусть $\{f_k(x)\}$ — произвольная система линейно-независимых функций и задана реперная плотность $p_0(x)$. Ортонормируем систему $\{f_k(x)\}$ с весом $p_0(x)$ (например, используя стандартную схему ортогонализации Гильберта–Шмидта). Полученную ортонормированную систему функций обозначим $\{\varphi_k(x)\}$. Имеем

$$\int_{-\infty}^{\infty} \varphi_i(x)\varphi_j(x)p_0(x)\,dx = \delta_{ij}.$$

Искомую плотность $p_\xi(x)$ представим в виде

$$p_\xi(x) = p_0(x)\sum_{k=0}^{\infty} C_k\varphi_k(x), \tag{4.9}$$

где $C_k = \int_{-\infty}^{\infty} \varphi_k(x)\,p_\xi(x)\,dx$.

Введем случайные величины $Y_k = \varphi_k(\xi)$. Имеем

$$\mathbf{M}\,Y_k = \int_{-\infty}^{\infty} \varphi_k(x)p_\xi(x)\,dx = C_k.$$

Поэтому по выборке (1.1) можно дать оценки коэффициентов C_k:

$$\widehat{C}_k = \frac{1}{n}\sum_{i=1}^{n} \varphi_k(X_i).$$

Ограничиваясь в разложении (4.9) конечным числом членов и заменяя неизвестные точные значения C_k на их оценки $\widehat{C}_k$, получаем проекционную оценку

$$\widehat{p}(x) = p_0(x)\sum_{k=0}^{m} \widehat{C}_k\varphi_k(x). \tag{4.10}$$

Проекционная оценка (4.10) обладает большей гибкостью по сравнению с оценкой (4.6), поскольку дает возможность выбирать подходящую систему разложения. При удачном выборе реперной плотности и системы $\{\varphi_k(x)\}$ можно достичь хорошей точности оценки (4.10) даже при небольшом числе членов разложения.

Если в (4.10) положить $p_0(x) = \text{const}$, то получаем оценку Ченцова

$$\widehat{p}(x) = \sum_{k=1}^{m} \widehat{C}_k \varphi_k(x), \tag{4.11}$$

где $\{\varphi_k(x)\}$ — ортонормированная система, $\widehat{C}_k = \dfrac{1}{n} \sum\limits_{i=1}^{n} \varphi_k(X_i)$.

Иногда удобно брать разложения вида (4.10) или (4.11) по системе функций $\{\varphi_k(x)\}$, не обладающих свойством ортогональности. Возьмем, например, в представлении (4.11) в качестве $\varphi_k(x)$ плотности распределения. Тогда плотность (4.11) представляет собой модель смеси с неизвестными весами $C_k \in [0,1]$, $\sum\limits_{k=1}^{m} C_k = 1$. Оценки $\widehat{C}_k$ весов C_k можно получить с помощью обобщенного метода максимального правдоподобия с учетом вышеуказанных ограничений на веса.

Проекционные методы представлены в работах [16, 57, 85, 86].

4.5. Восстановление плотности распределения регуляризованным методом гистограмм

Искомая плотность связана с функцией распределения соотношением $F(x) = \int\limits_{-\infty}^{x} p_\xi(s)\, ds$, которое запишем в виде

$$\int_{-\infty}^{\infty} \eta(x-s) p_\xi(s)\, ds = F(x). \tag{4.12}$$

При заданной функции $F(x)$ (4.12) — интегральное уравнение 1-го рода относительно неизвестной плотности, т. е. задача восстановления плотности по функции распределения — некорректно поставленная задача. Но $F(x)$ также неизвестна и по выборке (1.1) можно лишь дать оценку $\widehat{F}(x)$, взяв, например, в качестве оценки эмпирическую функцию распределения $F_n(x)$.

Итак, задача восстановления плотности по выборке сводится к задаче решения интегрального уравнения (4.12) в условиях, когда точная правая часть этого уравнения неизвестна и заменена оценкой $F_n(x)$.

Рассмотрим дискретный аналог уравнения (4.12), полученный с помощью сетки $\{x_j\}$, $j = 1, \ldots, N+1$, $\Delta x_j = x_{j+1} - x_j$:

$$\int_{x_j}^{x_{j+1}} p_\xi(s)\, ds = F(x_{j+1}) - F(x_j). \tag{4.13}$$

При замене неизвестной $F(x)$ на $F_n(x)$ имеем

$$\Delta F_{nj} = F_n(x_{j+1}) - F_n(x_j) = \frac{n_j}{n},$$

где n_j — число членов выборки, попавших в интервал $[x_j, x_{j+1}]$. Элементы K_{wij} ковариационной матрицы K_w случайного вектора

$$w = (\Delta F_{n1}, \ldots, \Delta F_{nN})^T K_{wii} = \frac{\Delta F_i - (\Delta F_i)^2}{n}, \quad K_{wij} = -\frac{\Delta F_i \Delta F_j}{n},$$

$$i \neq j, \quad i, j = 1, \ldots, N; \quad \Delta F_j = F(x_{j+1}) - F(x_j).$$

Поскольку $F(x)$ неизвестна, матрицу K_w заменим ее оценкой $\widehat{K}_w$, элементы которой вычисляются по формулам

$$\widehat{K}_{wij} = -\frac{n_i n_j}{n^3}, \quad i \neq j, \qquad \widehat{K}_{wii} = \frac{1}{n}\left(\frac{n_i}{n} - \left(\frac{n_i}{n}\right)^2\right).$$

При $n \to \infty$ вектор w асимптотически нормален. Это следует из интегральной теоремы Муавра–Лапласа, так как n_j — дискретные случайные величины, распределенные по биномиальному закону. Если $n_j/n \to 0$ при $n \to \infty$, то справедлива асимптотическая формула

$$\widehat{K}_w = \operatorname{diag}\left(\frac{n_1}{n^2}, \ldots, \frac{n_N}{n^2}\right)(I + o(n)),$$

где элементы матрицы $o(n)$ стремятся к нулю при $n \to \infty$.

Пусть априорная информация об искомой плотности имеет детерминированный характер: $p_\xi(x) \in G$, где G — некоторое множество плотностей задаваемого класса (множество G может задаваться совокупностью равенств и неравенств с вхождением в них $p_\xi(x)$).

Определим ОММП-оценку плотности

$$\widehat{p}(x) = \arg \min_{p(x) \in G} \sum_{j=1}^{N} \left(\int_{x_j}^{x_{j+1}} p(x)\, dx - \frac{n_j}{n} \right)^2 \frac{1}{n_j}. \tag{4.14}$$

Если априорная информация об искомой плотности отсутствует, то ОММП-оценка (4.14) совпадает с оценкой плотности по методу гистограмм.

Если G — выпуклое множество пространства непрерывных функций, то оценка (4.14) существует и единственна и может быть найдена с помощью методов выпуклого программирования, в частности, можно использовать метод сопряженных градиентов или методы штрафных функций.

Иногда множество G априори не задается. Тогда G можно выбирать в виде

$$\{G: \ p(x) = \sum_{k=1}^{m} \alpha_k \varphi_k(x)\},$$

где $\{\varphi_k(x)\}$ — система базисных функций. В качестве базисных функций можно брать ортогональную систему (аналогично методу Ченцова). Число членов разложения m должно быть согласовано с объемом выборки и может быть найдено с помощью способа невязки [54, 71] или с помощью метода упорядоченной минимизации риска [6, 22].

На рис. 4.2 представлен пример восстановления плотности распределения с учетом априорной информации. В качестве системы

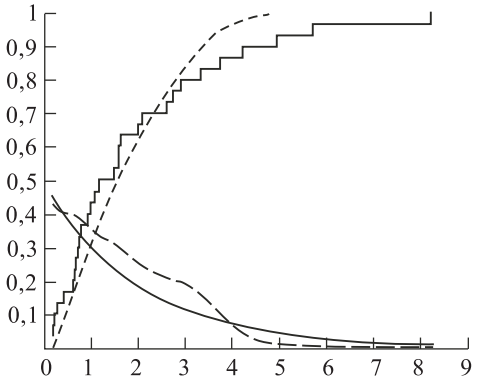

Рис. 4.2

базисных функций были использованы ортогональные полиномы Чебышева–Эрмита. Искомая плотность и эмпирическая функция распределения вероятностей представлены сплошной линией. Восстановленные плотность и функция распределения вероятностей обозначены пунктиром.

Методы непараметрической статистики представлены также в работах [66, 82].

4.6. Метод корневой оценки плотности распределения

В работе [13] был предложен новый метод оценки плотности, который связан с проекционными методами и носит название *метода корневой оценки плотности*. Суть корневого метода состоит в том, что вместо разложения искомой плотности предлагается разлагать по ортонормированной системе так называемую *пси-функцию* $\psi(x)$, связанную с искомой плотностью $p(x)$ равенством

$$p(x) = |\psi(x)|^2. \tag{4.15}$$

В квантовой механике функция $\psi(x)$ играет фундаментальную роль, называется *координатной пси-функцией* и является решением стационарного уравнения Шредингера.

Переход от разложения $p(x)$ к разложению $\psi(x)$ позволяет при применении метода максимального правдоподобия получить эффективную вычислительную схему оценки искомых параметров разложения.

Действительно, пусть

$$\psi(x) = \sum_{i=1}^{m} c_i\, \varphi_i(x), \tag{4.16}$$

где $\{\varphi_i(x)\}$ — ортонормированная система, $\{c_i\}$ — коэффициенты разложения, подлежащие оценке.

В дальнейшем предполагается, что функции $\varphi_i(x)$, $\psi(x)$ и коэффициенты c_i действительны. Из условия нормировки $\int\limits_{-\infty}^{\infty} p(x)\, dx = 1$ следует равенство

$$\sum_{i=1}^{m} c_i^2 = 1. \tag{4.17}$$

Следовательно, необходимо оценить $m-1$ независимых коэффициентов. Для их оценки используем метод максимального правдоподобия.

Если выборка (1.1) повторная, то

$$l_n(c) = \ln\, L_n(c) = \sum_{k=1}^{n} \ln \left(\sum_{i=1}^{m} \sum_{j=1}^{m} \varphi_i(X_k)\, \varphi_j(X_k)\, c_i\, c_j \right).$$

Для нахождения максимального значения логарифмической функции правдоподобия $l_n(\mathbf{c})$ с учетом ограничения (4.17) сформируем функцию Лагранжа:

$$L(c) = l_n(c) + \lambda\Big(1 - \sum_{i=1}^{m} c_i^2\Big). \tag{4.18}$$

Приравнивая к нулю частные производные $\partial L/\partial c_i$, получаем систему уравнений

$$\frac{\partial L}{\partial c_i} = 2 \sum_{k=1}^{n} \sum_{j=1}^{m} \frac{\varphi_i(X_k)\, \varphi_j(X_k)}{p(X_k)}\, c_j - 2\lambda\, c_i = 0, \quad i = 1, \ldots, m. \tag{4.19}$$

Умножая обе части последнего равенства на c_i и суммируя по i, получаем $\lambda = n$. Следовательно, для нахождения ММП-оценок коэффициентов $c_1, \ldots, c_n$ необходимо решить нелинейную систему уравнений

$$c_i = \frac{1}{n} \sum_{k=1}^{n} \varphi_i(X_k) \Big(\sum_{j=1}^{m} c_j \varphi_j(X_k) \Big)^{-1}, \qquad i = 1, \ldots, m. \tag{4.20}$$

Для решения системы (4.20) в [13] предложен итерационный процесс

$$c_i^{(l+1)} = \alpha\, c_i^{(l)} + \frac{1-\alpha}{n} \sum_{k=1}^{n} \varphi_i(X_k) \Big(\sum_{j=1}^{m} c_j^{(l)}\, \varphi_j(X_k) \Big)^{-1}, \qquad (4.21)$$

где $0 < \alpha < 1$ подбирается так, чтобы скорость сходимости итерационного процесса была наибольшей (4.21).

Поскольку из равенства (4.19) следует также равенство

$$c_i = \sum_{j=1}^{m} \Big(\frac{1}{n} \sum_{k=1}^{n} \frac{\varphi_i(X_k)\, \varphi_j(X_k)}{p(X_k)} \Big) c_j,$$

а

$$\frac{1}{n} \sum_{k=1}^{n} \frac{\varphi_i(X_k)\, \varphi_j(X_k)}{p(X_k)} \to \mathbf{M}\, \frac{\varphi_i(\xi)\, \varphi_j(\xi)}{p(\xi)} = \int_{-\infty}^{\infty} \frac{\varphi_i(x)\, \varphi_j(x)}{p(x)}\, p(x)\, dx = \delta_{ij},$$

то полученные оценки являются состоятельными.

В работе [13] приведены и другие свойства корневых оценок плотности, в частности выбор оптимального числа слагаемых m в представлении (4.16).

Как и в методе Грама–Шарлье, в качестве базисной системы функций $\varphi_i(x)$ можно взять систему функций Чебышева–Эрмита

$$\varphi_i(x) = \frac{1}{\sqrt{2^i\, i!\, \sqrt{\pi}}}\, H_i(x)\, e^{-x^2/2}, \quad i = 0,1,\ldots,$$

где $H_i(x)$ — полином Чебышева–Эрмита i-го порядка.

Глава 5
ПРОВЕРКА ГИПОТЕЗ О ЗАКОНЕ РАСПРЕДЕЛЕНИЯ

Восстановление закона распределения по выборке (1.1) в большинстве случаев основано на некоторых гипотезах. Например, в рамках параметрического подхода априори принимается гипотеза о принадлежности плотности распределения классу плотностей, определенных с точностью до конечного числа параметров. В рамках параметрического и непараметрического подходов могут использоваться дополнительные априорные гипотезы (априорная информация) о законе распределения. Но гипотезы могут быть ошибочны. Поэтому после принятия той или иной априорной гипотезы и восстановления закона распределения необходимо проверить выполнение этой гипотезы, т. е. проверить, насколько хорошо выборка (1.1) согласуется с восстановленным законом распределения. В частности, при применении параметрического подхода оценки параметров во многом зависят от принятого гипотетического класса плотностей. И поэтому проверка соответствия оцененной плотности этому классу, с точки зрения согласования выборки (1.1) с оцененным законом распределения, является гарантией правильности самих оценок параметров. Если же применяется непараметрический подход, который априори может не «навязывать» класс плотностей, то после получения оценки плотности часто возникает задача о проверке гипотез принадлежности оцененной плотности тому или иному классу плотностей.

5.1. Проверка гипотезы о законе распределения в рамках непараметрической статистики

Для проверки гипотезы о законе распределения можно использовать либо восстановленную функцию распределения (обычно эмпирическую функцию распределения $F_n(x)$), либо восстановленную плотность распределения $\widehat{p}(x)$.

При использовании эмпирической функции распределения $F_n(x)$ часто применяют критерий согласия Колмогорова и критерий ω^2.

Критерий согласия Колмогорова использует рассмотренную ранее в разделе (4.1) случайную величину

$$V_n = \max_x |F(x) - F_n(x)|, \tag{5.1}$$

где $F(x)$ — тестируемая функция распределения.

Если гипотеза H о соответствии тестируемой функции распределения истинной функции распределения верна, то $V_n \to 0$ при $n \to \infty$, в противном случае $V_n \to V^* > 0$.

На практике величину (5.1) удобно рассчитывать по формуле

$$V_n = \max_{i=1,\ldots,n} \left\{ \frac{i}{n} - F(X_{(i)}), F(X_{(i)}) - \frac{i-1}{n} \right\},$$

где $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ — ранжированная последовательность членов выборки (1.1).

Для заданных уровня значимости (обычно используют стандартные значения $p = 0{,}8;\ 0{,}9;\ 0{,}95;\ 0{,}99$) гипотеза H отвергается, если $V_n > V_{\text{крит}}(n, p)$, где $V_{\text{крит}}(n, p)$ — критическое значение величины (5.1), зависящее от объема выборки и заданного уровня значимости. Числовые значения $V_{\text{крит}}(n, p)$ можно найти в [14].

При применении критерия ω^2 используется статистика

$$\omega_n^2 = \int_{-\infty}^{\infty} (F(x) - F_n(x))^2 \, dF(x). \tag{5.2}$$

Если $F_n(x)$ — эмпирическая функция распределения, построенная по выборке (1.1), то для расчета ω_n^2 можно использовать формулу

$$\omega_n^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left(F(X_{(i)}) - \frac{2i-1}{2n} \right)^2, \tag{5.3}$$

где, как и прежде, $X_{(i)}$ — ранжированная последовательность членов выборки (1.1).

При выполнении гипотезы H о равенстве тестируемой функции распределения истинной функции распределения $\omega_n^2 \to 0$ при $n \to \infty$.

Для заданных уровня значимости и объема выборки гипотеза H отвергается, если $\omega_n^2 > \omega_{\text{крит}}(n, p)$, где $\omega_{\text{крит}}(n, p)$ — критическое значение величины (5.3), зависящее от объема выборки и уровня значимости. Числовые значения $\omega_{\text{крит}}(n, p)$ можно найти в [14].

Для проверки гипотезы H о соответствии тестируемой плотности распределения $p(x)$ выборке (1.1) часто используют критерий χ^2 (хи-квадрат) Пирсона.

Пусть область изменения значений случайной величины ξ разбита на N интервалов $[x_j, x_{j+1})$, $j = 1, \ldots, N$. Обозначим n_j — число членов выборки (1.1), попавших в интервал $[x_j, x_{j+1})$, $p_j = \int_{x_j}^{x_{j+1}} p(x)\, dx$ — вероятность попадания в интервал $[x_j, x_{j+1})$.

Определим статистику

$$\widehat{Z}_n = n \sum_{j=1}^{N} \frac{1}{p_j} \Big(\frac{n_j}{n} - p_j \Big)^2. \tag{5.4}$$

Теорема 5.1 (Пирсон). *Пусть*: 1° *выборка* (1.1) *повторная*; 2° $\sum\limits_{j=1}^{N} p_j = 1$. *Тогда статистика* $\widehat{Z}_n$ *из* (5.4) *при* $n \to \infty$ *имеет* χ^2*-распределение с* $N-1$ *степенями свободы.*

Для проверки гипотезы H о соответствии тестируемой плотности распределения $p(x)$ выборке (1.1) зададим $p_{\text{дов}}$ и по таблице χ^2-распределения найдем такое $z(p_{\text{дов}})$, что $\int\limits_0^{z(p_{\text{дов}})} P^2_{\chi^2_{N-1}}(z)\, dz = p_{\text{дов}}$. Если $\widehat{Z}_n < z(p_{\text{дов}})$, то считают, что гипотеза H подтверждается $100\, p_{\text{дов}}$-процентным уровнем доверия. Если же $\widehat{Z}_n > z(p_{\text{дов}})$, то считают, что гипотеза H о соответствии плотности распределения $p(x)$ выборке (1.1) не подтверждается.

5.2. Проверка гипотезы о законе распределения в рамках параметрической статистики

Проверка гипотез о законе распределения и согласования выборки (1.1) с оцененным законом в рамках параметрической статистики также производится с помощью различных критериев согласия. Наиболее часто используют критерий согласия χ^2 Фишера. Пусть, как и выше, область изменения значений случайной величины ξ разбита на N интервалов Δx_j, $j = 1, \ldots, N$. Обозначим n_j — число членов выборки (1.1), попавших в интервал Δx_j, $\widehat{p}_j = \int\limits_{x_j}^{x_{j+1}} p(x;\, \widehat{u})\, dx$ — вероятности попадания в интервал Δx_j, вычисленные по гипотетической оцененной плотности $p(x;\, \widehat{u})$. Определим статистику

$$\widehat{Z}_n = n \sum_{j=1}^{N} \frac{1}{\widehat{p}_j} \Big(\frac{n_j}{n} - \widehat{p}_j \Big)^2. \tag{5.5}$$

Теорема 5.2 (Фишер). *Пусть оценка* $\widehat{u}$ n*-мерного параметра и является асимптотически эффективной и асимптотически нормальной* (*например*, $\widehat{u} = \widehat{u}_{ММП}$). *Тогда статистика* $\widehat{Z}_n$ *при* $n \to \infty$ *имеет* χ^2*-распределение с* $N - m - 1$ *степенями свободы.*

Используя теорему Фишера, можно проверить соответствие гипотезы о принадлежности искомой плотности классу $p(x;\, u)$ выборки (1.1).

Действительно, зададим $p_{\text{дов}}$ и по таблице χ^2-распределения найдем такое $z(p_{\text{дов}})$, что

$$\int_0^{z(p_{\text{дов}})} p_{\chi^2_{N-m-1}}(z)\,dz = p_{\text{дов}}.$$

Если $\widehat{Z}_n < z(p_{\text{дов}})$, то считают, что гипотеза о принадлежности плотности классу $p(x;\ u)$ согласуется с выборкой (1.1). Если же $\widehat{Z}_n \geqslant z(p_{\text{дов}})$, то считают, что гипотеза о принадлежности плотности классу $p(x;\ u)$ не согласуется с выборкой (1.1) и имеет место значимое отклонение оцененной гипотетической плотности от выборки (1.1).

Критерий χ^2 применяется также для проверки гипотезы о совпадении законов распределения для серии выборок. Пусть вместо одной выборки (1.1) заданы l повторных выборок

$$X_{k1}, \ldots, X_{kn_k}, \qquad k = 1, \ldots, l, \tag{5.6}$$

где n_k — объем k-й выборки, $n = \sum\limits_{k=1}^{l} n_k$. Пусть вводится гипотеза о принадлежности плотности $p_\xi(x)$ классу $p(x;\ u)$. Обозначим $\widehat{u}$ оценку параметра u, сделанную по всей совокупности выборок (5.6):

$$\widehat{p}_j = \int_{x_j}^{x_{j+1}} p(x;\ \widehat{u})\,dx,$$

n_{kj} — число членов выборки (5.6) (с фиксированным k), попавших в интервал Δx_j. Введем статистику

$$\widehat{Z}_{ln} = \sum_{k=1}^{l} n_k \sum_{j=1}^{N} \frac{1}{\widehat{p}_j}\left(\frac{n_{kj}}{n_k} - \widehat{p}_j\right). \tag{5.7}$$

Аналогично теореме 5.2 справедливо следующее утверждение.

Теорема 5.3. *Пусть оценки $\widehat{u}$ являются асимптотически эффективными и асимптотически нормальными. Тогда при $n_k \to \infty$, $k = 1, \ldots, l$, закон распределения статистики $\widehat{z}_{ln}$ стремится к χ^2-распределению с $l(N-1) - m$ степенями свободы.*

Применим теорему 5.3 для проверки гипотезы о совпадении закона распределения для каждой из l выборок (5.6). По заданной $p_{\text{дов}}$ найдем такое $Z(p_{\text{дов}})$, что

$$\int_0^{Z(p_{\text{дов}})} p_{\chi^2_{l(N-1)-m}}(z)\,dz = p_{\text{дов}}.$$

Если $\widehat{Z}_{ln} < Z(p_{\text{дов}})$, то гипотеза о совпадении законов распределения для серии выборок (5.6) (гипотеза об однородности выборок) прини-

мается. Если $\widehat{Z}_{ln} \geqslant Z(p_{\text{дов}})$, то гипотеза об однородности выборок отвергается.

Если гипотеза об однородности выборок (5.6) проверяется по отношению к фиксированной плотности $p(x)$, то вместо статистики (5.7) вводится статистика

$$Z_{ln} = \sum_{k=1}^{l} n_k \sum_{j=1}^{N} \frac{1}{p_j} \left(\frac{n_{kj}}{n_k} - p_j \right), \tag{5.8}$$

где $p_j = \int\limits_{x_j}^{x_{j+1}} p(x)\, dx$.

Закон распределения статистики (5.8) при $n_k \to \infty$, $k = 1, \ldots, l$, стремится к χ^2-распределению с $l(N-l)$ степенями свободы. Поэтому критическая граница $z(p_{\text{дов}})$ проверки гипотезы однородности выборок (5.6) определяется из равенства

$$\int\limits_{0}^{z(p_{\text{дов}})} p_{\chi^2_{l(N-1)}}(z)\, dz = p_{\text{дов}}.$$

Проверка гипотез о законе распределения представлена в работах [28, 33, 34, 38, 40].

Глава 6
ЧИСЛЕННЫЕ МЕТОДЫ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Под статистическим моделированием в узком смысле (в самой математической статистике) обычно понимают способы моделирования случайных величин с помощью специально разработанных «датчиков» случайных величин и использования полученных с их помощью выборок для апробирования тех или иных статистических методов и численных алгоритмов, базирующихся на этих методах. Под статистическим моделированием в широком смысле понимается решение различных задач с применением методов математической статистики и моделирования случайных величин. В частности, методы Монте–Карло являются численными методами решения различных задач, и прежде всего задач математической физики, при помощи моделирования случайных величин [70].

6.1. Способы моделирования случайных величин

При решении задач методами статистического моделирования необходимо уметь моделировать реализации случайных величин с различными законами распределения. Оказывается, что достаточно уметь моделировать одну реперную случайную величину, а реализации случайных величин с другими законами распределения получать с помощью некоторых преобразований над реализациями значений реперной случайной величины. Вообще говоря, в качестве реперного можно взять различные распределения, но в настоящее время общепринято в качестве реперной брать равномерно распределенную на отрезке $[0,1]$ случайную величину γ. Действительно, пусть задана произвольная случайная величина ξ с плотностью $p(x)$ и функцией распределения

$$F(x) = \int\limits_{-\infty}^{x} p(s)\,ds.$$

Теорема 6.1. *Выполнение равенства*

$$F(\xi) = \gamma \tag{6.1}$$

необходимо и достаточно для того, чтобы $F(x)$ была функцией распределения случайной величины ξ.

Доказательство. Пусть $F(x)$ — функция распределения случайной величины ξ. Тогда случайная величина γ, определенная равен-

ством (6.1), распределена равномерно на промежутке $[0,1]$. Действительно, функция распределения $F_\gamma(y)$ случайной величины γ определится равенством

$$F_\gamma(y) = \mathbf{P}\{\gamma < y\} = \mathbf{P}\{F(\xi) < y\}.$$

Если $y < 0$, то $\mathbf{P}\{F(\xi) < y\} = 0$. Если $y > 1$, то $\mathbf{P}\{F(\xi) < y\} = 1$. Если $0 \leqslant y \leqslant 1$, то $\mathbf{P}\{F(\xi) < y\} = \mathbf{P}\{\xi < F^{-1}(y)\} = F(F^{-1}(y)) = y$.

Итак,

$$F_\gamma(y) = \begin{cases} 0, & y < 0, \\ y, & 0 \leqslant y \leqslant 1, \\ 1, & y > 1, \end{cases}$$

т. е. γ распределена равномерно на $[0,1]$.

Пусть γ распределена равномерно на $[0,1]$ и выполнено равенство (6.1), где $F(x)$ — функция распределения некоторой случайной величины. Покажем, что $F_\xi(x) = F(x)$. Действительно, $\xi = F^{-1}(\gamma)$. Поэтому $F_\xi(x) = \mathbf{P}\{\xi < x\} = \mathbf{P}\{F^{-1}(\gamma) < x\} = \mathbf{P}\{\gamma < F(x)\} = F_\gamma(F(x)) = F(x)$, поскольку $F(x) \in [0,1]$. Теорема доказана.

Из теоремы 6.1 следует, что повторная выборка

$$\gamma_1, \ldots, \gamma_n \tag{6.2}$$

равномерно распределенной на $[0,1]$ случайной величины порождает повторную выборку (1.1) случайной величины ξ, где $X_i = F_\xi^{-1}(\gamma_i)$ (мы ограничиваемся случаем, когда F_ξ^{-1} существует).

Пример 6.1. Пусть ξ равномерно распределена на отрезке $[a,b]$. Имеем

$$F_\xi(x) = \frac{x-a}{b-a}, \quad a \leqslant x \leqslant b.$$

Поэтому $X_i = a + (b-a)\gamma_i$.

Пример 6.2. Пусть ξ распределена по экспоненциальному закону. Имеем: $F_\xi(x) = 1 - e^{-\lambda x}$, $x \geqslant 0$. Поэтому $1 - e^{-\lambda X_i} = \gamma_i$, откуда

$$X_i = \frac{1}{\lambda} \ln \frac{1}{1-\gamma_i}.$$

Поскольку случайная величина $1 - \gamma_i$ также равномерно распределена на $[0,1]$, то повторную выборку (1.1) можно получить и по формуле

$$X_i = \frac{1}{\lambda} \ln \frac{1}{\gamma_i}.$$

Задача 6.1. Используя формулу (6.1), представить выборку (1.1) для нормально распределенной случайной величины ξ через выборку (6.2).

Для многих распределений использование формулы (6.1) для построения выборки (1.1) с помощью выборки (6.2) неудобно и связа-

но с достаточно большими вычислительными затратами. Ниже для некоторых стандартных широко используемых распределений будут даны другие более удобные и эффективные способы построения выборки (1.1) с помощью выборки (6.2).

Итак, для получения выборки (1.1) необходимо иметь эффективные способы получения выборки (6.2).

Существует несколько способов получения выборки (6.2) с помощью компьютера. Наиболее эффективным и надежным из них является способ псевдослучайных чисел, основанный на последовательном получении членов выборки (6.2) с использованием какой-либо формулы. В этом случае числа γ_i из (6.2) называют *псевдослучайными*.

Применяют несколько алгоритмов получения псевдослучайных чисел. Большая часть из них основана на формуле

$$\gamma_{i+1} = R(\gamma_i), \tag{6.3}$$

где $R(x)$ — достаточно быстро осциллирующая функция на промежутке $[0,1]$ со значениями на этом же промежутке.

Например, в методе вычетов Лемера $R(x) = D(g\,x)$, где $D(x)$ — дробная часть числа x, $g > 0$ — достаточно большое число, т. е. последовательность псевдослучайных чисел (6.2) с помощью метода вычетов получается по рекуррентной формуле

$$\gamma_{i+1} = D(g\gamma_i). \tag{6.4}$$

Если в (6.4) $\gamma_0 = m_0/M$, где m_0, M, g — целые взаимно простые числа, то все γ_i из (6.4) представимы в виде $\gamma_i = m_i/M$, где m_{i+1} и M взаимно просты и m_i равно остатку, полученному при делении $g\,m_i$ на M. Например, удовлетворительная последовательность псевдослучайных чисел получается при $m_0 = 1$, $M = 2^{42}$, $g = 5^{17}$. Заметим, что случайные числа γ_{i+1}, γ_i, связанные формулой (6.4), зависимы с коэффициентом корреляции, равным $1/g$. Ясно, что при достаточно большом g их можно считать практически некоррелируемыми. Более подробную информацию о псевдослучайных числах и методах их тестирования можно найти в [70].

Вернемся к методам построения выборки (1.1) по выборке (6.2). Кроме метода построения, основанного на формуле (6.1), на практике часто используют метод Неймана. Пусть значения случайной величины ξ принадлежат конечному отрезку $a \leqslant x \leqslant b$ и $p_\xi(x) \leqslant C$. Обозначим $\gamma^{(1)}$, $\gamma^{(2)}$ независимые равномерно распределенные на $[0,1]$ случайные величины. Тогда случайные величины $\xi^{(1)} = a + \gamma^{(1)}(b-a)$, $\xi^{(2)} = C\gamma^{(2)}$ независимы и равномерно распределены на отрезках $[a,b]$, $[0,C]$ соответственно, и поэтому их совместная плотность $p(x,y) = (b-a)^{-1}C^{-1}$, $a \leqslant x \leqslant b$, $0 \leqslant y \leqslant C$.

Теорема 6.2 (Нейман). *Случайная величина* $\xi = \xi^{(1)}$, *если* $\xi^{(2)} < p(\xi^{(1)})$, *имеет плотность вероятностей* $p(x)$.

Доказательство. Имеем

$$F_\xi(z) = \mathbf{P}\{\xi < z\} = \mathbf{P}\{\xi^{(1)} < z \mid \xi^{(2)} < p(\xi^{(1)})\} = \\ = \frac{\mathbf{P}\{(\xi^{(1)} < z) \cap (\xi^{(2)} < p(\xi^{(1)}))\}}{\mathbf{P}\{\xi^{(2)} < p(\xi^{(1)})\}}.$$

Числитель последнего выражения равен вероятности того, что случайная точка $(\xi^{(1)}, \xi^{(2)})$ попадает ниже кривой $y = p(x)$ и одновременно левее вертикальной прямой $x = z$. Поэтому

$$\mathbf{P}\{(\xi^{(1)} < z) \cap (\xi^{(2)} < p(\xi^{(1)}))\} = \\ = \int_a^z dx \int_0^{p(x)} \frac{1}{(b-a)C}\, dy = \frac{1}{C(b-a)} \int_a^z p(x)\, dx.$$

Знаменатель равен вероятности попадания случайной точки $(\xi^{(1)}, \xi^{(2)})$ ниже кривой $y = p(x)$ и поэтому

$$\mathbf{P}\{\xi^{(2)} < p(\xi^{(1)})\} = \int_a^b dx \int_0^{p(x)} \frac{1}{(b-a)C}\, dy = \frac{1}{C(b-a)}.$$

Итак,

$$F_\xi(z) = \left[\int_a^z \frac{p(x)}{C(b-a)}\, dx\right] \Big/ \frac{1}{C(b-a)} = \int_a^z p(x)\, dx,$$

т. е. $p(x)$ — плотность вероятностей случайной величины ξ. Теорема доказана.

Для получения одной реализации случайной величины ξ по методу Неймана требуются по крайней мере две независимые реализации случайной величины γ, причем, если случайная точка $(\xi^{(1)}, \xi^{(2)})$ попадает выше кривой $y = p(x)$, то реализации ξ не будет получено. Поэтому для снижения числа пар $(\xi^{(1)}, \xi^{(2)})$, для которых не происходит реализаций ξ, необходимо брать C равным наименьшему возможному значению $C = \max_x p(x)$.

Для получения выборки (1.1) для некоторых стандартных законов распределения используют специальные приемы, учитывающие специфику рассматриваемых распределений.

Пусть ξ — нормально распределенная случайная величина. Обозначим $\xi_0 = (\xi - \mathbf{M}\,\xi)/\sigma_\xi$ нормированную случайную величину. Пусть

$$\xi_0^{(N)} = \sum_{j=1}^N \left(\gamma_j - \frac{1}{2}\right) \Big/ \sqrt{\frac{N}{12}} = \sqrt{\frac{3}{N}} \sum_{j=1}^N (2\gamma_j - 1)$$

— нормированная сумма N независимых случайных величин, равномерно распределенных на $[0, 1]$. Согласно центральной предельной тео-

реме Линдеберга–Леви закон распределения $\xi_0^{(N)}$ при $N \to \infty$ стремится к $\mathcal{N}(0,1)$.

Поэтому при достаточно большом N закон распределения $\xi^{(N)} = \sigma_\xi \xi_0^{(N)} + \mathbf{M}\xi$ будет близок к $\mathcal{N}(\mathbf{M}\xi, \sigma_\xi^2)$. Следовательно, при достаточном большом N одну реализацию случайной величины ξ можно подсчитывать по N независимым реализациям γ согласно формуле

$$X_i = \mathbf{M}\xi + \sigma_\xi \sqrt{\frac{3}{n}} \sum_{j=1}^{N} (2\gamma_j - 1). \tag{6.5}$$

На практике ограничиваются $N = 12$ и тогда формула (6.5) принимает вид

$$X_i = \sigma_\xi \sum_{j=1}^{12} \gamma_j + \mathbf{M}\xi - 6\sigma_\xi. \tag{6.6}$$

Рассмотрим случайную величину

$$\xi_N = -\frac{1}{\lambda} \ln(\gamma_1 \dots \gamma_{N+1}),$$

где $\gamma_1, \dots, \gamma_{N+1}$ — независимые случайные величины, равномерно распределенные на $[0,1]$. Докажем, что

$$p_{\xi_N}(x) = \frac{\lambda^{N+1}}{N!} x^N e^{-\lambda x}, \quad x > 0,$$

т. е. ξ_N подчиняется γ-распределению. Действительно, для $N = 0$ утверждение доказано ранее. Предположим, что утверждение справедливо для N. Тогда

$$\xi_{N+1} = -\frac{1}{\lambda} \ln(\gamma_1 \dots \gamma_{N+2}) = \xi_N + \xi_0.$$

Поскольку ξ_N и ξ_0 независимы,

$$p_{\xi_{N+1}}(x) = \int_{-\infty}^{\infty} p_{\xi_N}(x-s) p_{\xi_0}(s)\, ds = \int_0^x \frac{\lambda^{N+1}}{N!} (x-s)^N e^{-\lambda(x-s)} \lambda e^{-\lambda s} ds =$$

$$= \frac{\lambda^{N+2}}{N!} e^{-\lambda x} \int_0^x (x-s)^N\, ds = \frac{\lambda^{N+2}}{(N+1)!} x^{N+1} e^{-\lambda x}.$$

Следовательно, если ξ подчиняется γ-распределению с плотностью вероятностей $p_\xi(x) = \frac{\lambda^{N+1}}{N!} x^N e^{-\lambda x}$, $x > 0$, то по $N+1$ независимым реализациям γ находим одну реализацию ξ согласно формуле $X_i = -\frac{1}{\lambda} \ln(\gamma_1 \dots \gamma_{N+1})$.

Рассмотрим теперь m-мерную нормально распределенную случайную величину с вектором математического ожидания $a =$

$= (a_1, \ldots, a_m)^T$ и ковариационной матрицей K_ξ. Представим $K_\xi = Q\,Q^T$, где Q — невырожденная $(m \times m)$-матрица. Матрица Q определена неоднозначно. Одну из таких матриц можно найти методом квадратного корня (тогда Q будет треугольной). Введем случайную величину $\eta = Q^{-1}(\xi - a)$. Векторная случайная величина η распределена нормально. Имеем: $\mathbf{M}\,\eta = Q^{-1}(\mathbf{M}\,\xi - a) = 0$, ковариационная матрица $K_\eta = Q^{-1}K_\xi(Q^{-1})^T = Q^{-1}\,Q\,Q^T(Q^T)^{-1} = I$ (единичная матрица). Следовательно, $\eta = (\eta_1, \ldots, \eta_m)^T$, где η_j, $j = 1, \ldots, m$ — независимые нормально распределенные величины с законом распределения $\mathcal{N}(0,1)$ и одна реализация Y_i случайной величины η состоит из m независимых реализаций скалярной случайной величины с законом $\mathcal{N}(0,1)$.

Из вышесказанного следует, что реализация X_i случайной величины ξ может быть получена по формуле

$$X_i = a + Q\,Y_i. \tag{6.7}$$

6.2. Применение метода статистического моделирования для решения некоторых прикладных задач

Ранее уже отмечалось, что метод статистического моделирования применяется как для решения «внутренних» задач самой математической статистики и обработки экспериментальных данных, так и для решения прикладных задач вне математической статистики. В частности, после создания алгоритма обработки экспериментальных данных можно «отработать» этот алгоритм еще до поступления самих экспериментальных данных путем их замены «псевдоэкспериментальными данными», полученными с помощью моделирования случайных компонент. Конечно, на этапе отработки алгоритма и формирования псевдоэкспериментальных данных необходимо учитывать всю имеющуюся у исследователя информацию о возможной области изменения будущих реальных экспериментальных данных.

При решении прикладных задач методом статистического моделирования, когда эти задачи не несут в себе элемента случайности, его вносят искусственно по следующей схеме. Обозначим через u решение задачи (u может быть скаляром, вектором, функцией, вектор-функцией и т. д. в зависимости от конкретной задачи). Подберем такую случайную величину ξ, что $\mathbf{M}\,\xi = u$. Тогда, осуществив достаточно большое число реализаций величины ξ и получив выборку ее значений $X_1, \ldots, X_n$, получают оценку $\widehat{u}$, например взяв в качестве $\widehat{u}$ среднеарифметическую $\widehat{u} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.

Основная трудность состоит в подборе такой случайной величины ξ, что $\mathbf{M}\,\xi = u$. Конечно, подбор ξ неоднозначен и, если имеется возможность выбора, то следует отбирать случайную величину ξ с наименьшим разбросом относительно среднего значения u и учитывать (часто в первую очередь) вычислительные затраты на реализацию каждого

значения X_i величины ξ. В конечном итоге выбор ξ диктуется объемом вычислительных затрат на получение оценки $\widehat{u}$ с заданной точностью.

Рассмотрим несколько приемов вычисления определенных интегралов с помощью метода статистического моделирования.

Пусть требуется вычислить интеграл

$$J = \int\limits_G f(x)\,dx, \tag{6.8}$$

где x — m-мерный вектор. Введем m-векторную случайную величину η с плотностью $p(x) > 0$, $x \in G$, и скалярную случайную величину $\xi = f(\eta)/p(\eta)$. Имеем

$$\mathbf{M}\,\xi = \int\limits_G \frac{f(x)}{p(x)}\,p(x)\,dx = J.$$

Поэтому получаем оценку интеграла

$$\widehat{J} = \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)}{p(X_i)}, \tag{6.9}$$

где X_i $i = 1, \ldots, n$ — повторная выборка значений случайной величины η. В дальнейшем предполагается, что интеграл (6.8) сходится абсолютно.

Оценка (6.9) несмещена и состоятельна (по вероятности) в силу закона больших чисел (теорема Хинчина). Оценка (6.9) определена для любой плотности $p(x) > 0$, $x \in G$, $\int\limits_G p(x)\,dx = 1$.

Точность оценки (6.9) определяется ее дисперсией. Имеем: $\mathbf{D}\,\widehat{J} = \frac{1}{n}\mathbf{D}\,\xi$. Таким образом, следует выбирать ξ с наименьшей дисперсией, или, что то же самое, с наименьшим $\mathbf{M}\,\xi^2$. Имеем

$$\mathbf{M}\,\xi^2 = \int\limits_G \frac{f^2(x)}{p^2(x)}\,p(x)\,dx = \int\limits_G \frac{f^2(x)}{p(x)}\,dx.$$

Следовательно, выбор наилучшей оценки (6.9) сводится к выбору такой $p(x) > 0$, $x \in G$, для которой интеграл $\int\limits_G \frac{f^2(x)}{p(x)}\,dx$ принимает наименьшее значение. Используя неравенство Коши–Буняковского, имеем

$$\Big(\int\limits_G |f(x)|\,dx\Big)^2 = \Big(\int\limits_G \frac{|f(x)|}{\sqrt{p(x)}}\sqrt{p(x)}\,dx\Big)^2 \leqslant$$

$$\leqslant \int\limits_G \Big(\frac{|f(x)|}{\sqrt{p(x)}}\Big)^2 dx \int\limits_G \big(\sqrt{p(x)}\big)^2 dx = \int\limits_G \frac{f^2(x)}{p(x)}\,dx.$$

Положим

$$p(x) = \frac{|f(x)|}{\int\limits_G |f(x)|\,dx}. \tag{6.10}$$

Имеем

$$\mathbf{M}\,\xi^2 = \int\limits_G \frac{f^2(x)}{p(x)}\,dx = \int\limits_G \frac{f^2(x)}{|f(x)|}\,dx \int\limits_G |f(x)|\,dx = \Big(\int\limits_G |f(x)|\,dx\Big)^2.$$

Поэтому выбор $p(x)$ согласно (6.10) обеспечивает наименьшее значение дисперсии оценки (6.9). К сожалению, определение плотности согласно (6.10) требует знания нормировочной константы $\int\limits_G |f(x)|\,dx$, которая так же, как исходный интеграл (6.8), неизвестна. Несмотря на эту принципиальную трудность определения оптимальной плотности, формула (6.10) подсказывает, какую $p(x)$ следует выбирать, а именно: $p(x)$ должна «повторять» глобальное изменение $|f(x)|$ в области G. Пусть функция $f_a(x)$ достаточно хорошо аппроксимирует глобальное изменение $f(x)$ и интеграл $\int\limits_G |f_a(x)|\,dx$ достаточно легко подсчитывается. Тогда плотность

$$p(x) = |f_a(x)| \cdot \Big(\int\limits_G |f_a(x)|\,dx\Big)^{-1}$$

обеспечивает близость оценки (6.9) к оптимальной оценке с $p(x)$ из (6.10).

Существуют другие приемы уменьшения дисперсии оценки (6.9). Все эти приемы основаны на использовании дополнительной информации об искомом интеграле или подынтегральной функции $f(x)$. Пусть, например, функция $f_0(x)$ близка к $f(x)$ в L_2-метрике, т. е.

$$\int\limits_G \frac{(f(x) - f_0(x))^2}{p(x)}\,dx \leqslant \varepsilon,$$

где ε мало, и интеграл $\int\limits_G f_0(x)\,dx = J_0$ известен. Введем скалярную случайную величину

$$\xi = J_0 + \frac{f(\eta) - f_0(\eta)}{p(\eta)}$$

и оценку

$$\widehat{J} = J_0 + \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i) - f_0(X_i)}{p(X_i)}. \tag{6.11}$$

Имеем

$$\mathbf{M}\,\xi = J,$$

$$\mathbf{D}\,\xi = \mathbf{D}\left(\frac{f(\eta) - f_0(\eta)}{p(\eta)}\right) = \int\limits_G \frac{(f(x) - f_0(x))^2}{p(x)}\,dx - (J - J_0)^2 \leqslant \varepsilon^2.$$

Поэтому для дисперсии оценки справедливо неравенство $\mathbf{D}\,\widehat{J} \leqslant \varepsilon^2/n$. Такой способ учета дополнительной информации с получением оценки (6.11) называется *методом выделения главной части.*

Подчеркнем, что применение методов Монте–Карло дает значительное преимущество по сравнению с классическими детерминированными методами вычисления интегралов при увеличении размерности m области G (при увеличении кратности интеграла (6.8)).

Рассмотрим теперь решение линейных интегральных уравнений вида

$$\varphi(x) = \int\limits_G K(x, x')\varphi(x')\,dx' + f(x) \tag{6.12}$$

методами Монте–Карло.

Введем линейный интегральный оператор

$$K\varphi(x) = \int\limits_G K(x, x')\varphi(x')\,dx'.$$

Обозначим $K^j\varphi$ результат j-кратного применения оператора K к функции $\varphi(x)$ и

$$(f, \varphi) = \int\limits_G f(x)\varphi(x)\,dx$$

— скалярное произведение функций $f(x)$ и $\varphi(x)$. Таким образом, $(f, K^j\varphi)$ — скалярное произведение функции $f(x)$ и функции $K^j\varphi$, полученной в результате j-кратного применения оператора K к функции $\varphi(x)$. Например,

$$K^2\varphi(x) = \int\limits_G K(x, x') \int\limits_G K(x', x'')\varphi(x'')\,dx''\,dx' \equiv \int\limits_G K_2(x, x')\varphi(x')\,dx',$$

где

$$K_2(x, x') = \int\limits_G K(x, x'')K(x'', x')\,dx''$$

— дважды итерированное ядро. Для произвольного j имеем

$$K^j\varphi(x) = \int\limits_G K_j(x, x')\varphi(x')\,dx',$$

где

$$K_j(x,x') = \int\limits_G K_{j-1}(x,x'')K(x'',x')\,dx''$$

— итерированные ядра.

Выберем в G такие плотность $p(x) > 0$ и условную плотность $p(x \mid x') > 0$, что $\int\limits_G p(x)\,dx = \int\limits_G p(x \mid x')\,dx = 1$.

Условную плотность $p(x \mid x')$ называют *плотностью вероятностей перехода из точки x' в точку x* и часто обозначают $p(x' \to x)$.

Определим в G случайную траекторию

$$T_k = \{x_0 \to x_1 \to \ldots \to x_k\},$$

где x_0 — реализация случайной величины ξ_0 с плотностью $p(x)$, x_j — реализация случайной величины ξ_j с плотностью $p(x_j \mid x_{j-1})$, $j = 1,\ldots,k$.

Введем совокупность скалярных статистик

$$w_j = \frac{K(x_0,x_1)}{p(x_1 \mid x_0)}\,\frac{K(x_1,x_2)}{p(x_2 \mid x_1)} \cdots \frac{K(x_{j-1},x_j)}{p(x_j \mid x_{j-1})},$$

определенных для $j = 1,\ldots,k$. Если положить $w_0 = 1$, то $w_j = w_{j-1} \cdot K(x_{j-1},x_j)/p(x_j \mid x_{j-1})$, $j = 1,\ldots,k$.

Определим совокупность скалярных случайных величин $\vartheta_j = w_j \dfrac{\varphi(x_0)}{p(x_0)}\,f(x_j)$. Покажем, что $\mathbf{M}\,\vartheta_j = (\varphi, K^j f)$. Действительно,

$$\mathbf{M}\,\vartheta_j = \mathbf{M}\,\frac{\varphi(x_0)}{p(x_0)}\,\frac{K(x_0,x_1)}{p(x_1 \mid x_0)} \cdots \frac{K(x_{j-1},x_j)}{p(x_j \mid x_{j-1})}\,f(x_j) =$$

$$= \int\limits_G \varphi(s_0)\,ds \int\limits_G \ldots \int\limits_G K(s_0,s_1)\ldots K(s_{j-1},s_j)f(s_j)\,ds_1\ldots ds_j =$$

$$= \int\limits_G \varphi(s_0)K^j f(s_0)\,ds = (\varphi, K^j f).$$

Таким образом, реализовав n случайных траекторий

$$T_k = (T_{k1},\ldots,T_{kn}),$$

получаем оценки $\widehat{u}_j$ скалярных произведений $u_j = (\varphi, K^j f)$ согласно формулам

$$\widehat{u}_j = \frac{1}{n}\sum_{i=1}^{n} \vartheta_{ji}, \tag{6.13}$$

где ϑ_{ji}, $i = 1,\ldots,n$ — реализации случайной величины ϑ_j на i-й траектории T_{ji}.

С помощью оценок (6.13) скалярных произведений $(\varphi, K^j f)$ можно получить оценки значений функции $K^j f(x_0)$. Действительно, пусть

$\varphi(x) = \delta(x - x_0)$. Тогда $(\delta(x - x_0), K^j f(x)) = K^j f(x_0)$. Возьмем в качестве $p(x) = \delta(x - x_0)$, т. е. все траектории T_{ji} «стартуют» из фиксированной точки x_0 и $\vartheta_{j0} = w_j f(x_j)$. Имеем: $\mathbf{M}\vartheta_{j0} = K^j f(x_0)$.

Теперь все готово для построения приближенного решения интегрального уравнения (6.12). Действительно, если спектральный радиус линейного интегрального оператора K из (6.12) меньше 1, то единственное решение $\varphi(x)$ уравнения (6.12) представимо в виде ряда Неймана:

$$\varphi(x) = f(x) + Kf(x) + K^2 f(x) + \ldots \tag{6.14}$$

Ограничимся в (6.14) конечным числом слагаемых. Тогда в фиксированной точке $x = x_0$

$$\varphi(x_0) \approx f(x_0) + Kf(x_0) + \ldots + K^k f(x_0) \tag{6.15}$$

и приближенные значения слагаемых $K^j f(x_0)$, $j = 1, \ldots, k$, правой части (6.15) могут быть оценены по формулам (6.13), где ϑ_{ji}, $i = 1, \ldots, n$, — реализации случайной величины ϑ_j на i-й траектории T_{ji} с фиксированной точкой «старта» $x = x_0$ для всех траекторий.

Если уравнение (6.12) описывает стационарное распределение $\varphi(x)$ частиц, мигрирующих в объеме G, а $K(x, x')$ — индикатриса рассеяния частиц и $f(x)$ — функция распределения независимых источников частиц, то $K^j f(x_0)$ определяет «долю» в стационарном распределении j раз рассеянных частиц. Координаты траектории T_{ki} в этом случае являются координатами точек изменения направления движения частиц за счет последовательных актов их рассеяния.

Глава 7
МЕТОД НАИМЕНЬШИХ КВАДРАТОВ ДЛЯ ЛИНЕЙНЫХ МОДЕЛЕЙ С НЕОПРЕДЕЛЕННЫМИ ДАННЫМИ

Метод наименьших квадратов (МНК) — один из наиболее широко используемых методов при решении многих задач восстановления регрессионных зависимостей. Впервые МНК был использован Лежандром в 1806 г. для решения задач небесной механики на основе экспериментальных данных астрономических наблюдений. В 1809 г. Гаусс изложил статистическую интерпретацию МНК и тем самым дал начало широкому применению статистических методов при решении задач восстановления регрессионных зависимостей.

В дальнейшем (кроме гл. 10) в монографии будут рассматриваться задачи линейного регрессионного анализа.

Пусть $u = (u_1, \ldots, u_m)^T$ — m-мерный вектор искомых параметров, $f = (f_1, \ldots, f_n)^T$ — n–мерный вектор, полученный из эксперимента и тем самым отличающийся от неизвестного точного вектора f_T на вектор случайных погрешностей $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, т. е. $f = f_T + \varepsilon$. Известно, что искомый вектор u связан с вектором f_T соотношением

$$Au = f_T, \tag{7.1}$$

где A — заданная $(n \times m)$-матрица. Заметим, что в § 9.5 будет рассмотрен случай, когда элементы матрицы A также имеют случайный характер.

Систему (7.1) запишем в виде

$$Au = f - \varepsilon. \tag{7.2}$$

Задача нахождения приближенных значений u, «удовлетворяющих» системе (7.2), и будет основной задачей линейного регрессионного анализа.

Подчеркнем, что значения вектора случайных погрешностей ε неизвестны, известен лишь закон распределения случайного вектора ε или даже не сам закон, а только некоторые его характеристики.

7.1. Примеры линейных моделей с неопределенными данными

Прежде чем переходить к изложению МНК в рамках классической схемы, приведем примеры прикладных задач, сводящихся к задачам линейного регрессионного анализа.

Пример 7.1. Пусть между двумя физическими величинами x, y существует функциональная зависимость вида $y = \sum_{j=1}^{m} u_j \varphi_j(x)$, где $\varphi_1(x), \ldots, \varphi_m(x)$ — известные функции, $u_1, \ldots, u_m$ — параметры, значения которых неизвестны. Предположим, что производится серия фиксаций x_i, $i = 1, \ldots, n$, переменной x и при $x = x_i$ проводятся измерения y_i, $i = 1, \ldots, n$, переменной y, причем измерения y_i содержат случайные погрешности ε_i. Таким образом, имеем систему равенств

$$y_i - \varepsilon_i = \sum_{j=1}^{m} u_j \varphi_j(x_i), \qquad i = 1, \ldots, n. \tag{7.3}$$

Введем векторы

$$f = (y_1, \ldots, y_n)^T, \quad \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T, \quad u = (u_1, \ldots, u_m)^T$$

и $(n \times m)$-матрицу A, $A_{ij} = \varphi_j(x_i)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Тогда система равенств (7.3) запишется в эквивалентном виде (7.2).

Пример 7.2. Пусть между величинами y и $x_1, \ldots, x_m$ существует функциональная зависимость вида $y = \sum_{j=1}^{m} u_j x_j$, где $(u_1, \ldots, u_m)^T$ — параметры, значения которых неизвестны. Для восстановления значений параметров u_j, $j = 1, \ldots, m$, фиксируются значения x_{ji}, $i = 1, \ldots, n$, переменных x_j, $j = 1, \ldots, m$, и при $x_j = x_{ji}$, $j = 1, \ldots, m$, производятся измерения y_i, $i = 1, \ldots, n$, переменной y, причем измерения y_i содержат случайные погрешности ε_i.

Таким образом, имеем систему равенств

$$y_i - \varepsilon_i = \sum_{j=1}^{m} u_j x_{ji}, \qquad i = 1, \ldots, n. \tag{7.4}$$

Если обозначить $f = (y_1, \ldots, y_n)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $u = (u_1, \ldots, u_m)^T$, A — $(n \times m)$-матрица с элементами $A_{ij} = x_{ji}$, то система (7.4) запишется в эквивалентном виде (7.2).

Пример 7.3. Рассмотрим интегральное уравнение первого рода

$$\int_a^b K(x, s) u(s)\, ds = f_T(x), \tag{7.5}$$

где ядро $K(x, s)$ задано, $u(x)$ — неизвестная искомая функция, а вместо $f_T(x)$ известна ее реализация $f(x)$, измеряемая с погрешностью, т. е. $f(x) = f_T(x) + \varepsilon(x)$, $\varepsilon(x)$ — случайная функция погрешностей.

К интегральному уравнению первого рода (7.5) сводятся многие задачи идентификации изучаемого объекта по измерениям косвенной

информации об этом объекте. В приложениях $K(x,s)$ часто играет роль так называемой *аппаратной функции.*

Введем сетку $a \leqslant x_1 < x_2 < \ldots < x_n \leqslant b$ и будем рассматривать равенство (7.5) только в узлах $x = x_i$, $i = 1, \ldots, m$. Проведем дискретизацию интегрального уравнения (7.5), заменяя интеграл в левой части (7.5) интегральной суммой

$$\sum_{j=1}^{m} K(x_i, s_j)\alpha_j u(s_j) = f_T(x_i),$$

где α_j зависят от выбранной интегральной формулы. Например, применяя формулу трапеций, имеем $\alpha_j = \Delta s_j = s_{j+1} - s_j$.

Поэтому дискретный аналог интегрального уравнения (7.5) имеет вид

$$\sum_{j=1}^{m} K(x_i, s_j)\alpha_j u(s_j) = f(x_i) - \varepsilon(x_i), \qquad i = 1, \ldots, n. \tag{7.6}$$

Обозначим

$$f = (f(x_1), \ldots, f(x_n))^T, \quad \varepsilon = (\varepsilon(x_1), \ldots, \varepsilon(x_n))^T,$$

$$u = (u(s_1), \ldots, u(s_m))^T,$$

A — $(n \times m)$-матрица с элементами $A_{ij} = K(x_i, s_j)\alpha_j$.

Тогда система (7.6) запишется в эквивалентном виде (7.2).

Пример 7.4. В приложениях часто приходится рассматривать системы интегральных равенств вида

$$\int_a^b K_i(s)u(s)\,ds = a_{iT}, \qquad i = 1, \ldots, n, \tag{7.7}$$

где $K_i(s)$ — известные функции, $u(s)$ — искомая функция, a_{iT} — числа.

К системе интегральных равенств вида (7.7) сводятся, например, многие задачи спектрометрии и вычислительной томографии.

Из эксперимента известны не точные значения a_{iT}, а их приближенные значения a_i, т. е. $a_i = a_{iT} + \varepsilon_i$, где ε_i — случайные погрешности.

Проведем дискретизацию системы интегральных равенств (7.7). Вместо системы (7.7) получим ее дискретный аналог

$$\sum_{j=1}^{m} K_i(s_j)\alpha_j u(s_j) = a_i - \varepsilon_i, \qquad i = 1, \ldots, n, \tag{7.8}$$

где известные коэффициенты α_j зависят от выбранной интегральной формулы.

Если обозначить $f = (a_1, \ldots, a_n)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, A — $(n \times m)$-матрица с элементами $A_{ij} = K_i(s_j)\alpha_j$, то систему (7.8) можно записать в эквивалентном виде (7.2).

Замечание 7.1. При дискретизации интегрального уравнения (7.5) и системы интегральных равенств (7.7) всегда необходимо соблюдать выполнение неравенства $n \geqslant m$.

7.2. Классическая схема МНК

Рассмотрим основное уравнение (7.2) линейного регрессионного анализа. В рамках классической схемы МНК вводятся следующие предположения.

1. Отсутствует какая-либо априорная информация о векторе искомых параметров u, т. е. априори предполагается, что $u \in R^m$;

2. Векторы ε и f случайные;

3. Отсутствуют систематические смещения в f, т. е. $\mathbf{M}\,\varepsilon = 0$ — нулевой вектор;

4. Ковариационная матрица вектора ε дается равенством $K_\varepsilon = \sigma^2 I_n$, где I_n — единичная матрица порядка n, т. е. компоненты ε_i попарно не коррелируют и имеют одинаковую дисперсию σ^2 (σ^2 может быть неизвестной);

5. Матрица A — детерминированная и полного ранга, т. е. $\operatorname{rank} A = m$.

Введем функционал

$$Q(u) = \|\varepsilon\|^2 \equiv \|f - Au\|^2, \tag{7.9}$$

где $\|\varepsilon\| = \left(\sum\limits_{i=1}^{n} \varepsilon_i\right)^{1/2}$ — евклидова норма.

Согласно МНК за оценку искомого вектора u принимается вектор, который минимизирует сумму квадратов отклонений:

$$\widehat{u}_{\text{МНК}} = \arg\min_u Q(u). \tag{7.10}$$

Найдем явный вид МНК-оценки. Имеем

$$Q(u) = (f - Au, f - Au) = \|f\|^2 - 2(f, Au) + (Au, Au) =$$

$$= \|f\|^2 - 2(u, A^T f) + (A^T Au, u),$$

где A^T — транспонированная к A матрица.

Поскольку на u априори не накладывается никаких ограничений, МНК-оценка является корнем уравнения $\operatorname{grad}_u Q(u) = 0$. Имеем

$$\operatorname{grad}_u Q(u) = -2A^T f + 2A^T Au.$$

Поэтому получаем равенство $A^T A\widehat{u}_{\text{МНК}} = A^T f$. Так как A — матрица полного ранга, то симметричная $(m \times m)$-матрица $A^T A$ — положительно определена. Следовательно,

$$\widehat{u}_{\text{МНК}} = (A^T A)^{-1} A^T f. \tag{7.11}$$

Очевидно, что при $u = \widehat{u}_{\text{МНК}}$ функционал $Q(u)$ имеет единственный (глобальный) минимум, поскольку $\dfrac{\partial^2 Q(u)}{\partial u^2} = 2A^TA$ — положительно определенная матрица.

Как и любая оценка, МНК-оценка $\widehat{u}_{\text{МНК}}$ — (векторная) случайная величина. Из (7.11) следует, что $\widehat{u}_{\text{МНК}}$ — линейная оценка относительно вектора экспериментальных измерений f. Исследуем некоторые свойства МНК-оценки (7.11). Имеем: $\mathbf{M}\,\widehat{u}_{\text{МНК}} = \mathbf{M}\,(A^TA)^{-1}A^Tf = (A^TA)^{-1}A^T\mathbf{M}\,f = (A^TA)^{-1}A^TAu = u$, т. е. МНК-оценка не смещена.

Найдем ковариационную матрицу K_u МНК-оценки $\widehat{u}_{\text{МНК}}$. Имеем: $K_u = BK_fB^T$, где $B = (A^TA)^{-1}A^T$, K_f — ковариационная матрица вектора измерений f. Но $K_f = K_\varepsilon = \sigma^2 I_n$, $B^T = A(A^TA)^{-1}$. Поэтому $K_u = (A^TA)^{-1}A^T\sigma^2 I_n A(A^TA)^{-1}$. Итак,

$$K_u = \sigma^2(A^TA)^{-1}. \tag{7.12}$$

Из (7.12), в частности, следует, что K_u стремится к нулевой матрице при $\sigma^2 \to 0$, т. е. МНК-оценка (7.11) сходится в среднем квадратичном к истинному значению вектора u искомых параметров при неограниченном увеличении точности измерения каждой компоненты вектора f.

Заметим, что сама МНК-оценка (7.11) не зависит от дисперсии σ^2 и тем самым определяется однозначно и при неизвестной σ^2. В то же время ковариационная матрица МНК-оценки (7.11), определяющая точность этой оценки, пропорциональна σ^2.

Итак, МНК-оценка (7.11) линейна и несмещена. Но существует бесконечно много линейных несмещенных оценок $\widehat{u}$ искомого вектора u (все они имеют вид $\widehat{u} = Rf$, где R — детерминированная $(n \times m)$-матрица). Спрашивается, какое место среди всех линейных несмещенных оценок занимает МНК-оценка (7.11)?

Теорема 7.1 (Гаусса–Маркова). *Пусть выполнены предположения* 1–5 *классической схемы* МНК. *Тогда* МНК-*оценка* (7.11) *эффективна в классе всех линейных несмещенных оценок.* [1])

Доказательство. Рассмотрим любую линейную несмещенную оценку $\widehat{u} = Rf$. Имеем

$$\mathbf{M}\,\widehat{u} = u = R\mathbf{M}\,f = RAu.$$

Поэтому $RA = I_m$. Обозначим $L = R - (A^TA)^{-1}A^T$. Имеем

$$LA = RA - (A^TA)^{-1}A^TA = I_m - I_m = 0$$

[1]) Напомним, что оценка $\widehat{u}^*$ называется *эффективной в классе оценок* $\{\widehat{u}\}$, если $K_{\widehat{u}^*} \leqslant K_{\widehat{u}}$ для любой оценки $\widehat{u} \in \{\widehat{u}\}$.

— нулевая матрица. Далее, для ковариационной матрицы $K_{\widehat{u}}$ оценки $\widehat{u}$ получаем

$$K_{\widehat{u}} = RK_f R^T = R\sigma^2 I_n R^T = \sigma^2 RR^T = \sigma^2[L + (A^TA)^{-1}A^T] \times$$
$$\times [L + (A^TA)^{-1}A^T]^T = \sigma^2[L + (A^TA)^{-1}A^T]\,[L^T + A(A^TA)^{-1}] =$$
$$= \sigma^2[LL^T + LA(A^TA)^{-1} + (L\,A(A^TA)^{-1})^T + (A^TA)^{-1}] =$$
$$= \sigma^2[(A^TA)^{-1} + LL^T].$$

Поскольку симметричная $(m \times m)$-матрица $L\,L^T$ неотрицательна, то $(A^TA)^{-1} + LL^T \geqslant (A^TA)^{-1}$, откуда $K_{\widehat{u}} \geqslant K_u$. Теорема доказана.

Из теоремы Гаусса–Маркова, в частности, следует, что дисперсия любой компоненты МНК-оценки (7.11) не превосходит дисперсию этой компоненты любой другой линейной несмещенной оценки, т. е. МНК-оценка совместно эффективна в классе линейных несмещенных оценок. Итак, МНК-оценка (7.11) является эффективной в классе линейных несмещенных оценок и по совокупности параметров $u_1, \ldots, u_m$, и по каждому параметру u_j, $j = 1, \ldots, m$, в отдельности.

Иногда необходимо получить оценку не самого вектора u, а вектора $v = Cu$, где C — заданная детерминированная $(k \times m)$-матрица. Справедливо следующее обобщение теоремы Гаусса–Маркова: при выполнении условий классической схемы МНК-оценка $\widehat{v}_{\text{МНК}} = C\widehat{u}_{\text{МНК}}$ эффективна в классе всех линейных несмещенных оценок.

Подчеркнем, что в этом разделе всюду идет речь о классе линейных несмещенных оценок, использующих лишь информацию об измерениях компонент $f_1, \ldots, f_n$ вектора f из (7.2).

Часто дисперсия σ^2 погрешностей измерений f_i неизвестна. Оказывается, что в этом случае можно получить оценки не только искомого вектора u, но и дополнительного неизвестного параметра σ^2. Обозначим $\widehat{f}_{\text{МНК}} = A\widehat{u}_{\text{МНК}}$, $e = f - \widehat{f}_{\text{МНК}}$. Компоненты $e_1, \ldots, e_n$ вектора e называют *невязками*, а сам вектор e — *вектором невязок*.

Введем положительную статистику

$$s^2 = \frac{\|e\|^2}{n - m}. \tag{7.13}$$

Теорема 7.2. *Статистика s^2 — несмещенная оценка дисперсии σ^2.*

Доказательство. Имеем

$$e = Au + \varepsilon - A\widehat{u}_{\text{МНК}} = Au + \varepsilon - A(A^TA)^{-1}A^Tf =$$
$$= Au + \varepsilon - A(A^TA)^{-1}A^T(Au + \varepsilon) =$$
$$= Au + \varepsilon - Au - A(A^TA)^{-1}A^T\varepsilon = (I_n - A(A^TA)^{-1}A^T)\varepsilon = C\varepsilon,$$

где $C = I_n - A(A^TA)^{-1}A^T$. Поскольку

$$C^2 = [I_n - A(A^TA)^{-1}A^T]^2 = I_n - 2A(A^TA)^{-1}A^T +$$
$$+ A(A^TA)^{-1}A^TA(A^TA)^{-1}A^T = I_n - A(A^TA)^{-1}A^T = C,$$

то C — проекционная матрица. Кроме того, C — симметричная матрица. Далее,

$$\mathbf{M}\,\|e\|^2 = \mathbf{M}\,e^Te = \mathbf{M}\,\varepsilon^TC^TC\varepsilon = \mathbf{M}\,\varepsilon^TC\varepsilon = \mathbf{M}\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}\varepsilon_i\varepsilon_j =$$

$$= \sigma^2\sum_{i=1}^{n}C_{ii} = \sigma^2\,\mathrm{tr}\;C = \sigma^2(n - \mathrm{tr}(A(A^TA)^{-1}A^T)) = \sigma^2(n-m),$$

так как $\mathrm{tr}\,(A(A^TA)^{-1}A^T) = \mathrm{tr}\,((A^TA)^{-1}A^TA) = \mathrm{tr}\;I_m = m$.

Итак, $\mathbf{M}\,s^2 = \sigma^2$. Теорема доказана.

Если σ^2 неизвестна, то в качестве несмещенной оценки ковариационной матрицы МНК-оценки (7.11) можно взять матрицу $K_{\widehat{u}} = s^2(A^TA)^{-1}$.

Заметим, что корреляционная матрица $\mathrm{cor}(\widehat{u})$ МНК-оценки (7.11) не зависит от σ^2 и полностью определяется матрицей A.

Рассмотрим вопрос о состоятельности МНК-оценки (7.11) при $n \to \infty$, т. е. при неограниченном увеличении числа уравнений системы (7.2). В задачах регрессии увеличению n соответствует увеличение числа измерений функции y. Ясно, что не всякое неограниченное увеличение числа уравнений системы (7.2) будет приводить к неограниченному увеличению точности МНК-оценки (7.11). Например, если в задаче восстановления функциональной зависимости $y = u_1 + u_2x$ неограниченное увеличение n происходит только за счет неограниченного увеличения числа измерений y_i при одном фиксированном значении $x_i = x_0$, то МНК-оценка (7.11) не будет состоятельной.

Поскольку при увеличении n изменяется матрица A (добавляются новые строки), то вместо одной системы (7.2) рассматривается последовательность систем вида

$$A_nu = f_n - \varepsilon_n, \tag{7.14}$$

где A_n — последовательность матриц с неограниченно увеличивающимся числом строк n.

Для того чтобы последовательность МНК-оценок

$$\widehat{u}_n = (A_n^TA)^{-1}A_n^Tf_n \tag{7.15}$$

обладала свойством состоятельности, необходимо наложить определенные ограничения на последовательность матриц A_n исходной системы (7.14).

В дальнейшем предполагается, что для модели (7.14) выполнены все пять предположений классической схемы МНК.

Напомним, что состоятельность в среднем квадратичном последовательности несмещенных оценок $\widehat{u}_n$ эквивалентна стремлению к нулевой матрице последовательности ковариационных матриц K_n оценок $\widehat{u}_n$. Поэтому состоятельность в среднем квадратичном последовательности МНК-оценок (7.15) при $n \to \infty$ эквивалентна соотношению $(A_n^T A_n)^{-1} \to 0$ — нулевая матрица при $n \to \infty$.

Теорема 7.3 (Эйкер). *Для того чтобы последовательность* МНК-*оценок* (7.15) *была состоятельной в среднем квадратичном, необходимо и достаточно выполнение соотношения* $\lambda_{\min}(A_n^T A_n) \to +\infty$ *при* $n \to \infty$. *Здесь* $\lambda_{\min}(A_n^T A_n)$ *— минимальное собственное значение симметричной положительно определенной матрицы* $A_n^T A_n$.

Доказательство. Пусть $\widehat{u}_n$ сходится в среднем квадратичном к u, т.е. $(A_n^T A_n)^{-1} \to 0$ при $n \to \infty$. Из последнего соотношения следует, что все собственные значения последовательности матриц $(A_n^T A_n)^{-1}$ стремятся к нулю, в частности $\lambda_{\max}(A_n^T A_n)^{-1} \to 0$. Но тогда $\lambda_{\min}(A_n^T A_n) = 1/\lambda_{\max}(A_n^T A_n)^{-1} \to \infty$. Пусть теперь $\lambda_{\min}(A_n^T A_n) \to$ $\to \infty$. Тогда $\lambda_{\max}(A_n^T A_n)^{-1} = 1/\lambda_{\min}(A_n^T A_n) \to 0$ при $n \to \infty$. Следовательно, все собственные значения последовательности матриц $(A_n^T A_n)$ стремятся к нулю, что эквивалентно стремлению к нулевой матрице последовательности матриц $(A_n^T A_n)^{-1}$. Теорема доказана.

Условие Эйкера,

$$\lambda_{\min}(A_n^T A_n) \to +\infty \quad \text{при} \quad n \to \infty, \tag{7.16}$$

являющееся необходимым и достаточным для состоятельности в среднем квадратичном, в общем случае довольно трудно проверить. Поэтому было предложено несколько более легко проверяемых условий, достаточных для состоятельности в среднем квадратичном последовательности МНК-оценок (7.15).

Теорема 7.4. *Пусть существуют такие последовательность положительных чисел* $\alpha_n \to 0$ *и положительно определенная симметричная* $(m \times m)$*-матрица* R *что* $\lim_{n\to\infty} \alpha_n A_n^T A_n = R$. *Тогда последовательность* МНК-*оценок* (7.15) *состоятельна в среднем квадратичном.*

Доказательство. Имеем: $K_n = (A_n^T A_n)^{-1} = \alpha_n (\alpha_n A_n^T A_n)^{-1}$. Но $(\alpha_n A_n^T A_n)^{-1} \to R^{-1}$ при $n \to \infty$. Поэтому $K_n \to 0$. Теорема доказана.

Пример 7.5. Рассмотрим задачу восстановления функциональной зависимости $y = u\,x$, где u — неизвестный параметр, x — независимая переменная. Пусть измерения производятся в точках $x_i = i\,\Delta x$, $i = 1, \ldots, n$, где Δx — расстояние между двумя соседними точками фиксации x_i. Имеем: $A_n = \Delta x(1, 2, \ldots, n)^T$ — матрица-столбец, $A_n^T = \Delta x(1, 2, \ldots, n)$ — матрица-строка, $A_n^T A_n = (\Delta x)^2 \sum_{i=1}^{n} i^2 =$ $= (\Delta x)^2 n(n+1)(2n+1)/6$ — положительное число. Имеем: $A_n^T A_n / n^3 \to (\Delta x)^2/3$ при $n \to \infty$. Поэтому условие теоремы 7.4

выполнено, и, следовательно, последовательность МНК-оценок $\widehat{u}_n = 6(n(n+1)(2n+1)\Delta x)^{-1}\sum_{i=1}^{n} iy_i$ сходится в среднем квадратичном к точному значению параметра u. Конечно, в данном примере состоятельность в среднем квадратичном легко проверяется непосредственно, так как

$$K_n = \sigma^2(\widehat{u}_n) = \frac{6\sigma^2}{n(n+1)(2n+1)(\Delta x)^2} \to 0 \quad \text{при} \quad n \to \infty.$$

Приведем еще один достаточный критерий состоятельности в среднем квадратичном последовательности МНК-оценок (7.15).

Определим матрицы сопряженности R_n равенством

$$(R_n)_{ij} = \frac{(A_n^T A_n)_{ij}}{\sqrt{(A_n^T A_n)_{ii}(A_n^T A_n)_{jj}}}, \quad i,j = 1,\ldots,m.$$

Матрица сопряженности R_n симметрична и положительно определена, причем на ее главной диагонали стоят единицы, а элемент $(R_n)_{ij}$ — косинус угла между двумя векторами — i-м и j-м столбцами матрицы A_n.

Теорема 7.5. *Пусть*

1° $(A_n^T A_n)_{jj} = \sum_{i=1}^{n} a_{ij}^2 \to \infty$ *при* $n \to \infty$ *для любого* $j = 1,\ldots,m$;

2° $\lim_{n\to\infty} R_n = R$, *где* R *— симметричная положительно определенная* $(m \times m)$*-матрица.*

Тогда последовательность МНК*-оценок* (7.15) *состоятельна в среднем квадратичном.*

Доказательство. Введем диагональные $(m \times m)$-матрицы

$$D_n = \operatorname{diag}\left(\sqrt{(A_n^T A_n)_{11}}, \ldots, \sqrt{(A_n^T A_n)_{mm}}\right).$$

Имеем: $R_n = D_n^{-1} A_n^T A_n D_n^{-1}$, откуда $A_n^T A_n = D_n R_n D_n$. Поскольку $|R| \neq 0$, то $\lambda_{\min}(R) > \delta > 0$. Поэтому существует такое n_0, что для всех $n \geqslant n_0$ выполнены неравенства $\lambda_{\min}(R_n) \geqslant \delta$. Имеем: $\lambda_{\min}(A_n^T A_n) = \lambda_{\min}(D_n R_n D_n) \geqslant \lambda_{\min}(R_n)\lambda_{\min}(D_n^2) \geqslant \delta \min (A_n^T A_n)_{jj} \to \infty$ при $n \to \infty$. Итак, выполнено условие Эйкера (7.16). Теорема доказана.

Пример 7.6. Рассмотрим задачу восстановления функциональной зависимости $y = u_1 + u_2 x$, где u_1, u_2 — неизвестные искомые параметры, x — независимая переменная. Пусть измерения производятся в точках $x_i = i\Delta x$, $i = 1,\ldots,n$, где Δx — расстояние между двумя соседними точками фиксации x_i. Имеем

$$A_n = \begin{pmatrix} 1 & \Delta x \\ 1 & 2\Delta x \\ \ldots & \ldots \\ 1 & n\Delta x \end{pmatrix}, \quad A_n^T A_n = \begin{pmatrix} n & \dfrac{n(n+1)}{2}\Delta x \\ \dfrac{n(n+1)}{2}\Delta x & \dfrac{n(n+1)(2n+1)}{6}(\Delta x)^2 \end{pmatrix}.$$

Здесь A_n — $(n \times 2)$-матрица, $A_n^T A_n$ — симметричная положительно определенная (2×2)-матрица. Матрицы сопряженности даются равенствами

$$R_n = \begin{pmatrix} 1 & \frac{\sqrt{6}}{2}\sqrt{\frac{n+1}{2n+1}} \\ \frac{\sqrt{6}}{2}\sqrt{\frac{n+1}{2n+1}} & 1 \end{pmatrix} \to \begin{pmatrix} 1 & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & 1 \end{pmatrix} \quad \text{при} \quad n \to \infty.$$

Следовательно, условия теоремы 7.5 выполнены.

Выше была доказана несмещенность оценки (7.13) неизвестной дисперсии σ^2. Рассмотрим теперь состоятельность оценки (7.13) при $n \to \infty$.

Теорема 7.6. *Если выполнены предположения классической схемы* МНК, *то оценка* (7.13) *состоятельна по вероятности при* $n \to \infty$ *без дополнительных условий.*

Доказательство. При доказательстве теоремы 7.2 было установлено равенство $e = C\varepsilon$, где $C = I_n - A_n(A_n^T A_n)^{-1} A_n^T$ — проекционная симметричная матрица. Поэтому

$$s^2 = \frac{1}{n-m}\, e^T e = \frac{1}{n-m}\varepsilon^T C^T C\varepsilon = \frac{1}{n-m}\varepsilon^T C\varepsilon =$$

$$= \frac{1}{n-m}\sum_{i=1}^{n}\varepsilon_i^2 - \frac{1}{n-m}\varepsilon^T A_n(A_n^T A_n)^{-1} A_n^T\varepsilon.$$

В силу закона больших чисел $\dfrac{1}{n-m}\sum\limits_{i=1}^{n}\varepsilon_i^2 \to \sigma^2$ по вероятности при $n \to \infty$. Покажем, что $\dfrac{1}{n-m}\varepsilon^T A_n(A_n^T A_n)^{-1} A_n^T\varepsilon \to 0$ в среднем квадратичном при $n \to \infty$. Действительно, поскольку матрица $(A_n^T A_n)^{-1}$ симметрична и положительно определена, то существует такая невырожденная $(m \times m)$-матрица P_n, что $(A_n^T A_n)^{-1} = P_n P_n^T$. Следовательно, $A_n^T A_n = (P_n^T)^{-1} P_n^{-1}$, откуда получаем равенство $P_n^T A_n^T A_n P_n = I_m$, где I_m — единичная матрица порядка m. Обозначим $r_n = (n-m)^{-1/2} P_n^T A_n^T\varepsilon$. Имеем

$$\|r_n\|^2 = r_n^T r_n = \frac{1}{n-m}\varepsilon^T A_n P_n P_n^T A_n^T\varepsilon = \frac{1}{n-m}\varepsilon^T A_n(A_n^T A_n)^{-1} A_n^T\varepsilon.$$

Следовательно, достаточно доказать соотношение $r_n \to 0$ в среднем квадратичном при $n \to \infty$. Имеем: $\mathbf{M}\, r_n = 0$ — нулевой вектор. Для ковариационной матрицы K_r вектора r_n получаем соотношения

$$K_r = \mathbf{M}\,(r_n r_n^T) = \frac{1}{n-m}\mathbf{M}\,(P_n^T A_n^T\varepsilon\varepsilon^T A_n P_n) = \frac{\sigma^2}{n-m}\, P_n^T A_n^T A_n P_n =$$

$$= \frac{\sigma^2}{n-m}\, I_m.$$

Следовательно, K_r стремится к нулевой матрице при $n \to \infty$, т. е. $r_n \to 0$ в среднем квадратичном при $n \to \infty$. Теорема доказана.

Последовательность МНК-оценок (7.15) при выполнении некоторых условий обладает еще одним замечательным свойством асимптотической нормальности.

Прежде чем переходить к выявлению этого свойства у последовательности МНК-оценок (7.15), напомним определение асимптотической нормальности для произвольной последовательности случайных векторов v_n размерности m.

Обозначим $\mathbf{M}\, v_n = a_n$, V_n — ковариационная матрица вектора v_n. Предполагается, что матрицы V_n положительно определенные. Тогда существуют такие положительно определенные симметричные матрицы $V_n^{-1/2}$, что $V_n^{-1/2} V_n^{-1/2} = V_n^{-1}$. Введем последовательность нормированных векторов $v_n^0 = V_n^{-1/2}(v_n - a_n)$. Имеем: $\mathbf{M}\, v_n^0 = 0$ — нулевой вектор.

Для ковариационной матрицы V_n^0 вектора v_n^0 получаем равенства $V_n^0 = V_n^{-1/2} V_n V_n^{-1/2} = V_n^{-1/2} V_n^{1/2} V_n^{1/2} V_n^{-1/2} = I_m$. Итак, ковариационная матрица нормированного вектора является единичной матрицей.

Говорят, что последовательность случайных векторов v_n *асимптотически нормальна*, если закон распределения нормированных векторов v_n^0 стремится к нормальному $\mathcal{N}(0, I_m)$ при $n \to \infty$.

Проведем нормировку последовательности МНК-оценок (7.15). Имеем

$$\widehat{u}_n^0 = \frac{1}{\sigma}\,(A_n^T A_n)^{1/2}(\widehat{u}_n - u) = \frac{1}{\sigma}\,(A_n^T A_n)^{1/2}((A_n^T A_n)^{-1} A_n^T f - u) =$$

$$= \frac{1}{\sigma}\,(A_n^T A_n)^{1/2}((A_n^T A_n)^{-1} A_n^T (A_n u + \varepsilon) - u) =$$

$$= \frac{1}{\sigma}\,(A_n^T A_n)^{-1/2} A_n^T \varepsilon = B_n \varepsilon,$$

где $B_n = \dfrac{1}{\sigma}\,(A_n^T A_n)^{-1/2} A_n^T$ — $(m \times n)$-матрица. Обозначим b_{nji}, $j = 1, \ldots, m$, $i = 1, \ldots, n$, — элементы матрицы B_n.

Теорема 7.7. *Пусть дополнительно к предположениям классической схемы* МНК *случайные величины* ε_i, $i = 1, \ldots, n$ *независимы и их моменты третьего порядка равномерно ограничены. Тогда для асимптотической нормальности последовательности* МНК*-оценок* (7.15) *достаточно выполнения соотношения*

$$b_{nji} \to 0 \tag{7.17}$$

при $n \to \infty$ *для всех* $j = 1, \ldots m$, $i = 1, \ldots, n$.

Справедливость теоремы 7.7 следует из центральной предельной теоремы Ляпунова.

Теорема 7.8. *Пусть выполнены условия теорем* 7.4, 7.7 *и* $\alpha_n a_{ij}^2 \to 0$ *при* $n \to \infty$ *для всех* $i = 1, \ldots, n$, $j = 1, \ldots, m$. *Тогда*

последовательность МНК-*оценок* (7.15) *асимптотически нормальна и справедливо соотношение* $\frac{1}{\sqrt{\alpha_n}}\,(\widehat{u}_n - u) \sim \mathcal{N}(0, \sigma^2 R^{-1})$ *при* $n \to \infty$.

Доказательство. Имеем

$$B_n = \frac{1}{\sigma}\,(A_n^T A_n)^{-1/2} A_n^T = \frac{1}{\sigma}\,(\alpha_n A_n^T A_n)^{-1/2} \sqrt{\alpha_n}\, A_n^T.$$

Но $(\alpha_n A_n^T A_n)^{-1/2} \to R^{-1/2}$, $\sqrt{\alpha_n}\, A_n^T \to 0$ при $n \to \infty$. Поэтому выполнено соотношение (7.17) теоремы 7.7. Далее, ковариационная матрица вектора $\frac{1}{\sqrt{\alpha_n}}\,(\widehat{u}_n - u)$ дается равенством

$$K_{\widehat{u}} = \frac{1}{\alpha_n}\,\sigma^2 (A_n^T A_n)^{-1} = \sigma^2 (\alpha_n A_n^T A_n)^{-1}.$$

Поэтому $K_{\widehat{u}} \to \sigma^2 R^{-1}$ при $n \to \infty$. Теорема доказана.

Заметим, что соотношение (7.17) означает неограниченное уменьшение доли каждого отдельного измерения при неограниченном увеличении числа измерений.

Наличие свойства асимптотической нормальности последовательности МНК-оценок (7.15) позволяет утверждать, что при достаточно большом n закон распределения МНК-оценки (7.11) близок к нормальному $\mathcal{N}(u, \sigma^2 (A^T A)^{-1})$. Последнее позволяет, например, строить приближенные доверительные интервалы для каждой компоненты вектора u с тем большей точностью, чем больше n. Действительно, при достаточно большом n можно утверждать, что с вероятностью $p_{\text{дов}}$ истинное значение u_j принадлежит интервалу

$$\widehat{u}_{j\text{МНК}} - \sigma t(p_{\text{дов}}) \sqrt{(A^T A)^{-1}_{jj}} \leqslant u_j \leqslant \widehat{u}_{j\text{МНК}} + \sigma t(p_{\text{дов}}) \sqrt{(A^T A)^{-1}_{jj}}\,,$$

где

$$\frac{2}{\sqrt{2\pi}} \int\limits_0^{t(p_{\text{дов}})} e^{-s^2/2}\, ds = p_{\text{дов}}, \quad j = 1, \ldots, m.$$

До сих пор предполагалось, что закон случайных величин ε_i, $i = 1, \ldots, n$, произволен. Предположим теперь, что случайные величины ε_i распределены по нормальному закону. Тогда с учетом предположений классической схемы МНК совместная плотность вероятностей $p_\varepsilon(x)$ совокупности случайных величин ε определяется равенством

$$p_\varepsilon(x) = \frac{1}{(2\pi)^{n/2}\sigma^n}\, \exp\left\{-\frac{x^T x}{2\sigma^2}\right\}.$$

Равенство (7.2) дает реализацию случайной величины $\varepsilon = f - Au$, и, следовательно, функция правдоподобия имеет вид

$$L(u) = \frac{1}{(2\pi)^{n/2}\sigma^n}\, \exp\left\{-\frac{(f - Au)^T (f - Au)}{2\sigma^2}\right\}.$$

ММП-оценка искомого вектора u определяется равенством $\widehat{u}_{\text{ММП}} = \arg\max_u L(u)$, которое в нашем случае эквивалентно равенству $\widehat{u}_{\text{ММП}} = \arg\min_u \|Au - f\|^2$.

Итак, если в рамках классической схемы МНК погрешности распределены по нормальному закону, то МНК-оценка (7.11) совпадает с ММП-оценкой. Поскольку случайная реализация f подчиняется нормальному закону $\mathcal{N}(f_T, \sigma^2 I_n)$, а МНК-оценка — линейная комбинация компонент вектора f, то МНК-оценка (7.11) также подчиняется нормальному закону $\mathcal{N}(u, \sigma^2(A^TA)^{-1})$.

Замечательным дополнительным свойством МНК-оценки в случае нормального закона распределения погрешностей является ее эффективность в классе всех несмещенных оценок.

Теорема 7.9. *Пусть* $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. *Тогда* МНК-*оценка* (7.11) *является эффективной в классе всех несмещенных* (*линейных и нелинейных*) *оценок.*

Доказательство. Для любой несмещенной оценки $\widehat{u}_n$ справедливо неравенство Фишера–Крамера–Рао

$$K_u \geqslant I^{-1},$$

где

$$I = \mathbf{M}((\operatorname{grad}_u \ln L(u))(\operatorname{grad}_u \ln L(u))^T)$$

— информационная матрица Фишера. Имеем

$$\ln L(u) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\|Au - f\|^2.$$

Поэтому

$$\operatorname{grad}_u \ln L(u) = \frac{1}{\sigma^2}A^T(f - Au) = \frac{1}{\sigma^2}A^T\varepsilon.$$

Отсюда

$$(\operatorname{grad}_u \ln L(u))^T = \frac{1}{\sigma^2}\varepsilon^T A,$$

следовательно,

$$(\operatorname{grad}_u \ln L(u))(\operatorname{grad}_u \ln L(u))^T = \frac{1}{\sigma^4}A^T\varepsilon\varepsilon^T A.$$

Поэтому

$$I = \mathbf{M}\left(\frac{1}{\sigma^4}A^T\varepsilon\varepsilon^T A\right) = \frac{1}{\sigma^4}A^T\mathbf{M}(\varepsilon\varepsilon^T)A = \frac{1}{\sigma^2}A^TA,$$

откуда

$$I^{-1} = \sigma^2(A^TA)^{-1},$$

т. е. в неравенстве Фишера–Крамера–Рао для МНК-оценки (7.11) достигается равенство. Теорема доказана.

Часто регрессионную модель (7.1) используют для подсчета значения линейной комбинации $y = \sum_{j=1}^{m} a_j u_j = a^T u$, где a — заданный вектор. Эту задачу называют *прогнозом* для y. Например, при восстановлении функциональной зависимости необходимо знать значения функции не только в точках фиксации независимой переменной $x_1, \ldots, x_n$, где производятся экспериментальные измерения функции, но и при других значениях независимой переменной. Прогноз по формуле

$$\widehat{y}_{\text{МНК}} = a^T \widehat{u}_{\text{МНК}} \tag{7.18}$$

будем называть МНК-*прогнозом*.

Теорема 7.10. МНК-*прогноз* (7.18) *несмещен и эффективен в классе всех линейных несмещенных прогнозов.*

Доказательство. Имеем

$$\mathbf{M}\,\widehat{y}_{\text{МНК}} = \mathbf{M}\,a^T \widehat{u}_{\text{МНК}} = a^T \mathbf{M}\,\widehat{u}_{\text{МНК}} = a^T u = y,$$

т. е. МНК-прогноз несмещен. Любой линейный несмещенный прогноз $\widehat{y}$ имеет форму $\widehat{y} = a^T \widehat{u}$, где $\widehat{u}$ — линейная несмещенная оценка u, т. е. $\widehat{u} = Rf$ и $RA = I_m$. Обозначим $L = R - (A^T A)^{-1} A^T$. При доказательстве теоремы 7.1 было показано, что $LA = 0$. Имеем

$$\sigma^2(\widehat{y}_{\text{МНК}}) = \sigma^2 a^T (A^T A)^{-1} a,$$

$$\sigma^2(\widehat{y}) = a^T R K_f R^T a = \sigma^2 a^T R R^T a = \sigma^2 a^T (L + A^T A)^{-1} A^T) \times$$

$$\times\, (L^T + A(A^T A)^{-1}) a = \sigma^2 a^T ((A^T A)^{-1} + L L^T) a =$$

$$= \sigma^2 a^T (A^T A)^{-1} a + \sigma^2 a^T L L^T a \geqslant \sigma^2(\widehat{y}_{\text{МНК}}).$$

Последнее неравенство следует из неотрицательности симметричной матрицы LL^T. Теорема доказана.

Если в предположениях классической схемы вектор погрешностей ε распределен нормально, то МНК-прогноз (7.18) распределен по одномерному нормальному закону $\mathcal{N}(y, \sigma^2 a^T (A^T A)^{-1} a)$ и является эффективным в классе всех несмещенных прогнозов (линейных и нелинейных).

7.3. Обобщения классической схемы МНК

В настоящем параграфе будут рассмотрены обобщения классической схемы МНК, связанные с отказом от выполнения тех или иных предположений классической схемы.

Начнем обобщения с отказа от предположения 4 классической схемы, вместо которого вводится следующее предположение: ковариационная матрица вектора ε дается равенством $K_\varepsilon = \sigma^2 \Omega$, где Ω — положительно определенная $(n \times n)$-матрица (σ^2 может быть неиз-

вестной). Таким образом, теперь возможно наличие корреляционных связей между компонентами случайного вектора ε.

Поскольку матрица Ω симметрична и положительно определена, существует такая невырожденная матрица S, что $\Omega = SS^T$. Введем обозначения $\varphi = S^{-1}f$, $\xi = S^{-1}\varepsilon$, $B = S^{-1}A$. Тогда регрессионная модель (7.2) эквивалентна системе

$$Bu = \varphi - \xi. \tag{7.19}$$

Имеем

$$\mathbf{M}\,\xi = \mathbf{M}\,S^{-1}\varepsilon = 0,$$

$$K_\xi = S^{-1}K_\varepsilon(S^{-1})^T = \sigma^2 S^{-1}\Omega(S^{-1})^T = \sigma^2 S^{-1}SS^T(S^T)^{-1} = \sigma^2 I_n.$$

Таким образом, регрессионная модель (7.19) удовлетворяет всем предположениям классической схемы МНК. Следовательно, МНК-оценка искомого вектора u определяется равенством

$$\widehat{u}_{\text{МНК}} = \arg\min Q(u), \tag{7.20}$$

где $Q(u) = \|Bu - \varphi\|^2$.

Преобразуем функционал $Q(u)$. Имеем

$$Q(u) = (Bu-\varphi)^T(Bu-\varphi) = (S^{-1}Au - S^{-1}f)^T(S^{-1}Au - S^{-1}f) =$$
$$= (Au-f)^T(S^{-1})^TS^{-1}(Au-f) = (Au-f)^T(SS^T)^{-1}(Au-f) =$$
$$= (Au-f)^T\Omega^{-1}(Au-f).$$

Итак, МНК-оценка (7.20) определяется равенством

$$\widehat{u}_{\text{МНК}} = \arg\min\,(Au-f)^T\Omega^{-1}(Au-f). \tag{7.21}$$

Имеем

$$\widehat{u}_{\text{МНК}} = (B^TB)^{-1}B^T\varphi = ((S^{-1}A)^TS^{-1}A)^{-1}(S^{-1}A)^TS^{-1}f =$$
$$= (A^T(S^{-1})^TS^{-1}A)^{-1}A^T(S^{-1})^TS^{-1}f = (A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1}f.$$

Следовательно

$$\widehat{u}_{\text{МНК}} = (A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1}f. \tag{7.22}$$

Оценку (7.22) принято называть *взвешенной* МНК-*оценкой Гаусса–Маркова.*

Взвешенная МНК-оценка (7.22) несмещена. Действительно,

$$\mathbf{M}\,\widehat{u}_{\text{МНК}} = (A^T\Omega^{-1}A^T)^{-1}A^T\Omega^{-1}f = (A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1}Au = u.$$

Найдем ковариационную матрицу оценки (7.22). Имеем

$$K_u = (A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1}K_f((A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1})^T =$$
$$= \sigma^2(A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1}\Omega\Omega^{-1}A(A^T\Omega^{-1}A)^{-1} = \sigma^2(A^T\Omega^{-1}A)^{-1}.$$

Здесь использована симметричность матриц $(A^T\Omega^{-1}A)^{-1}$ и Ω^{-1}.

Итак,

$$K_u = \sigma^2(A^T\Omega^{-1}A)^{-1}. \tag{7.23}$$

Теорема 7.11. *При сделанных выше предположениях обобщенная* МНК*-оценка эффективна в классе линейных несмещенных оценок.*

Схема доказательства теоремы 7.3 аналогична схеме доказательства теоремы 7.2.

Рассмотрим последовательность взвешенных МНК-оценок

$$\widehat{u}_n = (A_n^T\Omega_n^{-1}A_n)^{-1}A_n^T\Omega_n^{-1}f_n$$

для регрессионной модели $A_n u = f_n - \varepsilon_n$, $K_\varepsilon = \sigma^2\Omega_n$.

Теорема 7.12. *Пусть выполнено условие Эйкера* (7.16) *и* $\lambda_{\max}(\Omega_n) \leqslant r$, *где* r *не зависит от* n. *Тогда последовательность взвешенных* МНК*-оценок* $\widehat{u}_n$ *состоятельна в среднем квадратичном.*

Доказательство. Достаточно доказать, что $\lambda_{\min}(A_n^T\Omega_n^{-1}A_n) \to \infty$ при $n \to \infty$. Имеем

$$\lambda_{\min}(A_n^T\Omega^{-1}A_n) \geqslant \lambda_{\min}(A_n^TA_n)\lambda_{\min}(\Omega_n^{-1}) = \frac{\lambda_{\min}(A_n^TA_n)}{\lambda_{\max}(\Omega_n)} \geqslant$$

$$\geqslant \frac{1}{r}\lambda_{\min}(A_n^TA_n) \to \infty$$

при $n \to \infty$. Теорема доказана.

Пусть теперь дополнительно к вышеуказанным предположениям вектор погрешностей ε распределен по нормальному закону $\mathcal{N}(0, \sigma^2\Omega)$. Тогда взвешенная МНК-оценка (7.22) будет ММП-оценкой. Действительно, функция правдоподобия для этого случая имеет вид

$$L(u) = \frac{1}{(2\pi)^{n/2}|\sigma^2\Omega|^{1/2}} \exp\left\{-\frac{(f-Au)^T\Omega^{-1}(f-Au)}{(2\sigma^2)}\right\}.$$

Следовательно,

$$\widehat{u}_{\text{ММП}} = \arg\max L(u) = \arg\min (f-Au)^T\Omega^{-1}(f-Au) = \widehat{u}_{\text{МНК}}.$$

Более того, сама взвешенная МНК-оценка будет распределена по нормальному закону $\mathcal{N}(u, \sigma^2(A^T\Omega^{-1}A)^{-1})$.

Часто σ^2 неизвестна. Из формулы (7.22) следует, что взвешенная МНК-оценка не зависит от σ^2, но ковариационная матрица взвешенной МНК-оценки, определяющая точность этой оценки, пропорциональна σ^2 (см. формулу (7.23)). Если σ^2 неизвестна, то ее можно оценить по формуле

$$\widehat{\sigma}^2 = \frac{(A\widehat{u}_{\text{МНК}} - f)^T\Omega^{-1}(A\widehat{u}_{\text{МНК}} - f)}{n-m}. \tag{7.24}$$

Оценка (7.24) несмещена и состоятельна по вероятности при $n \to$ $\to \infty$.

При решении практических задач иногда ковариационная матрица Ω погрешностей неизвестна и тогда вместо оптимальной взвешенной МНК-оценки (7.22) применяют МНК-оценку (7.11). Ясно, что в этом случае эффективность оценивания снижается. Спрашивается, сохраняются ли тем не менее такие свойства оценки, как несмещенность и состоятельность?

Имеем: $\mathbf{M}\,\widehat{u}_{\text{МНК}} = \mathbf{M}\,(A^TA)^{-1}A^Tf = (A^TA)^{-1}A^T\mathbf{M}\,f = u$, т. е. оценка остается несмещенной без всяких дополнительных условий.

Предположим, что выполнены условия теоремы 7.12. Имеем

$$K_{\widehat{u}} = (A_n^TA_n)^{-1}A_n^TK_f[(A_n^TA_n)^{-1}A_n^T]^T = \\ = \sigma^2(A_n^TA_n)^{-1}A_n^T\Omega_nA_n(A_n^TA_n)^{-1}.$$

Поэтому

$$\lambda_{\max}(K_{\widehat{u}}) \leqslant \sigma^2\lambda_{\max}(\Omega_n)\lambda_{\max}((A_n^TA_n)^{-1}A_n^TA_n(A_n^TA_n)^{-1}) = \\ = \sigma^2\lambda_{\max}((A_n^TA_n)^{-1})\lambda_{\max}(\Omega_n) \leqslant \frac{\sigma^2r}{\lambda_{\min}(A_n^TA_n)} \to 0$$

при $n \to \infty$. Следовательно, при вышеуказанных условиях МНК-оценка (7.11) состоятельна в среднем квадратичном даже при наличии корреляций между компонентами вектора погрешностей.

При решении многих задач кроме регрессионной модели (7.2) исследователь располагает дополнительной информацией о векторе u искомых параметров. В гл. 9 будут подробно рассмотрены методы учета дополнительной априорной информации, и особенно в условиях, когда в исходной регрессионной модели (7.2) «заложено» недостаточно информации о параметрах. Здесь же предположим, что дополнительная информация задана в виде системы равенств

$$Bu = b, \tag{7.25}$$

где B — заданная $(k \times m)$-матрица ранга k, b — заданный вектор размерности k, $k < m$.

Требуется на основе регрессионной модели (7.2) получить оценку $\widehat{u}$, удовлетворяющую системе уравнений (7.25).

Пусть выполняются предположения классической схемы МНК.

Определим обобщенную МНК-оценку $\widehat{u}_{\text{ОБ}}$ соотношением

$$\widehat{u}_{\text{ОБ}} = \arg\min \|Au - f\|^2 \tag{7.26}$$

при условии (7.25).

Найдем оценку $\widehat{u}_{\text{ОБ}}$ и выясним ее свойства.

Теорема 7.13. *Оценка $\widehat{u}_{ОБ}$ (7.26) единственна, несмещена и дается равенством*

$$\widehat{u}_{ОБ} = \widehat{u}_{МНК} + (A^TA)^{-1}B^TC(b - B\widehat{u}_{МНК}), \tag{7.27}$$

где $\widehat{u}_{МНК}$ — МНК-*оценка* (7.11), $C = (B(A^TA)^{-1}B^T)^{-1}$. *Для ковариационной матрицы оценки* $\widehat{u}_{ОБ}$ *справедливо равенство*

$$K_{\widehat{u}} = \sigma^2 (A^TA)^{-1} (I_m - B^TCB(A^TA)^{-1}). \tag{7.28}$$

Доказательство. Для решения экстремальной задачи (7.26) с ограничениями (7.25) применим метод множителей Лагранжа. Введем функционал Лагранжа $Q(u) = \|Au - f\|^2 + \alpha^T(Bu - b)$, где α — вектор множителей Лагранжа размерности k.

Имеем: $Q(u) = (Au - f, Au - f) + \alpha^T(Bu - b)$,

$$\operatorname{grad}_u Q(u) = 2A^TAu - 2A^Tf + B^T\alpha = 0, \tag{7.29}$$

откуда $u = (A^TA)^{-1}(A^Tf - \frac{1}{2}B^T\alpha)$. Подставляя последнее выражение в (7.25), найдем вектор $\alpha = 2(B(A^TA)^{-1}B^T)^{-1}(B\widehat{u}_{\text{МНК}} - b)$. Если теперь в уравнении (7.29) заменить α последним выражением, то получаем оценку (7.27).

Далее,

$$\mathbf{M}\,\widehat{u}_{\text{ОБ}} = \mathbf{M}\,\widehat{u}_{\text{МНК}} + \mathbf{M}\,(A^TA)^{-1}B^TC(b - B\widehat{u}_{\text{МНК}}) = u + \\ + (A^TA)^{-1}B^TC(b - Bu) = u,$$

$$K_{\widehat{u}} = LK_{u_{\text{МНК}}}L^T,$$

где $L = I_m - (A^TA)^{-1}B^TCB$. Следовательно,

$$K_{\widehat{u}} = \sigma^2(I_m - (A^TA)^{-1}B^TCB)(A^TA)^{-1}(I_m - B^TCB(A^TA)^{-1}) = \\ = \sigma^2((A^TA)^{-1} - (A^TA)^{-1}B^TCB(A^TA)^{-1})(I_m - B^TCB(A^TA)^{-1}) = \\ = \sigma^2(A^TA)^{-1}(I_m - B^TCB(A^TA)^{-1})^2 = \\ = \sigma^2(A^TA)^{-1}(I_m - B^TCB(A^TA)^{-1}).$$

Теорема доказана.

Из доказательства теоремы 7.13 следует, что

$$P = I_m - B^TCB(A^TA)^{-1}$$

— проекционная матрица, и тем самым все ее собственные значения равны либо 1, либо 0. Можно показать, что P — вырожденная матрица, т. е. по крайней мере одно ее собственное значение равно нулю.

Из формулы (7.28) следует, что

$$K_{\widehat{u}} = \sigma^2(A^TA)^{-1} - \sigma^2(A^TA)^{-1}B^TCB(A^TA)^{-1} \leqslant \\ \leqslant \sigma^2(A^TA)^{-1} = K_{u_{\text{МНК}}},$$

поскольку матрица $(A^TA)^{-1}B^TCB(A^TA)^{-1}$ симметрична и неотрицательна.

Итак, обобщенная МНК-оценка (7.27) эффективней МНК-оценки (7.11). Последнее связано с тем, что оценка (7.27) использует дополнительную информацию об искомых параметрах, даваемую системой (7.25).

Можно показать, что обобщенная МНК-оценка (7.27) эффективна в классе всех линейных несмещенных оценок, удовлетворяющих системе (7.25).

Если ранг матрицы B равен m $(k = m)$, то оценка (7.27) совпадает с единственным решением системы (7.25), а $K_{\widehat{u}} = 0$ — нулевая матрица. Действительно, имеем

$$C = (B(A^TA)^{-1}B^T)^{-1} = (B^T)^{-1}(A^TA)B^{-1},$$

$$\widehat{u}_{\text{ОБ}} = \widehat{u}_{\text{МНК}} + (A^TA)^{-1}B^T(B^T)^{-1}(A^TA)B^{-1}(b - B\widehat{u}_{\text{МНК}}) =$$

$$= \widehat{u}_{\text{МНК}} + B^{-1}b - \widehat{u}_{\text{МНК}} = B^{-1}b,$$

$$K_{\widehat{u}} = \sigma^2(A^TA)^{-1}(I_m - B^T(B^T)^{-1}(A^TA)B^{-1}B(A^TA)^{-1}) =$$

$$= \sigma^2(A^TA)^{-1}(I_m - I_m) = 0.$$

Итак, если $k = m$, то система (7.25) содержит полную информацию об искомых параметрах, однозначно и точно (без погрешностей) их определяя, а регрессионная модель (7.2) становится ненужной.

В заключение этого параграфа сделаем несколько замечаний относительно численной реализации МНК.

Для нахождения МНК-оценок и ковариационных матриц, определяющих точность этих оценок, необходимо обратить матрицы:

— для МНК-оценки (7.11) — A^TA;

— для взвешенной МНК-оценки (7.22) — Ω, $A^T\Omega^{-1}A^T$;

— для обобщенной МНК-оценки (7.27) — A^TA, $B(A^TA)^{-1}B^T$.

Матрицы A^TA, $A^T\Omega^{-1}A^T$ — размерности $(m \times m)$, Ω — размерности $(n \times n)$, $B(A^TA)^{-1}B^T$ — размерности $(k \times k)$.

Поскольку все матрицы A^TA, Ω, $A^T\Omega^{-1}A$, $B(A^TA)^{-1}B^T$ симметричны и положительно определены, то для их обращения рекомендуется применять эффективный метод квадратного корня, созданный как раз для обращения симметричных матриц. В математическом обеспечении современных компьютеров имеются стандартные программы обращения симметричных матриц методом квадратного корня.

Большие вычислительные трудности возникают, когда матрица A^TA плохо обусловлена (в этом случае и матрица $A^T\Omega^{-1}A$ будет также плохо обусловлена). Плохая обусловленность матрицы A^TA означает, что регрессионная модель (7.1) содержит недостаточно информации об искомых параметрах u. В этих случаях необходимо использовать дополнительную априорную информацию о векторе u и применять

методы оценивания, позволяющие «соединять» всю информацию о векторе u. Подробно такие методы будут рассмотрены в гл. 9.

Метод наименьших квадратов и его применения представлены в работах [2, 6, 15, 27, 64, 65, 68].

7.4. Прогнозирование с помощью линейных регрессионных моделей

При решении прежде всего экономических задач, в частности задач прогнозирования в финансовой области, можно использовать многофакторную регрессионную модель вида

$$Y = u_0 + \sum_{j=1}^{m} u_j X_j + \varepsilon, \tag{7.30}$$

где $X_1, \ldots, X_m$ — наблюдаемые случайные величины (факторы); ε — помеха, носящая хаотический характер, причем $\mathbf{M}\,\varepsilon = 0$; $u_0, u_1, \ldots, u_m$ — параметры, подлежащие оценке; Y — искомая величина (отклик системы).

Модель (7.30) может быть использована для прогнозирования искомой величины Y для будущих моментов времени, если значения факторов для настоящего момента t_0 известны.

В этом случае необходимо реализовать два этапа. На первом из них на основе «исторических» данных оценивают значения параметров $u_0, u_1, \ldots, u_m$ (этап «обучения» модели (7.30)). На втором этапе используют равенство (7.30) для прогноза искомой величины Y при оцененных значениях $u_0, u_1, \ldots, u_m$ и известных «текущих» значениях факторов $X_1, \ldots, X_m$.

Замечание 7.2. Если исследуемый с помощью модели (7.30) процесс имеет нестационарный характер, то приходится периодически «переобучать» модель (7.30), переоценивая параметры $u_0, u_1, \ldots, u_m$ на основе последних новых «исторических» данных.

Для получения оптимальных оценок параметров $u_0, u_1, \ldots, u_m$ на этапах «обучения» и «переобучения» модели (7.30) применим МНК.

Запишем модель (7.30) в виде двух равенств:

$$\mathbf{M}\,Y = u_0 + \sum_{j=1}^{m} u_j\,\mathbf{M}\,X_j, \tag{7.31}$$

$$Y^0 = \sum_{j=1}^{m} u_j\,X_j^0 + \varepsilon, \tag{7.32}$$

где $Y^0 = Y - \mathbf{M}\,Y$, $X_j^0 = X_j - \mathbf{M}\,X_j$ — центрированные величины.

Формулу для определения $u_1, \ldots, u_m$ получаем из (7.32), минимизируя дисперсию $\sigma^2(\varepsilon) = \mathbf{M}(Y^0 - \sum\limits_{j=1}^{m} u_j\, X_j^0)^2$ по $u_1, \ldots, u_m$:

$$u = K_x^{-1}\, \mathrm{cov}\,(Y, X), \tag{7.33}$$

где $u = (u_1, \ldots, u_m)^T$, K_x — положительно определенная ковариационная $(m \times m)$-матрица факторов $X_1, \ldots, X_m$; $\mathrm{cov}\,(Y, X) = (\mathrm{cov}\,(Y, X_1), \ldots, \mathrm{cov}\,(Y,\ X_m))^T$ — вектор, компоненты которого $\mathrm{cov}\,(Y, X_j) = \mathbf{M}\,(Y^0 X_j^0)$ — ковариации величины Y и факторов X_j, $j = 1, \ldots, m$.

На практике на этапах «обучения» или «переобучения» модели (7.30) следует использовать формулу (7.33) с заменой K_x и $\mathrm{cov}\,(Y, X)$ их оценками, полученными на основе «исторических» данных.

Если база «исторических» данных, используемых для «обучения», состоит из реализованных значений

$$Y_i, \quad X_{ij}, \quad j = 1, \ldots, m, \quad i = 1, \ldots, n, \tag{7.34}$$

то оценки элементов матрицы K_x и вектора $\mathrm{cov}\,(Y, X)$ определяются равенствами

$$(\widehat{K}_x)_{rl} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ir} - \overline{X}_r)(X_{il} - \overline{X}_l), \quad \overline{X}_r = \frac{1}{n} \sum_{i=1}^{n} X_{ir},$$

$$\widehat{\mathrm{cov}}\,(Y, X_r) = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})(X_{ir} - \overline{X}_r), \quad \overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i.$$

Таким образом, окончательные оценки параметров $u_0, u_1, \ldots, u_m$, полученные на этапе «обучения» модели (7.30), определяются равенствами

$$\widehat{u}_0 = \overline{Y} - \sum_{j=1}^{m} \widehat{u}_j\, \overline{X}_j, \quad \widehat{u} = (\widehat{u}_1, \ldots, \widehat{u}_m)^T = \widehat{K}_x^{-1} \widehat{\mathrm{cov}}\,(Y, X), \tag{7.35}$$

где $\widehat{\mathrm{cov}}\,(Y, X) = (\widehat{\mathrm{cov}}\,(Y, X_1), \ldots, \widehat{\mathrm{cov}}\,(Y, X_m))^T$.

После этапа «обучения» модели (7.30), определяемого формулами (7.35), модель (7.30) можно использовать для прогнозирования искомой величины Y согласно равенству

$$Y_{\text{прог}} = \widehat{u}_0 + \sum_{j=1}^{m} \widehat{u}_j\, X_j,$$

где $X_1, \ldots, X_m$ — значения наблюдаемых факторов на текущий момент времени.

Глава 8
РОБАСТНЫЕ МЕТОДЫ ДЛЯ ЛИНЕЙНЫХ МОДЕЛЕЙ С НЕОПРЕДЕЛЕННЫМИ ДАННЫМИ

Ранее уже отмечалось, что МНК-оценки являются эффективными в классе всех несмещенных оценок, если погрешности измерений компонент вектора f распределены по нормальному закону. Однако на практике часто нормальность закона распределения погрешностей будет нарушаться. Некоторые нарушения нормальности закона распределения могут приводить к значительной потере эффективности МНК-оценки и ее отклонению от истинных значений искомых параметров. Подчеркнем, что истинный закон распределения погрешностей измерений остается неизвестным.

Особенно большая потеря эффективности МНК-оценок происходит при наличии даже небольшой доли больших выбросов. На практике большим выбросам соответствуют те измерения f_i, реальная погрешность которых значительно превосходит приписываемую им дисперсию σ_i^2, причем ни реальная величина таких погрешностей, ни их номера i не известны.

При наличии больших выбросов необходимо применять робастные методы оценивания [83], позволяющие значительно снизить вредное влияние больших выбросов на оценку и получить приемлемую итоговую оценку искомых параметров.

8.1. Робастные М-оценки параметров линейных моделей

Существует несколько робастных методов оценивания [69, 81, 83, 88]. На их основе можно сконструировать различные робастные методы восстановления. Один из методов робастного восстановления основан на замене квадратичной нормы в экстремальных задачах (7.10), (7.21) на L_p-норму.

Вместо экстремальной задачи (7.10) рассмотрим экстремальную задачу

$$\widehat{u}_{\text{роб}} = \arg\min_u \sum_{i=1}^{n} |(Au)_i - f_i|^p, \tag{8.1}$$

где $(Au)_i$ — i-я компонента вектора Au, $1 \leqslant p < 2$.

При уменьшении p вплоть до 1 робастность оценки (8.1) увеличивается. Поскольку при $p > 1$ функционал правой части (8.1) сильно выпуклый, то существует единственная робастная оценка (8.1).

При $p \neq 2$ оценка (8.1) нелинейна, а ее фактическое вычисление значительно сложнее, чем вычисление МНК-оценки (7.11). Ниже будут даны конкретные алгоритмы вычисления нелинейных робастных оценок, в том числе и оценки (8.1).

Пусть выполняются все предположения классической схемы МНК, за исключением предположения $K_f = \sigma^2 I_n$, которое заменено на $K_f = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.

Тогда взвешенная МНК-оценка будет определяться равенством

$$\widehat{u}_{\text{МНК}} = \arg\min_u \sum_{i=1}^{n} ((Au)_i - f_i)^2 \frac{1}{\sigma_i^2}. \tag{8.2}$$

При наличии больших выбросов некоторые измерения f_i будут иметь большие значения дисперсий, чем соответствующие им дисперсии σ_i^2, причем ни номера i больших выбросов, ни истинные значения дисперсий $\sigma^2(f_i)$ не известны.

Введем М-класс робастных оценок, определяемых равенством

$$\widehat{u}_{\text{роб}} = \arg\min_u \sum_{i=1}^{n} \rho\Big(\frac{(Au)_i - f_i}{\sigma_i}\Big), \tag{8.3}$$

где $\rho(x)$ — выпуклая четная положительная функция, $\rho(0) = 0$.

Для того чтобы оценка (8.3) обладала робастностью по отношению к большим выбросам и не «портила» измерения, не являющиеся большими выбросами, функция $\rho(x)$ должна быть близкой к x^2 при малых значениях $|x|$ и менее возрастающей по сравнению с x^2 при больших значениях $|x|$.

В качестве конкретных функций $\rho(x)$, порождающих конкретные робастные оценки (8.3), можно брать, например, функции Андрюса, Рамсея, Хьюбера [69, 81, 83, 88]. Заметим, что L_p-оценка (8.1) является М-оценкой с $\rho(x) = |x|^p$. Поскольку для любой робастной М-оценки функция $\rho(x)$ будет отлична от x^2, любая робастная М-оценка нелинейна.

В силу сильной выпуклости функции $\rho(x)$ робастная М-оценка (8.3) существует и единственна.

Рассмотрим более подробно робастную оценку Хьюбера, которая является М-оценкой с

$$\rho(x) = \begin{cases} x^2/2, & |x| \leqslant K, \\ -K^2/2 + K|x|, & |x| > K, \end{cases}$$

где K — параметр Хьюбера [83].

Часто робастная оценка Хьюбера при наличии больших выбросов предпочтительней других робастных оценок, так как она является

минимаксной оценкой в классе плотностей, образуемом моделью смеси [83].

Множество индексов $I = \{i : i = 1, \ldots, n\}$ разобьем на три подмножества:

$$I_0 = \{i : |f_i - (A\widehat{u}_{\text{роб}})_i| \leqslant K\sigma_i\},$$
$$I_- = \{i : f_i - (A\widehat{u}_{\text{роб}})_i < -K\sigma_i\}, \tag{8.4}$$
$$I_+ = \{i : f_i - (A\widehat{u}_{\text{роб}})_i > K\sigma_i\}.$$

Очевидно, что $I = I_0 \cup I_- \cup I_+$.

Уравнение $\operatorname{grad}_u \sum\limits_{i=1}^{n} \rho((A\widehat{u}_{\text{роб}})_i - f_i)/\sigma_i) = 0$ для определения робастной оценки Хьюбера имеет вид

$$\sum_{i\in I_0} \frac{a_{ij}}{\sigma_i^2}\,((A\widehat{u}_{\text{роб}})_i - f_i) + \sum_{i\in I_-} \frac{Ka_{ij}}{\sigma_i} - \sum_{i\in I_+} \frac{Ka_{ij}}{\sigma_i} = 0, \quad j = 1, \ldots, m.$$

Последние уравнения перепишем в эквивалентном виде

$$\sum_{i\in I_0} \frac{a_{ij}}{\sigma_i^2}\,((A\widehat{u}_{\text{роб}})_i - f_i) + \sum_{i\in I_-} \frac{K\sigma_i a_{ij} + a_{ij}(A\widehat{u}_{\text{роб}})_i - a_{ij}(A\widehat{u}_{\text{роб}})_i}{\sigma_i^2} -$$
$$- \sum_{i\in I_+} \frac{K\sigma_i a_{ij} + a_{ij}(A\widehat{u}_{\text{роб}})_i - a_{ij}(A\widehat{u}_{\text{роб}})_i}{\sigma_i^2} = 0.$$

Отсюда получаем

$$\sum_{i=1}^{n} \frac{a_{ij}}{\sigma_i^2}\,(A\widehat{u}_{\text{роб}})_i =$$
$$= \sum_{i\in I_0} \frac{a_{ij}}{\sigma_i^2}\,f_i + \sum_{i\in I_-} \frac{a_{ij}}{\sigma_i^2}\,[(A\widehat{u}_{\text{роб}})_i - K\sigma_i] + \sum_{i\in I_+} \frac{a_{ij}}{\sigma_i^2}\,[(A\widehat{u}_{\text{роб}})_i + K\sigma_i],$$

$j = 1, \ldots, m$. Из последней системы следует, что $\widehat{u}_{\text{роб}}$ определяется равенством

$$\widehat{u}_{\text{роб}} = (A^T W A)^{-1} A^T W \widetilde{f}, \tag{8.5}$$

где $W = \operatorname{diag}(1/\sigma_1^2, \ldots, 1/\sigma_n^2)$, $\widetilde{f} = (\widetilde{f}_1, \ldots, \widetilde{f}_n)^T$; $\widetilde{f}_i = f_i$, если $i \in I_0$; $\widetilde{f}_i = (A\widehat{u}_{\text{роб}})_i - K\sigma_i$, если $i \in I_-$; $\widetilde{f}_i = (A\widehat{u}_{\text{роб}})_i + K\sigma_i$, если $i \in I_+$.

8.2. Численные методы нахождения робастных оценок параметров линейных моделей

Нахождение робастной оценки (8.3) связано с минимизацией функционала $\sum\limits_{i=1}^{n} \rho((Au)_i - f_i)/\sigma_i)$, которую можно осуществить любыми стандартными итерационными методами решения экстремальных задач без ограничений, например градиентными методами или методом

Ньютона. Однако учитывая специфику экстремальной задачи (8.3), можно предложить новые эффективные численные алгоритмы нахождения робастной оценки (8.3).

Предположим, что $\rho(x)$ дифференцируема. Тогда для нахождения оценки (8.3) имеем уравнение $\operatorname{grad}_u \sum_{i=1}^{n} \rho((Au)_i - f_i)/\sigma_i) = 0$, которое эквивалентно системе нелинейных уравнений

$$\sum_{i=1}^{n} \rho'\Big(\frac{(Au)_i - f_i}{\sigma_i}\Big)\frac{a_{ij}}{\sigma_i} = 0, \qquad j = 1, \ldots, m. \tag{8.6}$$

Систему (8.6) с учетом четности $\rho(x)$ запишем в эквивалентном виде

$$\sum_{i=1}^{n} W_i((Au)_i - f_i)a_{ij} = 0, \qquad j = 1, \ldots, m, \tag{8.7}$$

$$W_i = \frac{|\rho'((Au)_i - f_i)/\sigma_i)|}{|(Au)_i - f_i|\sigma_i}. \tag{8.8}$$

Если бы «веса» W_i были известны, то решение системы (8.7) давалось бы равенством

$$u = (A^T W A)^{-1} A^T W f, \tag{8.9}$$

где $W = \operatorname{diag}(W_1, \ldots, W_n)$.

На основе (8.8), (8.9) предлагается итерационный метод нахождения робастной оценки (8.3):

$$W^{(l)} = \operatorname{diag}(W_1^{(l)}, \ldots, W_n^{(l)}), \quad W_i^{(l)} = \frac{|\rho'(((Au^{(l)})_i - f_i)/\sigma_i)|}{|(Au^{(l)})_i - f_i|\sigma_i}, \tag{8.10}$$

$$u^{(l+1)} = (A^T W^{(l)} A)^{-1} A^T W^{(l)} f.$$

В качестве первого приближения рекомендуется брать средневзвешенную МНК-оценку $u^{(1)} = (A^T W^{(0)} A)^{-1} A^T W^{(0)} f$, где $W^{(0)} = \operatorname{diag}(1/\sigma_1^2, \ldots, 1/\sigma_n^2)$.

Итерационный метод (8.10) нахождения робастной оценки (8.3) называется *итеративным* МНК.

Можно предложить еще один вариант итерационного метода нахождения робастной оценки (8.3).

Экстремальную задачу (8.3) запишем в эквивалентном виде

$$\widehat{u}_{\text{роб}} = \arg\min_u \sum_{i=1}^{n} \widetilde{W}_i((Au)_i - f_i)^2, \tag{8.11}$$

где

$$\widetilde{W}_i = \frac{\rho((Au)_i - f_i)/\sigma_i)}{((Au)_i - f_i)^2}. \tag{8.12}$$

Если бы «веса» $\widetilde{W}_i$ были известны, то решение экстремальной задачи (8.11) давалось бы равенством

$$u = (A^T \widetilde{W} A)^{-1} A^T \widetilde{W} f, \tag{8.13}$$

где $\widetilde{W} = \operatorname{diag}(\widetilde{W}_1, \ldots, \widetilde{W}_n)$.

На основе (8.12), (8.13) предлагается еще один итерационный метод нахождения робастной оценки (8.3):

$$\widetilde{W}^{(l)} = \operatorname{diag}(\widetilde{W}_1^{(l)}, \ldots, \widetilde{W}_n^{(l)}), \quad \widetilde{W}_i^{(l)} = \frac{\rho(((Au^{(l)})_i - f_i)/\sigma_i)}{((Au^{(l)})_i - f_i)^2}, \tag{8.14}$$

$$u^{(l+1)} = (A^T \widetilde{W}^{(l)} A)^{-1} A^T \widetilde{W}^{(l)} f.$$

В качестве первого приближения рекомендуется брать средневзвешенную МНК-оценку $u^{(1)} = (A^T W^{(0)} A)^{-1} A^T W^{(0)} f$, где $W^{(0)} = \operatorname{diag}(1/\sigma_1^2, \ldots, 1/\sigma_n^2)$.

Итерационный метод (8.14) нахождения робастной оценки (8.3) называется *методом вариационно-взвешенных квадратических приближений.*

Если $\rho(x) = |x|^p$, $1 \leqslant p < 2$, т. е. рассматривается робастная L_p-оценка, то итеративный МНК (8.10) и метод вариационно-взвешенных квадратических приближений (8.14) эквивалентны.

Сходимость итераций (8.10), (8.14) для случая $\rho(x) = |x|^p$, $1 \leqslant p < 2$, доказана В. И. Мудровым и В. Л. Кушко. В общем случае при решении конкретных задач метод вариационно-взвешенных квадратических приближений часто дает очень быструю сходимость — иногда достаточно 1-2 итераций для достижения приемлемой точности.

Форма записи системы (8.5), определяющей робастную оценку Хьюбера, подсказывает следующий итерационный процесс (ИП):

$$u^{(l+1)} = (A^T W A)^{-1} A^T W f^{(l)}, \tag{8.15}$$

где

$$f^{(l)} = (f_1^{(l)}, \ldots, f_n^{(l)})^T; \quad f_i^{(l)} = f_i, \quad \text{если } i \in I_0^{(l)};$$

$$f_i^{(l)} = (Au^{(l)})_i + K\sigma_i, \quad \text{если } i \in I_+^{(l)};$$

$$f_i^{(l)} = (Au^{(l)})_i - K\sigma_i, \quad \text{если } i \in I_-^{(l)};$$

$$I_0^{(l)} = \{i : |(Au^{(l)})_i - f_i| \leqslant K\sigma_i\}, \quad I_-^{(l)} = \{i : f_i - (Au^{(l)})_i < -K\sigma_i\},$$

$$I_+^{(l)} = \{i : f_i - (Au^{(l)})_i > K\sigma_i\}.$$

В качестве первого приближения рекомендуется брать средневзвешенную МНК-оценку.

Имеется принципиальное преимущество ИП (8.15) по сравнению с ИП (8.10) и (8.14). Действительно, на каждом шаге ИП (8.10), (8.14)

приходится пересчитывать матрицы $(A^T W^{(l)} A)^{-1} A^T W^{(l)}$, в том числе обращая матрицу $A^T W^{(l)} A$. В ИП (8.15) формирование матрицы $(A^T W A)^{-1} A^T W$ производится только один раз, а затем эта матрица «обслуживает» все итерации. На каждом шаге процесса (8.15) производится только перерасчет вектора $f^{(l)}$, что требует значительно меньших вычислительных затрат по сравнению с перерасчетом матрицы $(A^T W^{(l)} A)^{-1} A^T W^{(l)}$.

Итерационный процесс (8.15) легко программируется и сходится за конечное число итераций. Практика применения ИП (8.15) показала его высокую эффективность. Для достижения приемлемой точности требуется, как правило, 1–10 итераций. Рекомендации по выбору величин параметра K остаются такими же, как и в § 3.1.

Заметим, что ИП (8.15) позволяет автоматически выявлять большие выбросы, для этого в компьютерной программе достаточно предусмотреть фиксацию тех номеров i, для которых измерения f_i попадают в множества $I_+^{(l)}$ и $I_-^{(l)}$.

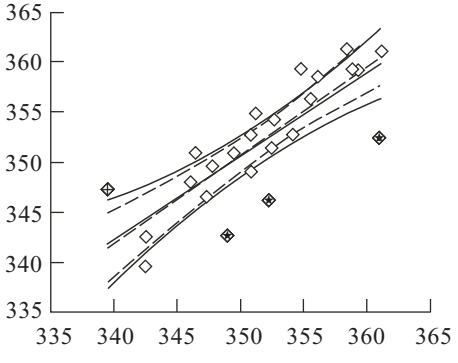

Рис. 8.1

На рис. 8.1 представлены результаты применения численных методов нахождения параметров линейных моделей. Сплошные линии соответствуют линейной регрессии по МНК и доверительному интервалу с уровнем доверия 95 %. Пунктиром показаны робастная линейная регрессия и доверительный интервал с уровнем доверия 95 %. Символами $(+, \bigstar)$ помечены значения с $i \in I_+$ и $i \in I_-$.

Робастные методы для линейных моделей представлены также в работах [7–9, 69, 84, 88].

Глава 9
УЧЕТ АПРИОРНОЙ ИНФОРМАЦИИ В ЛИНЕЙНЫХ МОДЕЛЯХ С НЕОПРЕДЕЛЕННЫМИ ДАННЫМИ

В настоящей главе будут рассмотрены классические и сравнительно новые методы учета дополнительной априорной информации об искомом векторе u. Эти методы позволяют получать приемлемые оценки вектора u даже в тех условиях, когда в исходной регрессионной модели (7.2) «заложено» недостаточно информации о векторе u.

Выше были рассмотрены только несмещенные (или, по крайней мере, асимптотически несмещенные) оценки искомого вектора u и дополнительных неизвестных параметров, если таковые имеются. Это и естественно, поскольку, если относительно u нет никакой другой информации, кроме «заложенной» в регрессионной модели (7.2), то смещенное оценивание не позволяет определить погрешность оценки. Действительно, вектор погрешности оценки $\delta = u - u_T$ можно представить в виде $\delta = b(\widehat{u}, u_T) + \eta(\widehat{u})$, где $b(\widehat{u}, u_T) = \mathbf{M}\,\widehat{u} - u_T$ — вектор смещения, $\eta(\widehat{u}) = \widehat{u} - \mathbf{M}\,\widehat{u}$ — вектор статистической погрешности оценки $\widehat{u}$. Вектор смещения зависит от искомого вектора u_T. Вектор же статистической погрешности оценки от искомого вектора не зависит, он определяется лишь выбранной оценкой. Если нет никакой априорной информации об искомом векторе, то нельзя оценить вектор смещения $b(\widehat{u}, u_T)$.

Если регрессионная модель (7.2) принадлежит к классу некорректно поставленных задач, то информационная матрица Фишера либо вырождена, либо плохо обусловлена, и поэтому даже оптимальная МНК-оценка либо не существует, либо дает слишком большое значение погрешности оценки. Поскольку любая другая несмещенная оценка будет давать еще большие значения погрешности оценки, то при решении некорректных задач восстановления в классе несмещенных оценок не существует «хороших» — все несмещенные оценки «плохие».

Подчеркнем еще раз, что речь идет о несмещенных оценках, использующих информацию об искомом векторе, «заложенную» только в регрессионной модели (7.2).

Итак, при решении некорректных задач восстановления необходимо искать разумные оценки (тем более «хорошие» оценки) в классе оценок, использующих дополнительную априорную информацию об искомом векторе u, помимо «заложенной» в регрессионной модели (7.2). Подчеркнем, что некорректность задачи восстановления, основанной на регрессионной модели (7.2), эквивалентна вырожденности или пло-

хой обусловленности информационной матрицы Фишера. Последнее означает, что в модели (7.2) «заложено» недостаточно информации об искомом векторе, и поэтому принципиально нельзя получить разумные оценки решения, оставаясь только в рамках модели (7.2).

Следовательно, отличие корректной задачи восстановления от некорректной заключается в том, что для корректной задачи существуют приемлемые оценки искомого вектора без привлечения априорной информации, для некорректной же задачи невозможно получение приемлемых оценок искомого вектора без привлечения априорной информации.

Априорная информация об искомом векторе u может быть двух видов — статистической и детерминированной. В зависимости от вида априорной информации используют два классических метода решения некорректных задач восстановления. Если в распоряжении исследователя имеется априорная информация об искомом векторе статистического характера, то можно использовать метод Байеса. Если же априорная информация об искомом векторе имеет детерминированный характер, то можно использовать минимаксный метод.

Подчеркнем, что любой метод оценивания искомого вектора u, основанный на использовании дополнительной априорной информации, можно применять и при решении корректных задач восстановления. Но если при решении корректных задач такие методы (в частности, метод Байеса и минимаксный метод) позволяют лишь улучшать качество оценок (на основе использования дополнительной априорной информации) и можно получать приемлемые оценки и без привлечения этих методов, то при решении некорректных задач без использования методов, позволяющих учесть априорную информацию об искомом векторе (в частности, методов Байеса и минимаксного), принципиально невозможно получение приемлемых оценок искомого вектора.

9.1. Метод Байеса в рамках линейных моделей

Метод Байеса основан на трактовке искомого вектора u как случайного вектора, для которого априорная информация задается в виде априорной плотности вероятностей $p(u)$. Тогда апостериорная плотность вероятностей $p(u \mid f)$ подсчитывается по формуле Байеса

$$p(u \mid f) = \frac{p(u)p(f \mid u)}{\int p(u)p(f \mid u)\, du}, \tag{9.1}$$

где $p(f \mid u)$ — плотность вероятностей случайного вектора f регрессионной модели (7.2).

В качестве байесовской оценки (Б-оценки) искомого вектора берут подходящую характеристику апостериорного распределения (при той реализации вектора f, которая зафиксирована в регрессионной модели (7.2)). Наиболее часто в качестве Б-оценки выбирают: среднее апо-

стериорное значение, т. е. $\mathbf{M}\,u$, где u — случайный вектор с плотностью $p(u \mid f)$; медиану случайного вектора u; наиболее вероятное значение, т. е. моду случайного вектора u.

Подчеркнем, что при применении байесовского подхода необходимо задание двух законов: распределения вектора f (совпадающего с точностью до математического ожидания с законом распределения вектора случайных погрешностей измерений ε) и априорного. Если законы неизвестны, то приходится вводить гипотезы, согласно которым постулируются законы распределения f и априорный. Конечно, в этом случае правильность полученных оценок во многом зависит от правильности «угадывания» законов распределения.

Предположим, что законы распределения вектора f и априорный нормальные, т. е.

$$p(f \mid u) = \frac{1}{(2\pi)^{n/2}\sigma^n|\Omega|^{1/2}} \exp\Big\{-\frac{1}{2\sigma^2}(f-Au)^T\Omega^{-1}(f-Au)\Big\},$$

$$p(u) = \frac{1}{(2\pi)^{m/2}\sigma_a^m|K_a|^{1/2}} \exp\Big\{-\frac{1}{2\sigma_a^2}(u-u_0)^T K_a^{-1}(u-u_0)\Big\},$$

где u_0, $\sigma_a^2 K_a$ — центр и ковариационная матрица априорного распределения.

Для рассматриваемого случая, когда законы распределения вектора f и априорный нормальные, все три вышеуказанных варианта Б-оценок искомого вектора u совпадают и определяются формулой

$$\widehat{u}_{\text{Б}} = \Big(\frac{1}{\sigma^2}A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}K_a^{-1}\Big)^{-1}\Big(\frac{1}{\sigma^2}A^T\Omega^{-1}f + \frac{1}{\sigma_a^2}K_a^{-1}u_0\Big). \quad (9.2)$$

В частности, если $u_0 = 0$, то

$$\widehat{u}_{\text{Б}} = \Big(\frac{1}{\sigma^2}A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}K_a^{-1}\Big)^{-1}A^T\Omega^{-1}f. \quad (9.3)$$

Если $u_0 \neq u_T$, то Б-оценка смещена. Действительно,

$$\mathbf{M}\,\widehat{u}_{\text{Б}} = \left(\frac{1}{\sigma^2}A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}K_a^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}A^T\Omega^{-1}Au_T + \frac{1}{\sigma_a^2}K_a^{-1}u_0\right) =$$

$$= u_T + \left(\frac{1}{\sigma^2}A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}K_a^{-1}\right)^{-1}\frac{1}{\sigma_a^2}K_a^{-1}(u_0-u_T) \neq u_T,$$

если $u_0 \neq u_T$.

Таким образом, смещение Б-оценки дается равенством

$$b(\widehat{u}_{\text{Б}}, u_T) = \frac{1}{\sigma_a^2}\Big(\frac{1}{\sigma^2}A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}K_a^{-1}\Big)^{-1}K_a^{-1}(u_0-u_T).$$

Чем лучше «угадан» вектор u_0, тем меньше величина смещения.

Подсчитаем ковариационную матрицу Б-оценки (9.2). Имеем

$$K_{\hat{u}_\text{Б}} = \left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1}\frac{1}{\sigma^2}\,A^T\Omega^{-1}\Omega\Omega^{-1}A \times$$

$$\times\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1} = \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma^2}\,K_a^{-1}\right)^{-1} \times$$

$$\times\,A^T\Omega^{-1}A\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1} =$$

$$= \left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1} - \frac{1}{\sigma_a^2}\,K_a^{-1}\right) \times$$

$$\times\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1} = \left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1} -$$

$$-\,\frac{1}{\sigma_a^2}\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1}K_a^{-1}\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A^T + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1}.$$

Но симметричная матрица

$$\frac{1}{\sigma_a^2}\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1}K_a^{-1}\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1}$$

положительно определена. Поэтому

$$K_{\hat{u}_\text{Б}} < \left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}\right)^{-1} < \sigma^2(A^T\Omega^{-1}A)^{-1} = K_{\hat{u}_\text{МНК}}.$$

Итак, ковариационная матрица Б-оценки меньше ковариационной матрицы взвешенной МНК-оценки, и тем самым статистическая погрешность Б-оценки меньше статистической погрешности взвешенной МНК-оценки. Последнее связано с тем, что Б-оценка, кроме информации, «заложенной» в регрессионной модели (7.2), использует дополнительную информацию об искомом векторе u, задаваемую в виде априорного распределения $p(u)$, а взвешенная МНК-оценка использует информацию, «заложенную» только в регрессионной модели (7.2).

Б-оценка (9.2) всегда существует, вне зависимости от того, корректна или некорректна задача восстановления (7.2). Действительно, матрица $\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\,K_a^{-1}$ положительно определена вне зависимости от обусловленности матрицы Фишера $\frac{1}{\sigma^2}\,A^T\Omega^{-1}A$ за счет слагаемого $\frac{1}{\sigma_a^2}\,K_a^{-1}$.

Теоретические основы метода Байеса представлены в работах [26, 33].

В настоящее время метод Байеса широко применяется при решении некорректных задач, в том числе некорректных задач обработки экспериментальных данных и задач восстановления [7, 22, 30, 31, 47, 55, 74].

9.2. Минимаксный метод учета детерминированной априорной информации

Рассмотрим теперь второй классический метод учета дополнительной априорной информации — *минимаксный метод*.

Минимаксный метод используют тогда, когда априорная информация об искомом векторе u задается в детерминированной форме.

Пусть априорная информация задана соотношением $u \in R_a$, где R_a — некоторое множество пространства R^m.

Рассмотрим класс $\{\widehat{u}\}$ линейных оценок искомого вектора $\widehat{u} = Sf + u^*$, где S — детерминированная $(m \times n)$-матрица, u^* — m-мерный вектор, причем S, u^* пока произвольны.

Введем функционал

$$J(\widehat{u}) = \max_{u \in R_a} \mathbf{M} \|\widehat{u} - u\|^2. \tag{9.4}$$

Заметим, что функционал $J(u)$ определяется, в частности, выбранной нормой; взяв в (9.4) конкретную норму, получаем один из вариантов минимаксного метода. Чаще всего в (9.4) берут евклидову норму.

Согласно минимаксному методу в качестве оценки u, определяемой матрицей S и вектором u^*, берут такой вектор, который минимизирует функционал (9.4), т. е. минимаксная оценка (ММ-оценка) определяется равенством

$$\widehat{u}_{\text{ММ}} = \arg \min_{\widehat{u} \in \{\widehat{u}\}} J(\widehat{u}). \tag{9.5}$$

Если априорная информация отсутствует, т. е. $R_a = R^m$, и в (9.4) взята евклидова норма, то ММ-оценка (9.5) совпадает с МНК-оценкой.

Минимаксный метод дает наилучшую оценку при наихудшем возможном расположении искомого вектора u в априорном множестве R_a.

Нахождение ММ-оценки для конкретных множеств R_a — трудоемкая вычислительная задача. В предположении, что норма в (9.4) евклидова, для двух частных видов множества R_a можно найти явный вид ММ-оценки.

Пусть R_a — m-мерный эллипсоид с центром в u_0 и «радиуса» ρ:

$$R_a = \{u : (u - u_0)^T B (u - u_0) \leqslant \rho^2\}, \tag{9.6}$$

где B — симметричная положительно определенная матрица.

Тогда ММ-оценка дается равенством

$$\widehat{u}_{\text{ММ}} = u_0 + \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\rho^2} B'\right)^{-1} A^T \Omega^{-1} (f - A u_0), \tag{9.7}$$

где модифицированная матрица B' такова, что $\widehat{u}_{\text{ММ}}$ принадлежит границе множества R_a.

В частности, если $u_0 = 0$, то

$$\widehat{u}_{\text{ММ}} = \left(A^T\Omega^{-1}A + \frac{\sigma^2}{\rho^2}\,B'\right)^{-1} A^T\Omega^{-1}f. \tag{9.8}$$

Пусть теперь R_a — m-мерный параллелепипед с центром в u_0:

$$R_a = \{u:\ (u-u_0,\varphi_k)^2 \leqslant \lambda_k^2,\ k=1,\ldots,m\}, \tag{9.9}$$

где $\{\varphi_k\}$ — ортонормированный базис пространства R^m, $\{\lambda_k^2\}$ — заданные положительные числа,

Обозначим

$$B = U\,\mathrm{diag}\,(1/\lambda_1^2,\ldots,1/\lambda_m^2)U^T,$$

где $U = (\varphi_1,\ldots,\varphi_m)$ — ортогональная матрица, k-й столбец которой образован m компонентами вектора φ_k. Матрица B симметрична и положительно определена, причем $\{1/\lambda_k^2\}$, $\{\varphi_k\}$, $k=1,\ldots,m$ — совокупность собственных значений и соответствующих им собственных векторов матрицы B.

Для параллелепипеда (9.9) ММ-оценка дается равенством

$$\widehat{u}_{\text{ММ}} = u_0 + \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\,A^T\Omega^{-1}A + B'\right)^{-1} A^T\Omega^{-1}(f - Au_0), \tag{9.10}$$

где B' - модифицированная матрица B [79].

В частности, если $u_0 = 0$, то

$$\widehat{u}_{\text{ММ}} = (A^T\Omega^{-1}A + \sigma^2 B')^{-1}A^T\Omega^{-1}f. \tag{9.11}$$

В приложениях априорное множество R_a часто задается в виде совокупности неравенств:

$$a_i \leqslant u_i \leqslant b_i, \qquad i=1,\ldots,m, \tag{9.12}$$

где a_i, b_i — заданные числа.

Совокупность неравенств (9.12) определяет параллелепипед (9.9), для которого

$$u_0 = \frac{1}{2}\,(a_1+b_1,\ldots,a_m+b_m)^T,$$

$$B = 4\,\mathrm{diag}\left(\frac{1}{(b_1-a_1)^2},\ldots,\frac{1}{(b_m-a_m)^2}\right),$$

$$U = I_m, \quad \varphi_k = l_k = (0,\ldots,1,\ldots,0)^T, \quad \lambda_k^2 = \frac{1}{4}\,(b_k-a_k)^2.$$

Если множество R_a ограничено и выпукло, то ММ-оценка существует и единственна для любой регрессионной модели (7.2), в том числе, когда задача восстановления некорректна.

ММ-оценки смещены, причем смещение не всегда приближает оценку к точному вектору u_T. Чем «у́же» априорное множество R_a, тем ближе ММ-оценка к u_T. Например, ММ-оценки (9.7), (9.10) стремятся к u_T в среднем квадратичном, если радиус эллипсоида $\rho \to 0$ (для оценки (9.7)) или $\max \lambda_k^2 \to 0$ (для оценки (9.10)). Наоборот, если $\rho \to \infty$ (для оценки (9.7)) или $\min \lambda_k^2 \to \infty$ (для оценки (9.10)), то ММ-оценки (9.7), (9.10) стремятся в среднем квадратичном к МНК-оценке (7.22).

Если A — матрица неполного ранга ($\operatorname{rank} A < m$), т. е. задача восстановления (7.2) некорректна, то при $\sigma^2 \to 0$ ММ-оценки сходятся в среднем квадратичном к одному из решений системы $Au = f_T$.

Поскольку ММ-оценки смещены, точность ММ-оценки можно определить с помощью симметричной неотрицательной $(m \times m)$-матрицы $\mathbf{D}(\widehat{u}_{\text{ММ}}) = \mathbf{M}(\widehat{u}_{\text{ММ}} - u_T)(\widehat{u}_{\text{ММ}} - u_T)^T$, в частности, i-й элемент главной диагонали матрицы $\mathbf{D}(\widehat{u}_{\text{ММ}})$ равен $\mathbf{M}(\widehat{u}_{i\text{ММ}} - u_i)^2$, где $\widehat{u}_{i\text{ММ}}$, u_i — i-е компоненты векторов $\widehat{u}_{\text{ММ}}$ и u_T соответственно. Следовательно, $\operatorname{Sp} \mathbf{D}(\widehat{u}_{\text{ММ}}) = \sum\limits_{i=1}^{m} \mathbf{M}(\widehat{u}_{i\text{ММ}} - u_i)^2$ — среднеквадратическая ошибка ММ-оценки.

Для ММ-оценок (9.7), (9.10)

$$\mathbf{D}(\widehat{u}_{\text{ММ}}) = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\rho^2} B'\right)^{-1},$$

$$\mathbf{D}(\widehat{u}_{\text{ММ}}) = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + B'\right)^{-1}$$

соответственно.

Поскольку матрицы B' в обеих ММ-оценках (9.7), (9.10) положительно определены, то справедливо неравенство $\mathbf{D}(\widehat{u}_{\text{ММ}}) < K_{\widehat{u}_{\text{МНК}}}$ и для оценки (9.7), и для оценки (9.10), где $K_{\widehat{u}_{\text{МНК}}} = \sigma^2 (A^T \Omega^{-1} A)^{-1}$ — ковариационная матрица МНК-оценки (7.22).

Итак, ММ-оценки (9.7), (9.10) эффективней МНК-оценки (7.22). Это и неудивительно, так как ММ-оценки используют дополнительную информацию об искомом векторе u_T.

Для фактического нахождения ММ-оценок (9.7), (9.10) необходимо обратить симметричные положительно определенные $(m \times m)$-матрицы $\left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\rho^2} B'\right)$, $\left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + B'\right)$ соответственно. Для обращения этих матриц рекомендуется пользоваться методом квадратного корня.

ММ-оценки (9.8), (9.11) эквивалентны смещенным ридж-оценкам, широко используемым, согласно публикациям в иностранной периодике, при решении некорректных задач восстановления. К сожалению, приходится констатировать, что в публикациях по ридж-оцениванию введение стабилизатора (эквивалентного положительно определенным

симметричным матрицам B' в ММ-оценках (9.7), (9.10)) никак не связывается с использованием априорной информации об искомом векторе u, а производится формально с единственной целью — выйти из класса несмещенных оценок [27].

Заметим также, что если в ММ-оценках (9.8), (9.11) положить $\Omega = I_n$, то формально получаем приближенное решение согласно вариационному методу регуляризации А. Н. Тихонова (с квадратичным стабилизатором) [71].

Минимаксный метод решения некорректно поставленных задач представлен в [79].

9.3. Обобщенный метод максимального правдоподобия учета априорной стохастической информации в рамках линейных моделей

В предыдущем параграфе уже отмечалось, что невозможно получить приемлемые оценки искомого вектора u некорректной задачи восстановления без учета дополнительной априорной информации о векторе u, причем априорная информация может задаваться в двух формах — статистической или детерминированной.

В настоящем параграфе используется схема учета априорной информации, отличная от традиционных: байесовской — в случае статистического характера априорной информации, и минимаксной — в случае детерминированного характера априорной информации. Предлагаемая схема базируется на классическом методе максимального правдоподобия (ММП) Фишера.

Рассмотрим сначала классическую схему применения ММП для нахождения ММП-оценки искомого вектора и регрессионной модели (7.2).

Пусть, как и прежде, $p_\varepsilon(x)$ — плотность вероятностей вектора погрешности ε. Введем функцию правдоподобия $L(u) = p_\varepsilon(f - Au)$, где f — известная реализация из (7.2).

Согласно ММП за оценку искомого вектора u принимается вектор $\widehat{u}_{\text{ММП}}$, максимизирующий функцию правдоподобия:

$$\widehat{u}_{\text{ММП}} = \arg\max_u L(u). \tag{9.13}$$

Часто вместо $L(u)$ удобно максимизировать логарифмическую функцию правдоподобия $\ln L(u)$. Ясно, что решения экстремальных задач (9.13) и $\widehat{u} = \arg\max_u \ln L(u)$ совпадают, поскольку функция $\ln v$, $v > 0$, — монотонно возрастающая и выпуклая.

Введем информационную $(m \times m)$-матрицу Фишера M,

$$M_{ij} = -\mathbf{M}\,\frac{\partial^2 \ln L(u)}{\partial u_i \partial u_j}, \quad i, j = 1, \ldots, m.$$

Матрица Фишера представима также в виде

$$M = \mathbf{M}\,(\operatorname{grad}_u \ln\, L(u))(\operatorname{grad}_u \ln\, L(u))^T.$$

Поскольку $\mathbf{M}\operatorname{grad}_u \ln\, L(u) = 0$, то M — ковариационная матрица случайного вектора $\operatorname{grad}_u \ln\, L(u)$. Таким образом, матрица Фишера симметрична и неотрицательна.

Если информационная матрица Фишера положительно определена, то ММП-оценка существует. Более того, если плотность $p_\varepsilon(x)$ принадлежит к экспоненциальному классу [33], то ММП-оценка единственна и является единственным решением системы уравнений правдоподобия

$$\operatorname{grad}_u \ln\, L(u) = 0. \tag{9.14}$$

Напомним, что ММП-оценки обладают оптимальными свойствами среди всех асимптотически несмещенных оценок. Оптимальность ММП-оценок означает их асимптотическую эффективность, т. е. приближение при увеличении объема выборки n к нижней границе точности несмещенных оценок, определяемой неравенством Фишера–Крамера–Рао

$$K_{\hat{u}} \geqslant M^{-1}, \tag{9.15}$$

где $K_{\hat{u}}$ — ковариационная матрица любой несмещенной оценки искомого вектора u_T, построенной на основе регрессионной модели (7.2).

Более того, выше было показано, что для нормальных законов распределения погрешностей ε, когда

$$p_\varepsilon(x) = \frac{1}{(2\pi)^{n/2}\sigma^n|\Omega|^{1/2}}\,\exp\left[-\frac{x^T\Omega^{-1}x}{2\sigma^2}\right],$$

МНК-оценка (7.22) совпадает с ММП-оценкой, $M = \dfrac{1}{\sigma^2}\,A^T\Omega^{-1}A$, и ММП-оценка является эффективной, т. е. для оценки (7.22) неравенство Фишера–Крамера–Рао (9.15) обращается в равенство.

Если задача восстановления (7.2) некорректна, то информационная матрица Фишера плохо обусловлена или вырождена. Последнее означает, что в регрессионной модели (7.2) «заложено» недостаточно информации об искомом векторе u_T. Из неравенства Фишера–Крамера–Рао (9.15) следует, что в этом случае нельзя получить приемлемые оценки искомого вектора, используя только информацию, «заложенную» в (7.2).

Итак, в условиях некорректности задачи восстановления (7.2) для получения приемлемых оценок искомого вектора необходима дополнительная априорная информация об искомом векторе, причем «объем» дополнительной информации должен быть таков, что при ее присоединении к информации Фишера, определяемой только по модели (7.2), суммарной информации должно быть достаточно для получения приемлемой оценки искомого вектора.

Рассмотрим случай статистического характера априорной информации. Пусть задана случайная величина ξ, независимая от случайной

величины ε из (7.2), плотность вероятностей которой параметрически зависит от искомого вектора u. Пусть далее известна выборка

$$Y_1, \ldots, Y_N \tag{9.16}$$

значений случайной величины ξ. Обозначим $p(y_1, \ldots, y_N;\ u)$ совместную плотность вероятностей случайных величин $Y_1, \ldots, Y_N$. В частности, если выборка (9.16) повторная, то

$$p(y_1, \ldots, y_N;\ u) = \prod_{i=1}^{N} p_\xi(y_i;\ u),$$

где $p(y;\ u)$ — плотность вероятностей случайной величины ξ.

Выборку $Y = (Y_1, \ldots, Y_N)^T$ и функцию $L_a(u) = p(Y_1, \ldots, Y_N;\ u)$ назовем *априорной выборкой* и *априорной функцией правдоподобия* соответственно.

Совместная плотность вероятности совокупности случайных величин $\varepsilon, Y_1, \ldots, Y_N$ равна $p_\varepsilon(x)p(y_1, \ldots, y_N;\ u)$. Функцию

$$L_0(u) = p_\varepsilon(f - Au)p(Y_1, \ldots, Y_N;\ u)$$

назовем *обобщенной функцией правдоподобия.*

Имеем

$$L_0(u) = L(u)\, L_a(u), \tag{9.17}$$

где $L(u)$ — функция правдоподобия регрессионной модели (7.2), $L_a(u)$ — априорная функция правдоподобия.

Из (9.17) следует, что логарифмическая обобщенная функция правдоподобия равна сумме логарифмических функций правдоподобия модели (7.2) и априорной выборки,

$$\ln\, L_0(u) = \ln\, L(u) + \ln\, L_a(u). \tag{9.18}$$

Введем априорную и обобщенную информационные $(m \times m)$-матрицы Фишера M_a, M_0,

$$(M_a)_{ij} = -\mathbf{M}\, \frac{\partial^2 \ln\, L_a(u)}{\partial u_i \partial u_j}, \quad (M_0)ij = -\mathbf{M}\, \frac{\partial^2 \ln\, L_0(u)}{\partial u_i \partial u_j},$$

$i, j = 1, \ldots, m$.

Из (9.18) следует формула

$$M_0 = M + M_a, \tag{9.19}$$

где M — информационная матрица Фишера регрессионной модели (7.2).

Обобщенная информационная матрица Фишера M_0 суммирует информацию об искомом векторе u_T, «заложенную» в исходной регрессионной модели (7.2), и априорную информацию.

Введем обобщенный метод максимального правдоподобия (ОММП), согласно которому за оценку искомого вектора u_T при-

нимается вектор $\widehat{u}_{\text{ОММП}}$, максимизирующий обобщенную функцию правдоподобия:

$$\widehat{u}_{\text{ОММП}} = \arg\max_u L_0(u). \tag{9.20}$$

Очевидно, что решения экстремальных задач (9.20) и $\widehat{u} = \arg\max_u \ln L_0(u)$ совпадают.

ОММП-оценка является решением системы обобщенных уравнений правдоподобия,

$$\operatorname{grad}_u \ln L(u) + \operatorname{grad}_u \ln L_a(u) = 0. \tag{9.21}$$

При решении практических задач часто неизвестными являются не только компоненты вектора u, но и некоторые характеристики законов распределения случайных величин ε и ξ. Тогда согласно ОММП вместо (9.20) необходимо решать экстремальную задачу

$$(\widehat{u}_{\text{ОММП}}, \widehat{\alpha}_{\text{ОММП}}) = \arg\max_{(u,\alpha)} L_0(u, \alpha), \tag{9.22}$$

где $\alpha = (\alpha_1, \ldots, \alpha_l)^T$ — вектор совокупности дополнительных неизвестных параметров, $L_0(u, \alpha) = L(u, \alpha) + L_a(u, \alpha)$, а вместо информационных $(m \times m)$-матриц M_0, M, M_a необходимо рассматривать расширенные информационные $((m + l) \times (m + l))$-матрицы $\widetilde{M}_0$, $\widetilde{M}$, $\widetilde{M}_a$, определяемые равенствами

$$(\widetilde{M}_0)_{ij} = -\mathbf{M}\,\frac{\partial^2 \ln L_0(u, \alpha)}{\partial v_i \partial v_j}, \quad (\widetilde{M})_{ij} = -\mathbf{M}\,\frac{\partial^2 \ln L(u, \alpha)}{\partial v_i \partial v_j},$$

$$(\widetilde{M}_a)_{ij} = -\mathbf{M}\,\frac{\partial^2 \ln L_a(u, \alpha)}{\partial v_i \partial v_j}, \qquad i, j = 1, \ldots, m + l,$$

$$v = (u_1, \ldots, u_m, \alpha_1, \ldots, \alpha_l)^T \equiv (v_1, \ldots, v_{m+l})^T.$$

Рассмотрим несколько наиболее важных при решении практических задач частных случаев ОММП-оценок.

Пусть $\varepsilon \sim \mathcal{N}(0, \sigma^2\Omega)$, каждый член Y_i априорной выборки (9.16) распределен по нормальному закону $Y_i \sim \mathcal{N}(u, \sigma_i^2 K_i)$, K_i — положительно определенные матрицы. Тогда

$$L(u) = \frac{1}{(2\pi)^{n/2}\sigma^n |\Omega|^{1/2}} \exp\Big\{ -\frac{1}{2\sigma^2} (f - Au)^T \Omega^{-1} (f - Au) \Big\},$$

$$L_a(u) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{m/2}\sigma_i^m |K_i|^{1/2}} \exp\Big\{ -\frac{1}{2\sigma_i^2} (Y_i - u)^T K_i^{-1} (Y_i - u) \Big\},$$

$$M = \frac{1}{\sigma^2} A^T \Omega^{-1} A, \quad M_a = \sum_{i=1}^{N} \frac{1}{\sigma^2} K_i^{-1}.$$

Имеем

$$M_0 = \frac{1}{\sigma^2} A^T \Omega^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1}. \tag{9.23}$$

Из формулы (9.23) видно, что обобщенная информационная матрица равна сумме информационной матрицы исходной регрессионной модели (7.2) и априорных информационных матриц, соответствующих каждому члену априорной выборки.

Даже если исходная задача восстановления (7.2) некорректно поставлена и тем самым информационная матрица M плохо обусловлена или вырождена, обобщенная информационная матрица M_0, суммирующая априорную информацию об искомом векторе u_T и информацию, «заложенную» в регрессионной модели (7.2), является обратимой, причем ее обусловленность улучшается либо при уменьшении каждого из σ_i^2, либо при увеличении объема априорной выборки N.

Для рассматриваемого случая решение системы обобщенных уравнений правдоподобия (9.21) существует, единственно и дается формулой

$$\widehat{u}_{\text{ОММП}} = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1}\right)^{-1} \times \\ \times \left(\frac{1}{\sigma^2} A^T \Omega^{-1} f + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1} Y_i\right). \tag{9.24}$$

В частности, если $Y_i \sim \mathcal{N}(u, \sigma_a^2 K_a)$, то

$$\widehat{u}_{\text{ОММП}} = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{N}{\sigma_a^2} K_a^{-1}\right)^{-1} \left(\frac{1}{\sigma^2} A^T \Omega^{-1} f + \frac{1}{\sigma_a^2} K_a^{-1} \sum_{i=1}^{N} Y_i\right). \tag{9.25}$$

Наконец, если $N = 1$, т. е. априорная выборка состоит из одной реализации, u_0, то

$$\widehat{u}_{\text{ОММП}} = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} K_a^{-1}\right)^{-1} \left(\frac{1}{\sigma^2} A^T \Omega^{-1} f + \frac{1}{\sigma_a^2} K_a^{-1} u_0\right). \tag{9.26}$$

Сравнение ОММП-оценки (9.26) и Б-оценки (9.2) показывает их полную внешнюю идентичность, хотя смысл входящих в них характеристик σ_a^2, K_a, u_0 различен. Таким образом, для частного случая, когда имеется лишь одна реализация априорной случайной величины, а законы распределения случайных величин ε, ξ и априорного (в рамках байесовского подхода) нормальны, ОММП-оценка (9.26) формально совпадает с Б-оценкой (9.2). Формальное совпадение ОММП- и Б-оценок при одной реализации априорной случайной величины имеет место и для законов распределения, отличных от нормального, если в качестве Б-оценки берется мода апостериорного распределения.

Иногда известна не одна реализация f правой части регрессионной модели (7.2), а несколько ее реализаций $f_1, \dots, f_r$. Пусть, например, $f_j \sim \mathcal{N}(f_T, \sigma_{fj}^2 K_{fj})$. Тогда

$$L(u) = \frac{1}{(2\pi)^{nr/2}} \prod_{j=1}^{r} \frac{1}{\sigma_{fj}^n |K_{fj}|^{1/2}} \exp\Big\{ -\frac{1}{2\sigma_{fj}^2} (f_j - Au)^T K_{fj}^{-1} (f - Au) \Big\}$$

и ОММП-оценка дается равенством

$$\widehat{u}_{\text{ОММП}} = \Big(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1} \Big)^{-1} \times$$

$$\times \Big(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} f_j + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1} Y_i \Big). \quad (9.27)$$

ОММП-оценку (9.27) можно записать в эквивалентном виде

$$\widehat{u}_{\text{ОММП}} = \Big(A^T K_{\widehat{f}}^{-1} A + K_a^{-1} \Big)^{-1} \Big(A^T K_{\widehat{f}}^{-1} \widehat{f} + K_a^{-1} \widehat{u}_a \Big), \quad (9.28)$$

где $K_{\widehat{f}} = \Big(\sum\limits_{j=1}^{r} \dfrac{1}{\sigma_{fj}^2} K_{fj}^{-1} \Big)^{-1}$ — ковариационная матрица ММП-оценки $\widehat{f}$ вектора f по реализациям $f_1, \dots, f_r$,

$$\widehat{f} = \Big(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} K_{fj}^{-1} \Big)^{-1} \sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} K_{fj}^{-1} f_j, \quad K_a = \Big(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1} \Big)^{-1}$$

— ковариационная матрица ММП-оценки $\widehat{u}_a$ искомого вектора u_T по реализации априорной выборки $Y_1, \dots, Y_N$,

$$\widehat{u}_a = \Big(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1} \Big)^{-1} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1} Y_i.$$

Эквивалентность оценок (9.27), (9.28) означает, что процедуру вычисления ОММП-оценки (9.27) можно разбить на два этапа: на первом этапе по реализациям $f_1, \dots, f_r$ находится ММП-оценка $\widehat{f}$ и ее ковариационная матрица $K_{\widehat{f}}$, а по реализациям априорной выборки (9.16) находится ММП-оценка $\widehat{u}_a$ и ее ковариационная матрица K_a; на втором этапе по формуле (9.28) окончательно находится ОММП-оценка искомого вектора u.

В частности, если $f_j \sim \mathcal{N}(f_T, \sigma^2\Omega)$, $j = 1, \ldots, r$, и $Y_i \sim \sim \mathcal{N}(u_T, \sigma_a^2 K_a)$, то оценка (9.27) принимает вид

$$\widehat{u}_{\text{ОММП}} = \Big(\frac{r}{\sigma^2}\, A^T\Omega^{-1}A + \frac{N}{\sigma_a^2}\, K_a^{-1}\Big)^{-1} \times \\ \times \Big(\frac{1}{\sigma^2}\, A^T\Omega^{-1}\sum_{j=1}^{r} f_j + \frac{1}{\sigma_a^2}\, K_a^{-1}\sum_{i=1}^{N} Y_i\Big). \quad (9.29)$$

Все вышеприведенные ОММП-оценки (9.24)–(9.29) несмещены. Заметим, что несмещенность этих оценок предполагает усреднение не только по реализациям f, но и по реализациям априорной случайной величины ξ. Если, например, усреднение производится только по реализациям f при фиксированном числе реализаций ξ, то ОММП-оценки, вообще говоря, смещены.

Ковариационные матрицы ОММП-оценок (9.24)–(9.29) даются, соответственно, формулами

$$K_{\widehat{u}} = \Big(\frac{1}{\sigma^2}\, A^T\Omega^{-1}A + \sum_{i=1}^{N}\frac{1}{\sigma_i^2}\, K_i^{-1}\Big)^{-1},$$

$$K_{\widehat{u}} = \Big(\frac{1}{\sigma^2}\, A^T\Omega^{-1}A + \frac{N}{\sigma_a^2}\, K_a^{-1}\Big)^{-1},$$

$$K_{\widehat{u}} = \Big(\frac{1}{\sigma^2}\, A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}\, K_a^{-1}\Big)^{-1},$$

$$K_{\widehat{u}} = \Big(\sum_{j=1}^{r}\frac{1}{\sigma_{fj}^2}\, A^T K_{fj}^{-1}A + \sum_{i=1}^{N}\frac{1}{\sigma_i^2}\, K_i^{-1}\Big)^{-1},$$

$$K_{\widehat{u}} = (A^T K_{\widehat{f}}^{-1}A + K_a^{-1})^{-1}, \quad K_{\widehat{u}} = \Big(\frac{r}{\sigma^2}\, A^T\Omega^{-1}A + \frac{N}{\sigma_a^2}\, K_a^{-1}\Big)^{-1}.$$

Все вышеприведенные ОММП-оценки состоятельны в среднем квадратичном. Рассмотрим, например, ОММП-оценку (9.29).

Теорема 9.1. *Пусть $n \to \infty$. Тогда* ОММП-*оценка* (9.29) *сходится в среднем квадратичном к u_T при любом числе r реализаций f.*

Доказательство. Представим ОММП-оценку (9.29) в виде

$$\widehat{u}_{\text{ОММП}} = u_T + R_1\Delta f + R_2\Delta u,$$

где

$$R_1 = \Big(\frac{r}{\sigma^2}\, A^T\Omega^{-1}A + \frac{N}{\sigma_a^2}\, K_a^{-1}\Big)^{-1}\frac{r}{\sigma^2}\, A^T\Omega^{-1},$$

$$R_2 = \left(\frac{r}{\sigma^2} A^T \Omega^{-1} A + \frac{N}{\sigma_a^2} K_a^{-1} \right)^{-1} \frac{N}{\sigma_a^2} K_a^{-1},$$

$$\Delta f = \widehat{f} - f_T, \quad \Delta u = \widehat{u} - u_T, \quad \widehat{f} = \frac{1}{r} \sum_{j=1}^{r} f_j, \quad \widehat{u} = \frac{1}{N} \sum_{i=1}^{N} Y_i.$$

Имеем

$$K_{\widehat{f}} = \frac{\sigma^2}{r} \Omega, \quad K_{\widehat{u}} = \frac{\sigma_a^2}{N} K_a, \quad \mathbf{M} \Delta f = 0, \quad \mathbf{M} \Delta u = 0.$$

Из независимости случайных векторов Δf, Δu следует

$$K_{\widehat{u}_{\text{оммп}}} = R_1 K_{\Delta f} R_1^T + R_2 K_{\Delta u} R_2^T = R_1 K_{\widehat{f}} R_1^T + R_2 K_{\widehat{u}} R_2^T.$$

Для пары симметричных матриц

$$\frac{1}{\sigma^2} A^T \Omega^{-1} A^T, \quad \frac{1}{\sigma_a^2} K_a^{-1},$$

из которых первая неотрицательна, а вторая положительно определена, существует такая невырожденная матрица U, что

$$U^T \frac{1}{\sigma^2} A^T \Omega^{-1} A U = \operatorname{diag}(\lambda_1, \ldots, \lambda_l, 0, \ldots, 0), \quad U^T \frac{1}{\sigma_a^2} K_a^{-1} U = I_m,$$

где $\lambda_1, \ldots, \lambda_l > 0$ — сингулярные числа матрицы $\frac{1}{\sigma^2} A^T \Omega^{-1} A$, l — ранг матрицы A, $l \leqslant m$.

Имеем

$$U^{-1} R_1 K_{\widehat{f}} R_1^T (U^T)^{-1} = r U^{-1} \left(\frac{r}{\sigma^2} A^T \Omega^{-1} A + \frac{N}{\sigma_a^2} K_a^{-1} \right)^{-1} \times$$

$$\times (U^T)^{-1} U^T \frac{1}{\sigma^2} A^T \Omega^{-1} A U U^{-1} \left(\frac{r}{\sigma^2} A^T \Omega^{-1} A + \frac{N}{\sigma_a^2} K_a^{-1} \right)^{-1} (U^T)^{-1} =$$

$$= r \left(r U^T \frac{1}{\sigma^2} A^T \Omega^{-1} A U + N U^T \frac{1}{\sigma_a^2} K_a^{-1} U \right)^{-1} \times$$

$$\times U^T \frac{1}{\sigma^2} A^T \Omega^{-1} A U \left(r U^T \frac{1}{\sigma^2} A^T \Omega^{-1} A U + N U^T \frac{1}{\sigma_a^2} K_a^{-1} U \right)^{-1} =$$

$$= r \left(\operatorname{diag}(N + r\lambda_1, \ldots, N + r\lambda_l, N, \ldots, N) \right)^{-1} \times$$

$$\times \operatorname{diag}(\lambda_1, \ldots, \lambda_l, 0, \ldots, 0) \times$$

$$\times \left(\operatorname{diag}(N + r\lambda_1, \ldots, N + r\lambda_l, N, \ldots, N) \right)^{-1} =$$

$$= \operatorname{diag}\left(\frac{r\lambda_1}{(N + r\lambda_1)^2}, \ldots, \frac{r\lambda_l}{(N + r\lambda_l)^2}, 0, \ldots, 0 \right).$$

Следовательно, $R_1 K_{\widehat{f}} R_1^T$ стремится к нулевой матрице при $N \to \infty$ для любого r.

Аналогично получаем равенство

$$U^{-1}R_2K_{\hat{u}}R_2^T(U^T)^{-1}=\operatorname{diag}\left(\frac{N}{(N+r\lambda_1)^2},\ldots,\frac{N}{(N+r\lambda_l)^2},\frac{1}{N},\ldots,\frac{1}{N}\right).$$

Следовательно, $R_2K_{\hat{u}}R_2^T$ стремится к нулевой матрице при $N\to\infty$ для любого r. Итак, $K_{\hat{u}_{\text{оммп}}}$ стремится к нулевой матрице при $N\to\infty$ для любого r. Теорема доказана.

Теорема 9.2. *Пусть $r\to\infty$. Тогда* ОММП-*оценка* (9.29) *сходится в среднем квадратичном к одному из решений системы $Au=f_T$. В частности, если A — матрица полного ранга, то* ОММП-*оценка сходится к u_T.*

Доказательство. Если одновременно с $r\to\infty$ и $N\to\infty$, то ОММП-оценка (9.29) сходится к u_T в силу теоремы 9.1. Поэтому достаточно рассмотреть случай ограниченного N. Не умаляя общности, можно считать N фиксированным. Рассмотрим представление $K_{\hat{u}_{\text{оммп}}}=$ $=R_1K_{\hat{f}}R_1^T+R_2K_{\hat{u}}R_2^T$ из предыдущей теоремы. Из доказательства теоремы 9.1 следует, что $R_1K_{\hat{f}}R_1^T$ стремится к нулевой матрице не только при $N\to\infty$, но и при $r\to\infty$, а $R_2K_{\hat{u}}R_2^T$ не стремится к нулевой матрице при $r\to\infty$. Последнее означает, что случайный вектор $R_2\Delta u$ не сходится к детерминированному вектору при ограниченном N и $r\to\infty$. Обозначим $\Delta u=U\Delta v$, где U введена при доказательстве теоремы 9.1. Имеем

$$U^{-1}R_2\Delta u=U^{-1}\left(\frac{r}{\sigma^2}A^T\Omega^{-1}A+\frac{N}{\sigma_a^2}K_a^{-1}\right)^{-1}\frac{N}{\sigma_a^2}K_a^{-1}U\Delta v=$$

$$=\operatorname{diag}\left(\frac{N}{N+r\lambda_1},\ldots,\frac{N}{N+r\lambda_l},1,\ldots,1\right)\Delta v.$$

Следовательно, $R_2\Delta u$ сходится в среднем квадратичном при $r\to\infty$ к случайному вектору

$$u^*=U\operatorname{diag}(0,\ldots,0,1,\ldots,1)U^{-1}\left(\frac{1}{N}\sum_{i=1}^N Y_i-u_T\right).$$

В частности, если A — матрица полного ранга ($l=m$), то $u^*=0$. Имеем

$$U^T\frac{1}{\sigma^2}A^T\Omega^{-1}Au^*=U^T\frac{1}{\sigma^2}A^T\Omega^{-1}AU\operatorname{diag}(0,\ldots,0,1,\ldots,1)U^{-1}\times$$

$$\times\left(\frac{1}{N}\sum_{i=1}^N Y_i-u_T\right)=\operatorname{diag}(\lambda_1,\ldots,\lambda_l,0,\ldots,0)\times$$

$$\times\operatorname{diag}(0,\ldots,0,1,\ldots,1)U^{-1}\left(\frac{1}{N}\sum_{i=1}^N Y_i-u_T\right)=0.$$

Поэтому $A^T\Omega^{-1}Au^* = 0$. Но из последнего равенства следует $Au^* = 0$. Действительно, представим Ω в виде $\Omega^{1/2}\Omega^{1/2}$, где $\Omega^{1/2}$ симметрична и положительно определена. Имеем

$$0 = (A^T\Omega^{-1}Au^*, u^*) = (A^T\Omega^{-1/2}\Omega^{-1/2}Au^*, u^*) =$$

$$= (\Omega^{-1/2}Au^*, \Omega^{-1/2}Au^*) = \|\Omega^{-1/2}Au^*\|^2.$$

Следовательно, $Au^* = 0$. Теорема доказана.

Теоремы 9.1, 9.2 гарантируют состоятельность ОММП-оценки при неограниченном возрастании числа реализаций либо f, либо априорной случайной величины. Неограниченной точности ОММП-оценок можно добиться не только путем неограниченного роста объема выборок, но и при $\sigma^2 \to 0$ либо при $\sigma_a^2 \to 0$.

Теорема 9.3. *Пусть* $\sigma_a^2 \to 0$. *Тогда* ОММП-*оценка* (9.29) *сходится в среднем квадратичном к* u_T.

Задача 9.1. Доказать теорему 9.3, используя схему доказательства теоремы 9.1.

Теорема 9.4. *Пусть* $\sigma^2 \to 0$. *Тогда* ОММП-*оценка* (9.29) *сходится в среднем квадратичном к одному из решений системы* $Au = f_T$. *В частности, если* A — *матрица полного ранга, то* ОММП-*оценка* (9.29) *сходится к* u_T.

Задача 9.2. Доказать теорему 9.4, используя схему доказательства теоремы 9.2.

Аналогично доказывается состоятельность других вышеприведенных ОММП-оценок.

Вернемся к ОММП-оценке (9.20) для общего случая. Обозначим $p(z;\, u) = \prod_{j=1}^{r} p_\varepsilon(x_j)p(y_1, \ldots, y_N;\, u)$ совместную плотность случайных величин $X_1 = f_1 - Au, \ldots, X_r = f_r - Au$ и случайных величин, образующих априорную выборку (9.16).

Введем обобщенные условия Фишера–Крамера–Рао:

1° $u \in R$, где R либо все пространство R^m, либо множество, для которого определена $p(z;\, u)$;

2° $\dfrac{\partial p(z;\, u)}{\partial u_k}$, $k = 1, \ldots, m$, существуют для любого $u \in R$;

3° $\displaystyle\int \left|\frac{\partial p(z;\, u)}{\partial u_k}\right| dz$, $k = 1, \ldots, m$, существуют и конечны для любого $u \in R$;

4° все элементы $(M_0)_{ij}$, $i, j = 1, \ldots, m$, обобщенной информационной матрицы существуют, конечны, а сама матрица M_0, положительно определена для любого $u \in R$.

Рассмотрим любую несмещенную оценку $\widehat{u}$, построенную на основе реализации $f_1, \ldots, f_r$ правой части регрессионной модели (7.2) и априорной выборки (9.16).

Теорема 9.5. *Пусть выполнены обобщенные условия Фишера–Крамера–Рао и интегралы* $\int \left| \widehat{u}_i(z) \dfrac{\partial p(z;\, u)}{\partial u_j} \right| dz$ *ограничены для любого* $u \in R$ *и всех* $i, j = 1, \ldots, m$. *Тогда для ковариационной матрицы* $K_{\widehat{u}}$ *оценки* $\widehat{u}$ *выполнено обобщенное неравенство Фишера–Крамера–Рао*

$$K_{\widehat{u}} \geqslant M_0^{-1}. \tag{9.30}$$

Обобщенное неравенство Фишера–Крамера–Рао определяет верхнюю границу точности любой несмещенной оценки, построенной на основе реализаций $f_1, \ldots, f_r$ регрессионной модели (7.2) и априорной выборки (9.16). Заметим, что если информационная матрица M регрессионной модели (7.2) вырождена, то верхняя граница точности для оценок, построенных только на основе реализаций $f_1, \ldots, f_r$, теряет смысл (не существует). Чтобы в этом случае была возможность получать приемлемые оценки искомого вектора u, необходимо к реализациям $f_1, \ldots, f_r$ исходной регрессионной модели (7.2) присоединить такую априорную выборку (9.16), чтобы обобщенная информационная матрица M_0 стала положительно определенной.

Теорема 9.5 остается в силе, даже если обе информационные матрицы M, M_a вырождены (последнее означает, что и в исходной регрессионной модели (7.2), и в априорной выборке (9.16) в отдельности «заложено» недостаточно информации об искомом векторе u), а их сумма, т. е. обобщенная информационная матрица M_0, положительно определена (последнее означает, что суммарной информации, «заложенной» в регрессионной модели (7.2) плюс в априорной выборке (9.16), уже достаточно для получения оценки с ограниченной ковариационной матрицей).

Из теоремы 9.5 следует, что наилучшими по точности несмещенными оценками вектора u, полученными на основе реализаций $f_1, \ldots, f_r$ регрессионной модели (7.2) и априорной выборки (9.16), являются те, для которых обобщенное неравенство Фишера–Крамера–Рао (9.30) обращается в равенство. Такие оценки назовем *обобщенными совместно эффективными.*

Все приведенные выше ОММП-оценки (9.24)–(9.29) — обобщенные совместно эффективные. Действительно, например, в условиях, когда была получена ОММП-оценка (9.27),

$$M = \sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A, \quad M_a = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1}, \tag{9.31}$$

т. е.

$$M_0 = \sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1}.$$

Но для ОММП-оценки (9.27)

$$K_{\widehat{u}} = \left(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1} \right)^{-1},$$

и поэтому $K_{\widehat{u}} = M_0^{-1}$.

Напомним, что ОММП-оценки (9.24)–(9.29) были получены, когда законы распределения $f_1, \ldots, f_r$ и членов априорной выборки (9.16) нормальные. Следовательно, в условиях, когда законы распределения реализаций $f_1, \ldots, f_r$ и членов априорной выборки (9.16) нормальные, ОММП-оценки являются наилучшими по точности оценками среди всех несмещенных оценок вектора u вне зависимости от числа реализаций r и N.

Если законы распределений $f_1, \ldots, f_r$ и членов априорной выборки (9.16) отличны от нормального, то можно доказать, что при выполнении условий, аналогичных обобщенным условиям Фишера–Крамера–Рао, ОММП-оценки являются асимптотически обобщенно совместно эффективными.

ОММП, в отличие от байесовского подхода, позволяет довольно просто получать оптимальные оценки не только вектора u, но и неизвестных характеристик законов распределения случайных величин $f_1, \ldots, f_r$ и членов априорной выборки (9.16).

Например, при решении практических задач часто неизвестны уровни шума σ^2 и σ_a^2.

Пусть для определенности выполнены условия, при которых получена ОММП-оценка (9.29), но σ^2 и σ_a^2 неизвестны. Тогда, согласно ОММП, необходимо решить экстремальную задачу (9.22), где $\alpha = (\sigma^2, \sigma_a^2)^T$. Решение экстремальной задачи (9.22) для рассматриваемого случая единственно и дается формулами

$$\widehat{u}_{\text{ОММП}} = \left(\frac{r}{\widehat{\sigma}^2} A^T \Omega^{-1} A + \frac{N}{\widehat{\sigma}_a^2} K_a^{-1} \right)^{-1} \left(\frac{r}{\widehat{\sigma}^2} A^T \Omega^{-1} \widehat{f} + \frac{N}{\widehat{\sigma}_a^2} K_a^{-1} \widehat{u}_a \right),$$

$$\widehat{\sigma}^2 = \frac{1}{rn} \sum_{j=1}^{r} \| A\widehat{u}_{\text{ОММП}} - f_i \|_{\Omega^{-1}}^2, \quad \widehat{\sigma}_a^2 = \frac{1}{Nm} \sum_{i=1}^{N} \| \widehat{u}_{\text{ОММП}} - Y_i \|_{K_a^{-1}}^2,$$

$$\widehat{f} = \frac{1}{r} \sum_{j=1}^{r} f_j, \quad \widehat{u}_a = \frac{1}{N} \sum_{i=1}^{N} Y_i.$$

ОММП-оценку для u можно записать в эквивалентном виде

$$\widehat{u}_{\text{ОММП}} = (A^T \Omega^{-1} A + \widehat{\alpha} K_a^{-1})^{-1} (A^T \Omega^{-1} \widehat{f} + \widehat{\alpha} K_a^{-1} \widehat{u}_a),$$

где

$$\widehat{\alpha} = \frac{N^2 m \sum\limits_{j=1}^{r} \| A\widehat{u}_{\text{ОММП}} - f_j \|^2_{\Omega^{-1}}}{r^2 n \sum\limits_{i=1}^{N} \| \widehat{u}_{\text{ОММП}} - Y_i \|^2_{K_a^{-1}}}. \tag{9.32}$$

Параметр $\widehat{\alpha}$ по аналогии с вариационным методом регуляризации А. Н. Тихонова можно назвать *параметром регуляризации*. Формула (9.32) определяет оптимальное значение параметра регуляризации.

Заметим, что в рамках байесовского подхода при неизвестных параметрах законов распределения эти параметры должны рассматриваться как случайные величины и необходимо задавать априорную плотность вероятности этих случайных величин. В частности, определение подходящего значения параметра регуляризации в рамках байесовского подхода невозможно, если не задавать априорное распределение возможных значений этого параметра.

ОММП-оценки, обладая всеми преимуществами Б-оценок, и, кроме того, имея вышеуказанные дополнительные преимущества, не имеют недостатков байесовского подхода. Действительно, основой байесовского подхода является трактовка искомого вектора u как случайной величины. Конечно, любая оценка искомого вектора u, полученная при применении любого статистического подхода, является случайной величиной, но элемент случайности должен относиться именно к оценкам искомого вектора, а не к самим векторам, являющимся детерминированными величинами. Тот факт, что оценки детерминированных величин оказываются величинами случайными, просто отражает ту объективную реальность, что мы не располагаем достоверной информацией об искомом векторе, а решаем задачу восстановления в условиях наличия случайных погрешностей в задаваемых входных величинах. При применении же байесовского подхода мы всегда относим элемент случайности не только к оценкам искомого детерминированного вектора u, но и к самому вектору u, толкуя его как случайную величину. Искусственно навязанная байесовским подходом трактовка искомых детерминированных величин как величин случайных и является причиной критики в адрес байесовского подхода. ОММП, в отличие от байесовского подхода, позволяет учитывать априорную информацию статистического характера об искомом векторе u, не требуя его трактовки как случайной величины.

Заметим, что в рамках ОММП возможен учет априорной информации экспертного характера об искомом векторе, когда на основе опыта исследователя (опыта решения предыдущих аналогичных задач или иных соображений экспертного характера) задается возможное случайное априорное экспертное расположение искомого вектора. В рамках байесовского подхода экспертное расположение искомого вектора задается априорной плотностью вероятности самого искомого вектора. В рамках же ОММП экспертное расположение искомого вектора за-

дается выборкой значений априорной случайной величины, в которую искомый детерминированный вектор входит как детерминированный параметр, значение которого неизвестно. Возможность учета априорной экспертной вероятностной оценки искомого вектора в рамках ОММП без трактовки искомого детерминированного вектора как вектора случайного позволяет более компактно, ясно и полностью, по сравнению с байесовским подходом, учесть априорные оценки различных экспертов, трактуя априорную оценку каждого эксперта как один член априорной выборки (9.16).

Вернемся к ОММП-оценке (9.25). При решении практических задач матрицы Ω и K_a часто неизвестны. Тогда вместо оптимальной оценки (9.25) естественно рассмотреть оценку

$$\widehat{u} = \Big(\frac{r}{\sigma^2}\, A^T A + \frac{N}{\sigma_a^2}\, B\Big)^{-1}\Big(\frac{r}{\sigma^2}\, A^T \widehat{f} + \frac{N}{\sigma_a^2}\, B\widehat{u}_a\Big), \qquad (9.33)$$

где B — симметричная положительно определенная матрица, вообще говоря отличная от K_a^{-1}.

Оценка (9.33) несмещена. Ясно, что оценка (9.33) не является оптимальной. Сохраняется ли тем не менее состоятельность оценки (9.33)?

Ранее уже отмечалась состоятельность вышеприведенных ОММП-оценок при четырех вариантах предельных переходов: $r \to \infty$; $N \to \infty$; $\sigma^2 \to 0$; $\sigma_a^2 \to 0$.

Оказывается, что при любом выборе положительно определенной матрицы B состоятельность оценки (9.33) сохраняется для любого из четырех вариантов предельного перехода.

Теорема 9.6. *Оценка* (9.33) *сходится в среднем квадратичном*: *к* u_T *либо при* $N \to \infty$, *либо при* $\sigma_a^2 \to 0$; *к одному из решений системы* $Au = f_T$ *либо при* $r \to \infty$, *либо при* $\sigma^2 \to 0$. *В частности, если* A — *матрица полного ранга, то оценка* (9.33) *сходится в среднем квадратичном к* u_T *и при* $r \to \infty$, *и при* $\sigma^2 \to 0$.

Задача 9.3. Используя схемы доказательств теорем 9.1, 9.2, докажите теорему 9.6.

9.4. Обобщенный метод максимального правдоподобия учета априорной детерминированной информации в рамках линейных моделей

Рассмотрим теперь ОММП-оценки, использующие априорную информацию детерминированного характера.

Пусть априорная информация задана в виде $u \in R_a$, где R_a — заданное априорное множество пространства R^m.

Согласно ОММП за оценку искомого вектора u берется вектор $\widehat{u}_{\text{ОММП}}$, максимизирующий функцию правдоподобия $L(u)$ при ограни-

чении $u \in R_a$, т. е.

$$\widehat{u}_{\text{ОММП}} = \arg \max_{u \in R_a} L(u). \tag{9.34}$$

Если $L(u)$ — сильно выпуклая функция, R_a — выпуклое множество, то ОММП-оценка (9.34) существует и единственна. Численно ОММП-оценки для общего случая можно находить методами нелинейного программирования, применяя, в частности, градиентные методы или методы штрафных функций.

Рассмотрим важный для практики частный случай, когда можно в явном аналитическом виде найти ОММП-оценку (9.34).

Пусть закон распределений реализаций $f_1, \ldots, f_r$ правой части регрессионной модели (7.2) нормальный — $f_j \sim \mathcal{N}(f_T, \sigma_{fj}^2 K_{fj})$, а сами реализации независимы. Тогда ОММП-оценка (9.34) является решением экстремальной задачи

$$\widehat{u}_{\text{ОММП}} = \arg \min_{u \in R_a} \sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} (f_j - Au)^T K_{fj}^{-1} (f_j - Au). \tag{9.35}$$

Пусть R_a — эллипсоид (9.6). Тогда ОММП-оценка (9.35) существует, единственна и дается равенством

$$\widehat{u}_{\text{ОММП}} = \Big(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A + \alpha^* B \Big)^{-1} \Big(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} f_j + \alpha^* B u_0 \Big), \tag{9.36}$$

где множитель Лагранжа $\alpha^* > 0$ находится из равенства

$$\|\widehat{u}_{\text{ОММП}} - u_0\|_B^2 = \rho^2. \tag{9.37}$$

Если все реализации $f_1, \ldots, f_r$ независимы и одинаково распределены по закону $\mathcal{N}(f_T, \sigma^2 \Omega)$, то ОММП-оценка (9.36) принимает вид

$$\widehat{u}_{\text{ОММП}} = \Big(\frac{r}{\sigma^2} A^T \Omega^{-1} A + \alpha^* B \Big)^{-1} \Big(\frac{r}{\sigma^2} A^T \Omega^{-1} \widehat{f} + \alpha^* B u_0 \Big), \tag{9.38}$$

где $\widehat{f} = \dfrac{1}{r} \sum\limits_{j=1}^{r} f_j$.

В частности, если имеется только одна реализация f, то

$$\widehat{u}_{\text{ОММП}} = (A^T \Omega^{-1} A + \alpha^* B)^{-1} (A^T \Omega^{-1} f + \alpha^* B u_0), \tag{9.39}$$

где $\alpha^* > 0$ находится из равенства (9.37).

ОММП-оценки (9.36)–(9.39), вообще говоря, смещены, т. е. $\mathbf{M}\, \widehat{u}_{\text{ОММП}} = u_T - \Delta u$, где смещение Δu для оценок (9.36)–(9.39) имеет, соответственно, величину

$$\Delta u = \alpha^* \Big(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A + \alpha^* B \Big)^{-1} B (u_0 - u_T),$$

$$\Delta u = \alpha^* \left(\frac{r}{\sigma^2} A^T \Omega^{-1} A + \alpha^* B \right)^{-1} B (u_0 - u_T),$$

$$\Delta u = \alpha^* \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \alpha^* B \right)^{-1} B (u_0 - u_T).$$

Рассмотрим, например, семейство оценок (9.39) с $\alpha^* = \alpha > 0$. При $\alpha = 0$

$$\widehat{u} = \widehat{u}_{\text{ММП}} = (A^T \Omega^{-1} A)^{-1} A^T \Omega^{-1} f.$$

При $\alpha \to \infty$

$$\widehat{u} \to u_0.$$

Если $\widehat{u}_{\text{ММП}} \notin R_a$, то ОММП-оценка (9.39) вычисляется по формуле (9.39) при условии $\|\widehat{u}_{\text{ОММП}} - u_0\|_B^2 = \rho^2$. Если $\widehat{u}_{\text{ММП}} \in R_a$, то ОММП-оценка (9.39) совпадает с $\widehat{u}_{\text{ММП}}$ (в (9.39) $\alpha^* = 0$).

Поэтому для вычисления ОММП-оценки (9.39) целесообразно применить следующий алгоритм.

Если при малых значениях $\alpha > 0$

$$(A^T \Omega^{-1} A + \alpha B)^{-1} (A^T \Omega^{-1} f + \alpha B u_0) \notin R_a,$$

то находим такое значение α^*, при котором

$$(A^T \Omega^{-1} A + \alpha^* B)^{-1} (A^T \Omega^{-1} f + \alpha^* B u_0) \in \overline{R}_a \setminus R_a,$$

и полагаем

$$\widehat{u}_{\text{ОММП}} = (A^T \Omega^{-1} A + \alpha^* B)^{-1} (A^T \Omega^{-1} f + \alpha^* B u_0).$$

Если при $\alpha = 0$ $\widehat{u}_{\text{ММП}}$ существует и принадлежит R_a, то полагаем $\widehat{u}_{\text{ОММП}} = \widehat{u}_{\text{ММП}}$.

Заметим, что для случая $\widehat{u}_{\text{ММП}} \in R_a$ имеем $\alpha^* = 0$ и поэтому $\Delta u = 0$, т. е. смещение отсутствует. Итак, если $\widehat{u}_{\text{ММП}} \in R_a$, то дополнительная априорная информация $u \in R_a$ «не срабатывает», если же $\widehat{u}_{\text{ММП}}$ не существует или не удовлетворяет по точности (матрица $A^T \Omega^{-1} A$ вырождена или плохо обусловлена) или $\widehat{u}_{\text{ММП}} \notin R_a$, то $\widehat{u}_{\text{ОММП}}$ «подтягивает» ММП-оценку до границы априорного множества R_a.

ОММП-оценки (9.36)–(9.39) обладают свойством состоятельности. Например, оценка (9.38) сходится в среднем квадратичном к одному из решений системы $Au = f_T$ либо при $\sigma^2 \to 0$, либо при $r \to \infty$. В частности, если A — матрица полного ранга, то оценка (9.38) сходится к u_T. Если $\rho \to 0$, то ОММП-оценки (9.36)–(9.39) сходятся в среднем квадратичном к u_T. Если $\rho \to \infty$, то ОММП-оценка (9.39) сходится к ММП-оценке (7.22) (вернее, существует такое ρ_0, что $\widehat{u}_{\text{ОММП}} = \widehat{u}_{\text{ММП}}$, если $\rho > \rho_0$).

Задача 9.4. Докажите перечисленные выше варианты сходимости ОММП-оценки (9.38).

Задача 9.5. Сходятся ли ОММП-оценки (9.38) при $\rho \to \infty$, если A — матрица неполного ранга?

Заметим, что при $u_0 = 0$ и $\Omega = I_m$ ОММП-оценка (9.39) формально совпадает с приближенным решением, полученным с помощью метода регуляризации А. Н. Тихонова или других методов решения некорректно поставленных задач [35, 46, 49, 51–54, 58, 59, 71–74, 109].

Обобщенный метод максимального правдоподобия представлен в публикациях [9, 59, 91].

9.5. Регуляризованный метод наименьших квадратов

До сих пор предполагалось, что в регрессионной модели (7.2) матрица A детерминированная, т. е. задана без случайных погрешностей. В настоящем параграфе рассматривается регрессионная модель (7.2) в условиях наличия погрешностей при задании элементов матрицы A. При решении практических задач такая ситуация возникает часто. Например, если регрессионная модель (7.2) порождена задачей восстановления функциональной зависимости, то наличие случайных погрешностей в задании элементов матрицы A эквивалентно наличию случайных погрешностей при фиксации x_i значений независимой переменной x.

Введем расширенную $(n \times (m+1))$-матрицу $B = (A_T, -f_T)$ и векторы $x^T = (u^T, 1)$ размерности $m+1$. В пространстве R^{m+1} векторов $v \in R^{m+1}$, $v^T = (u^T, y)$, y — число, множество векторов x образует гиперплоскость Π_m размерности m.

Систему (7.1) запишем в эквивалентном виде

$$(b_i, x) = 0, \qquad i = 1, \ldots, n, \tag{9.40}$$

где b_i^T — i-я строка матрицы B.

При фиксированном $x \in \Pi_m$ множество векторов $S_m = \{v\colon v \in \in R^{m+1},\ (v, x) = 0\}$ образует подпространство размерности m. Для произвольного $v \in R^{m+1}$ имеет место ортогональное разложение $v = = v_S + v_N$, где $v_S \in S_m$, $v_N \perp S_m$.

Имеем

$$v_N = \frac{(v,x)x}{\|x\|^2}, \quad x \neq 0, \quad \rho^2(v, S_m) = \|v_N\|^2 = \frac{(v,x)^2}{\|x\|^2},$$

где $\rho(v, S_m)$ — расстояние между v и S_m.

Обозначим $J_T(u)$ сумму квадратов расстояний между b_i и S_m, $i = 1, \ldots, n$.

Имеем

$$J_T(u) = \sum_{i=1}^{n} \frac{(b_i, x)^2}{\|x\|^2},$$

где $|(b_i, x)|$ — расстояние между b_i и S_m вдоль прямой, параллельной оси Oy (на рис. 9.1 показан случай для $m = 2$).

Используя равенство $\sum_{i=1}^{n} (b_i, x)^2 = \|Au - f_T\|^2$, получаем

$$J_T(u) = \frac{\|Au - f_T\|^2}{\|u\|^2 + 1}.$$

Пусть теперь вместо точной правой части f_T и точной матрицы A_T известны их случайные реализации f и A.

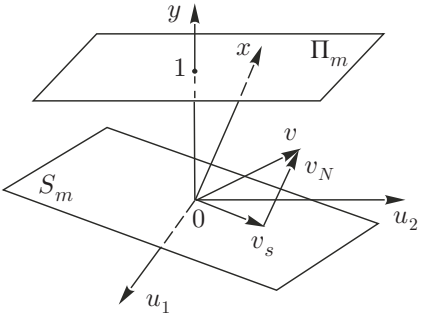

Рис. 9.1

Рассмотрим функционал

$$J(u) = \frac{\|Au - f\|^2}{\|u\|^2 + 1}. \tag{9.41}$$

Напомним, что согласно МНК в рамках классической схемы требуется найти вектор u, минимизирующий функционал невязки $\|Au - f\|^2$.

Из вышесказанного следует, что нахождение МНК-оценки эквивалентно нахождению вектора $x \in \Pi_m$, который определит такое подпространство S_m $(x \perp S_m)$, сумма квадратов расстояний до которого вдоль оси Oy от точек b_i, $i = 1, \ldots, n$, будет минимальной.

Введем *метод ортогональных проекций* (МОП) [41], согласно которому требуется найти вектор u, минимизирующий функционал ортогональных проекций (9.41).

Из вышесказанного следует, что нахождение u согласно МОП эквивалентно нахождению вектора $x \in \Pi_m$, который определит такое подпространство S_m $(x \perp S_m)$, сумма квадратов расстояний до которого от точек b_i, $i = 1, \ldots, n$ (сумма квадратов длин перпендикуляров, опущенных из точек b_i на S_m), будет минимальной.

Заметим, что подпространства S_m, определяемые МНК и МОП, в общем случае различны (и тем самым будут различны МНК-оценка и МОП-оценка вектора u). Эти подпространства будут совпадать, когда система $Au = f$ непротиворечива, т. е. существует ее точное решение.

В этом частном случае все точки b_i принадлежат общему подпространству S_m, определяемому и МНК, и МОП.

Найдем вектор u, минимизирующий функционал ортогональных проекций (9.41). Подсчет градиента функционала (9.41) дает формулу

$$\operatorname{grad}_u J(u) = \frac{2}{\|u\|^2+1}\Big(A^TAu - A^Tf - \frac{\|Au-f\|^2}{\|u\|^2+1}\,u\Big).$$

Следовательно, МОП-оценка $\widehat{u}_{\text{МОП}}$ является решением нелинейного уравнения

$$A^TAu - \alpha^* u = A^Tf, \tag{9.42}$$

где $\alpha^* = \min_u J(u) = J(\widehat{u}_{\text{МОП}})$.

Нелинейность уравнения (9.42) обусловлена зависимостью параметра $\alpha^* > 0$ от искомой МОП-оценки.

Можно оценить снизу значение α^*. Действительно,

$$\alpha^* = \min_{x\in\Pi_m} J(u) =$$

$$= \min_{x\in\Pi_m} \sum_{i=1}^n \frac{(b_i,x)^2}{\|x\|^2} \geqslant \min_{v\in R^{m+1}} \sum_{i=1}^n \frac{(b_i,v)^2}{\|v\|^2} = \min_{\|v\|=1} \sum_{i=1}^n (b_i,v)^2.$$

Уравнение Эйлера для последней экстремальной задачи с ограничением имеет вид

$$B^TBv = \lambda v, \tag{9.43}$$

где λ — множитель Лагранжа.

Уравнение (9.43) — задача на собственные значения симметричной неотрицательной $(m+1)\times(m+1)$-матрицы B^TB. Пусть $\lambda^* \geqslant 0$ — наименьшее собственное значение, а v^* ($\|v^*\| = 1$) — соответствующий ему собственный вектор матрицы B^TB. Поскольку

$$\sum_{i=1}^n (b_i,v)^2 = (B^TBv, v),$$

то

$$\min_{\|v\|=1} \sum_{i=1}^n (b_i,v)^2 = \min_{\|v\|=1} (B^TBv, v) = \lambda^*.$$

Следовательно, доказано неравенство $\alpha^* \geqslant \lambda^*$.

Если система $Au = f$ совместна, то все $b_i \in S_m$ (оба подпространства S_m, определяемые МНК и МОП, совпадают). Следовательно, $\alpha^* = \lambda^* = 0$, уравнения, определяющие МНК-оценку и МОП-оценку, совпадают.

Предположим теперь, что элементы a_{ij} матрицы A и элементы f_i вектора f — независимые нормально распределенные случайные величины $a_{ij} \sim \mathcal{N}((A_T)_{ij}, \sigma_u^2)$, $f_i \sim \mathcal{N}(f_{Ti}, \sigma_f^2)$, $i = 1,\ldots,n$, $j = 1,\ldots,m$.

Обозначим $f_{\text{Н}} = \frac{\sigma_a}{\sigma_f} f$, $u_{\text{Н}} = \frac{\sigma_a}{\sigma_f} u$, $f_{T\text{Н}} = \frac{\sigma_a}{\sigma_f} f_T$, $B_{\text{Н}} = (A, -f_{\text{Н}})$, $B_{T\text{Н}} = (A_T, -f_{T\text{Н}})$.

Все элементы $b_{\text{Н}ij}$ нормализованной расширенной $n \times (m+1)$-матрицы $B_{\text{Н}}$ — независимые нормально распределенные случайные величины с одинаковой дисперсией σ_a^2. Функция правдоподобия относительно всей совокупности случайных величин $b_{\text{Н}ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, имеет вид

$$L(B_{T\text{Н}}) = (2\pi\sigma_a^2)^{-n(m+1)/2} \exp\Big\{ -\frac{1}{2\sigma_a^2} \sum_{i=1}^{n} \sum_{j=1}^{m+1} (b_{\text{Н}ij} - b_{T\text{Н}ij})^2 \Big\},$$

где $b_{T\text{Н}ij}$ — неизвестные элементы нормализованной расширенной матрицы $B_{T\text{Н}}$.

Согласно ММП для нахождения ММП-оценок элементов $b_{T\text{Н}ij}$ необходимо максимизировать $L(B_{T\text{Н}})$, что эквивалентно минимизации функционала

$$J(B_{T\text{Н}}) = \sum_{i=1}^{n} \sum_{j=1}^{m+1} (b_{\text{Н}ij} - b_{T\text{Н}ij})^2 \equiv \sum_{i=1}^{n} \|b_{\text{Н}i} - b_{T\text{Н}i}\|^2,$$

где $b_{\text{Н}i}^T$, $b_{T\text{Н}i}^T$ — i-е строки матриц $B_{\text{Н}}$ и $B_{T\text{Н}}$ соответственно.

Согласно вышесказанному, векторы $b_{T\text{Н}i}$ принадлежат подпространству $S_{\text{Н}m}$, ортогональному искомому нормализованному вектору $x_{\text{Н}}^T = (u_{\text{Н}}^T, 1)$. Следовательно, функционал $J(B_{T\text{Н}})$ является функционалом ортогональных проекций для нормализованной системы $Au_{\text{Н}} = f_{\text{Н}}$. Итак, для нормализованной системы ММП эквивалентен МОП при выполнении вышеуказанных условий относительно случайных погрешностей. Уравнение (9.42) для нормализованной системы относительно ММП-оценки $\widehat{u}_{\text{Н}}$ искомого нормализованного вектора $u_{\text{Н}}$ имеет вид

$$A^T A\widehat{u}_{\text{Н}} - \alpha^* \widehat{u}_{\text{Н}} = A^T f_{\text{Н}}, \quad \text{где } \alpha^* = \min_{\widehat{u}_{\text{Н}} \in E^m} \frac{\|Au_{\text{Н}} - f_{\text{Н}}\|^2}{\|\widehat{u}_{\text{Н}}\|^2 + 1}.$$

Возвращаясь к исходным векторам, получаем

$$A^T A\widehat{u}_{\text{МОП}} - \alpha^* \widehat{u}_{\text{МОП}} = A^T f, \quad \text{где } \alpha^* = \sigma_a^2 \min_{u \in R^m} \frac{\|Au - f\|^2}{\sigma_a^2 \|u\|^2 + \sigma_f^2}. \quad (9.44)$$

Из последнего равенства, в частности, следует, что $\alpha^* \to 0$ при $\sigma_a^2 \to 0$, т. е. МОП-оценка стремится к МНК-оценке при неограниченном уменьшении уровня погрешностей в элементах матрицы A.

МОП позволяет найти не только оценку искомого вектора u, но и переоценить неизвестные значения матрицы A_T и вектора f_T. Действительно, МОП-оценки $\widehat{b}_{T\text{Н}i}$ неизвестных векторов $b_{T\text{Н}i}$ принадлежат одному подпространству $\widehat{S}_{\text{Н}m}$, ортогональному вектору $\widehat{x}_{\text{Н}}^T = (\widehat{u}_{\text{Н}}^T, 1)$, а сами $\widehat{b}_{T\text{Н}i}$ равны проекциям векторов $b_{\text{Н}i}$ на подпространство $\widehat{S}_{\text{Н}m}$.

Следовательно, $\widehat{b}_{T\mathrm{H}i} = b_{\mathrm{H}i} - (b_{\mathrm{H}i}, \widehat{x}_{\mathrm{H}})\widehat{x}_{\mathrm{H}}/\|\widehat{x}_{\mathrm{H}}\|^2$, $i = 1, \ldots, n$. Последние равенства запишем в эквивалентном виде

$$\widehat{a}_{Ti} = a_i - \frac{(a_i, \widehat{u}_{\mathrm{H}}) - f_{\mathrm{H}i}}{\|\widehat{u}_{\mathrm{H}}\|^2 + 1}\widehat{u}_{\mathrm{H}}, \quad \widehat{f}_{\mathrm{TH}i} = f_{\mathrm{H}i} + \frac{(a_i, \widehat{u}_{\mathrm{H}}) - f_{\mathrm{H}i}}{\|\widehat{u}_{\mathrm{H}}\|^2 + 1}, \quad i = 1, \ldots, n, \tag{9.45}$$

где a_i^T — i-я строка матрицы A, $f_{\mathrm{H}i}$ — i-й элемент нормализованного вектора f_{H}.

Векторно-матричная форма системы равенств (9.45) относительно исходных векторов u, f имеет вид

$$\widehat{A}_{\mathrm{МОП}} = A + \sigma_a^2 \frac{\rho(\widehat{u}_{\mathrm{МОП}})\widehat{u}_{\mathrm{МОП}}^T}{\sigma_a^2\|\widehat{u}_{\mathrm{МОП}}\|^2 + \sigma_f^2}, \quad \widehat{f}_{\mathrm{МОП}} = f - \sigma_f^2 \frac{\rho(\widehat{u}_{\mathrm{МОП}})}{\sigma_a^2\|\widehat{u}_{\mathrm{МОП}}\|^2 + \sigma_f^2}, \tag{9.46}$$

где $\rho(\widehat{u}_{\mathrm{МОП}}) = f - A\widehat{u}_{\mathrm{МОП}}$.

Заметим, что МОП-оценки A_T и f_T сходятся в среднем квадратичном к A_T и $\widehat{f}_{\mathrm{МНК}} = A_{\mathrm{T}}\widehat{u}_{\mathrm{МНК}}$ соответственно при $\sigma_a^2 \to 0$.

Если σ_a^2 неизвестна, то, используя ММП, можно получить оценку $\widehat{\sigma}_a^2$. Действительно,

$$\frac{\partial \ln L(B_{T\mathrm{H}})}{\partial \sigma_a^2} = -\frac{n(m+1)}{2\widehat{\sigma}_a^2} + \frac{1}{2\widehat{\sigma}_a^4}\sum_{i=1}^{n}\sum_{j=1}^{m+1}(b_{\mathrm{H}ij} - \widehat{b}_{T\mathrm{H}ij})^2 = 0.$$

Отсюда $\|\widehat{u}_{\mathrm{МОП}}\|^2\widehat{\sigma}_a^2 = (n(m+1))^{-1}\|A\widehat{u}_{\mathrm{МОП}} - f\|^2 - \sigma_f^2$.

Последнее равенство можно записать в виде

$$\|A\widehat{u}_{\mathrm{МОП}} - f\|^2 - \mu^2\|\widehat{u}_{\mathrm{МОП}}\|^2 = \delta^2, \tag{9.47}$$

где $\mu^2 = \widehat{\sigma}_a^2\, n(m+1)$, $\delta^2 = \sigma_f^2\, n(m+1)$.

Пусть теперь задана априорная информация об искомом векторе u вида $\|u\| \leqslant K$, где K — положительная константа.

Введем *регуляризованный метод ортогональных проекций* (РМОП), согласно которому за оценку искомого вектора u принимается вектор

$$\widehat{u}_\gamma = \arg\min_{\|u\| \leqslant K} \frac{\|Au - f\|^2}{\sigma_a^2\|u\|^2 + \sigma_f^2}. \tag{9.48}$$

Вектор $\widehat{u}_\gamma$ является решением нелинейного уравнения Эйлера

$$A^T A\widehat{u}_\gamma + \gamma\widehat{u}_\gamma = A^T f, \tag{9.49}$$

где

$$\gamma = \frac{\alpha(\sigma_f^2 + \sigma_a^2\|\widehat{u}_\gamma\|^2)}{\sigma_f^2} - \frac{\sigma_a^2\|A\widehat{u}_\gamma - f\|^2}{\sigma_a^2\|\widehat{u}_\gamma\|^2 + \sigma_f^2},$$

$\alpha > 0$ — множитель Лагранжа.

Из (9.49) следует, что при $\sigma_a^2 \to 0$ параметр регуляризации может принимать как положительные, так и отрицательные значения в зависимости от соотношения между σ_a^2 и σ_f^2. Значение параметра γ определяется согласно равенству (9.47), записанному относительно $\widehat{u}_\gamma$:

$$\| A\widehat{u}_\gamma - f\|^2 - \mu^2 \|\widehat{u}_\gamma\|^2 = \delta^2. \tag{9.50}$$

Формулы (9.46), (9.49), (9.50), определяющие РМОП-оценки искомых A_T, f_T, u_T, полностью соответствуют аналогичным формулам регуляризованного метода наименьших квадратов (РМНК) А. Н. Тихонова.

Из формул (9.49), (9.50) следует, что при $\sigma_a^2 = 0$ РМОП-оценка искомого вектора u определяется системой уравнений $A^T A\widehat{u} + \alpha\widehat{u} = A^T f$, где параметр регуляризации $\alpha > 0$ определяется из равенства $\| A\widehat{u} - f\|^2 = \delta^2$. Таким образом, при отсутствии погрешностей в элементах матрицы A РМОП-оценки искомого вектора u совпадают с приближенным решением, даваемым вариационным методом регуляризации А. Н. Тихонова с выбором значения параметра регуляризации согласно способу невязки.

В заключение параграфа отметим, что МОП аналогичен методу ортогональной регрессии, используемому для восстановления функциональных зависимостей [27, 38].

Глава 10
МЕТОД НАИМЕНЬШИХ КВАДРАТОВ ДЛЯ НЕЛИНЕЙНЫХ МОДЕЛЕЙ С НЕОПРЕДЕЛЕННЫМИ ДАННЫМИ

До сих пор рассматривались задачи восстановления, в которых основное уравнение (7.2) является линейным относительно искомого вектора u. Однако для многих прикладных задач основное регрессионное уравнение может быть нелинейным относительно u.

Пусть, например, рассматривается задача восстановления функциональной зависимости, теоретическая модель которой имеет вид $y = \varphi(x;\ u)$, где функция $\varphi(x;\ u)$ нелинейно зависит от компонент u_j искомого вектора $u = (u_1, \ldots, u_m)^T$. Предполагается, что при $x = x_i$, $i = 1, \ldots, n$, производится экспериментальное измерение y_i со случайной погрешностью ε_i. Тогда справедливы равенства $y_i = \varphi(x_i;\ u) + \varepsilon_i$, которые можно записать в виде

$$A(u) = f - \varepsilon. \tag{10.1}$$

Для рассматриваемой задачи $A(u) = (\varphi(x_1;\ u), \ldots, \varphi(x_n;\ u))^T$, $f = (y_1, \ldots, y_n)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$.

Нелинейную вектор-функцию $A(u)$ можно рассматривать как нелинейный оператор, действующий из пространства R^m в пространство R^n.

К нелинейной задаче восстановления (10.1) сводятся также: многомерные регрессионные задачи, теоретические модели которых имеют вид $y = \varphi(x_1, \ldots, x_k;\ u)$, где $x_1, \ldots, x_k$ — регрессоры (независимые переменные), а функция φ нелинейно зависит от компонент искомого вектора u; дискретизированные аналоги нелинейных интегральных уравнений, включая нелинейные интегральные уравнения первого рода.

10.1. МНК-оценки параметров нелинейных моделей и их свойства

Пусть компоненты ε_i вектора погрешностей ε имеют одинаковую дисперсию σ^2 и попарно не коррелируют.

Определим МНК-оценку искомого вектора u нелинейной регрессионной модели (10.1) равенством

$$\widehat{u}_{\text{МНК}} = \arg \min_{u \in R^m} \|A(u) - f\|^2, \tag{10.2}$$

где $Q(u) \equiv \|A(u) - f\|^2 \equiv \sum_{i=1}^{n} (A_i(u) - f_i)^2$, $A_i(u)$ — компоненты вектор-функции $A(u)$.

Если ε распределен по нормальному закону $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$, то МНК-оценка (10.2) является одновременно ММП-оценкой. Действительно, в этом случае функция правдоподобия дается равенством

$$L(u) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{Q(u)}{2\sigma^2}\right],$$

и поэтому максимизация $L(u)$ эквивалентна минимизации $Q(u)$.

Встречаются прикладные задачи, в которых между компонентами вектора погрешностей ε имеются корреляционные связи и известна корреляционная матрица $K_\varepsilon = \sigma^2\Omega$ (σ^2 может быть неизвестной). Тогда МНК-оценку определим равенством

$$\widehat{u}_{\text{МНК}} = \arg\min (A(u) - f)^T \Omega^{-1} (A(u) - f). \tag{10.3}$$

Если ε распределен по нормальному закону, т. е. $\varepsilon \sim \mathcal{N}(0, \sigma^2\Omega)$, то МНК-оценка (10.3) является одновременно ММП-оценкой.

Задача 10.1. Докажите последнее утверждение.

В отличие от линейной задачи восстановления, нахождение МНК-оценки для нелинейных задач восстановления является намного более трудоемким процессом, поскольку уравнение $\operatorname{grad}_u Q(u) = 0$, определяющее МНК-оценку для нелинейных задач восстановления, является нелинейным.

Более того, для нелинейных задач восстановления функционал $Q(u)$ может иметь несколько локальных минимумов или вообще не иметь ни одного. Таким образом, для нелинейных задач восстановления существование и единственность МНК-оценки приобретает существенное и первостепенное значение как с точки зрения качественного анализа, так и с точки зрения вычисления самой МНК-оценки (если она существует).

Рассмотрим последовательность МНК-оценок $\widehat{u}_n$ нелинейной регрессионной модели (10.1) при $n \to \infty$.

Теорема 10.1. *Пусть*:

1° *компоненты вектора погрешностей ε независимы и одинаково распределены*;

2° $\|A(u_1) - A(u_0)\|^2/n$ *равномерно относительно* $u_1, u_2 \in R$ *сходится к* $\varphi(u_1, u_0)$, *где* R — *ограниченное множество*;

3° $\varphi(u_1, u_0) = 0$ *тогда и только тогда, когда* $u_1 = u_0$;

4° $A(u)$ *непрерывен по* $u \in R$.

Тогда последовательность МНК-*оценок* $\widehat{u}_n$ *сходится по вероятности к* u_T *при* $n \to \infty$.

Доказательство. Имеем

$$\frac{1}{n}\|A(u) - f\|^2 = \frac{2}{n}(A(u_T) - A(u), \varepsilon) + \frac{1}{n}\|\varepsilon\|^2 + \frac{1}{n}\|A(u) - A(u_T)\|^2.$$

Далее,

$$\mathbf{M}\,\frac{1}{n}\,(A(u_T)-A(u),\varepsilon)=0,$$

$$\sigma^2\left(\frac{1}{n}\,(A(u_T)-A(u),\varepsilon)\right)=\frac{\sigma^2}{n^2}\|A(u_T)-A(u)\|^2\to 0$$

при $n\to\infty$. Поэтому $(A(u_T)-A(u),\varepsilon)/n$ сходится в среднем квадратичном к нулю. Так как ε_i^2, $i=1,\dots,n$, распределены одинаково, то $\|\varepsilon\|^2/n$ сходится по вероятности к σ^2 при $n\to\infty$. Итак, $\|A(u)-f\|^2/n$ сходится по вероятности к $\sigma^2+\varphi(u,u_T)$ равномерно по $u\in R$.

Из определения МНК-оценки имеем

$$\frac{1}{n}\|A(\widehat{u}_n)-f\|^2\leqslant\frac{1}{n}\|A(u_T)-f\|^2\equiv\frac{1}{n}\|\varepsilon\|^2.$$

Переходя в последнем неравенстве к пределу при $n\to\infty$, получаем $\sigma^2+\varphi(u^*,u_T)\leqslant\sigma^2$, где $u^*=\lim_{n\to\infty}\widehat{u}_n$. Из последнего неравенства имеем $u^*=u_T$. Теорема доказана.

Условие 2° теоремы 10.1 является аналогом сильной регулярности матриц A_n последовательности линейных регрессионных задач (7.14). Действительно, для линейной задачи

$$\frac{1}{n}\|A(u_1)-A(u_0)\|^2=\frac{1}{n}\|A_n(u_1-u_0)\|^2=(u_1-u_0)^T\frac{1}{n}A_n^TA_n(u_1-u_0).$$

Таким образом, выполнение условия 2° эквивалентно существованию предела $A_n^TA_n/n\to L$ при $n\to\infty$, где L — положительно определенная матрица. Отметим, что неограниченность последовательности норм $\|A(u)-A(u_0)\|$ для любых u,u_0 при $n\to\infty$ является необходимым условием состоятельности последовательности МНК-оценок $\widehat{u}_k$ при условии независимости и одинаковой распределенности компонент вектора погрешностей ε.

Пусть равномерно относительно $u\in R$ существует предел

$$\lim_{n\to\infty}\frac{1}{n}\,P_n^TP_n=L,$$

где P_n — матрица первых производных, элементы которой

$$(P_n)_{ij}=\frac{\partial A_i}{\partial u_j},\quad i=1,\dots,n,\quad j=1,\dots,m,$$

и $L(u_T)$ — положительно определенная матрица. Тогда при выполнении некоторых дополнительных условий к условиям теоремы 10.1 последовательность $\widehat{u}_n$ МНК-оценок будет асимптотически нормальной, а именно: закон распределения последовательности $\sqrt{n}\,(\widehat{u}_n-u_T)$ будет стремиться к $\mathcal{N}(0,\sigma^2L^{-1}(u_T))$ при $n\to\infty$.

Следовательно, при достаточно большом n с хорошим приближением можно полагать, что $\widehat{u}_n\sim\mathcal{N}(u_T,\sigma^2(P_n^T(\widehat{u}_n)P_n(\widehat{u}_n))^{-1})$.

Последнее, в частности, позволяет после нахождения МНК-оценки $\widehat{u}_{\text{МНК}}$ строить приближенные доверительные интервалы для компонент $\widehat{u}_{Tj}$, $j = 1, \ldots, m$, искомого вектора u_T:

$$\widehat{u}_{\text{МНК}j} - t(p_{\text{дов}})\sigma((P^T(\widehat{u}_{\text{МНК}})P(\widehat{u}_{\text{МНК}}))^{-1}_{jj})^{1/2} \leqslant$$

$$\leqslant u_{Tj} \leqslant \widehat{u}_{\text{МНК}j} + t(p_{\text{дов}})\sigma((P^T(\widehat{u}_{\text{МНК}})P(\widehat{u}_{\text{МНК}}))^{-1}_{jj})^{1/2},$$

где $p_{\text{дов}}$ — доверительная вероятность, $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^{t(p_{\text{дов}})} e^{-s^2/2}\, ds = \frac{p_{\text{дов}}}{2}$.

Если σ^2 неизвестна, то в качестве состоятельной оценки можно взять статистику $s^2 = \|A(\widehat{u}_{\text{МНК}}) - f\|^2/n$.

Заметим, что МНК-оценки нелинейной регрессионной модели (10.1), как правило, смещены. Ясно, что для состоятельных МНК-оценок смещение стремится к нулю при $n \to \infty$.

10.2. Численные методы нахождения МНК-оценок параметров нелинейных моделей

Вычисление МНК-оценки (10.2) можно осуществить, используя общие методы минимизации знакоопределенных функционалов, например градиентные методы или классический метод Ньютона. Однако для решения экстремальной задачи (10.2), учитывая специальный вид минимизируемого функционала $Q(u)$, можно сконструировать специальные эффективные вычислительные алгоритмы. Одним из таких специальных вычислительных методов является *метод Ньютона–Гаусса*.

Предположим, что функции $A_i(u)$ непрерывно дифференцируемы. Пусть задано приближение $u^{(k)}$ искомого вектора u. Для нахождения следующего приближения $u^{(k+1)}$ применим следующую схему.

Представим приближенно

$$A(u) = A(u^{(k)}) + P_k(u - u^{(k)}), \tag{10.4}$$

где P_k — $(n \times m)$-матрица первых производных компонент вектор-функции $A(u)$, вычисленных для значения $u = u^{(k)}$, т. е.

$$(P_k)_{ij} = \frac{\partial A_i(u^{(k)})}{\partial u_j}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m.$$

Формула (10.4) определяет линеаризацию вектор-функции $A(u)$ в окрестности точки $u = u^{(k)}$. Заменяя в (10.1) $A(u)$ согласно (10.4), получаем вместо нелинейной задачи (10.1) линейную задачу восстановления:

$$P_k \delta u = \delta f - \varepsilon, \tag{10.5}$$

где $\delta u = u - u^{(k)}$ — искомый вектор, $\delta f = f - A(u^{(k)})$.

МНК-оценка $\delta u_{\text{МНК}}$ линейной регрессионной модели (10.5) дается равенством

$$\delta u_{\text{МНК}} = (P_k^T P_k)^{-1} P_k^T \delta f. \tag{10.6}$$

Положим $u^{(k+1)} = u^{(k)} + \delta u_{\text{МНК}}$. Из (10.6) получаем

$$u^{(k+1)} = u^{(k)} + (P_k^T P_k)^{-1} P_k^T (f - A(u^{(k)})). \tag{10.7}$$

Формула (10.7) определяет итерационный процесс (ИП) Ньютона–Гаусса. Таким образом, последующее приближение ИП Ньютона–Гаусса является МНК-оценкой искомого вектора u для линеаризованной в окрестности предыдущего приближения исходной нелинейной регрессионной модели.

Задача 10.2. Пусть ковариационная матрица компонент вектора погрешностей ε определяется равенством $K_\varepsilon = \sigma^2 \Omega$, где Ω — положительно определенная матрица. Покажите, что ИП Ньютона–Гаусса для этого случая принимает вид

$$u^{(k+1)} = u^{(k)} + (P_k^T \Omega^{-1} P_k)^{-1} P_k^T \Omega^{-1} (f - A(u^{(k)})).$$

Для сходимости ИП (10.7) к стационарной точке, обращающей в нуль $\operatorname{grad}_u Q(u)$, достаточно, чтобы матрицы $P_k^T P_k$ были хорошо обусловлены, а нелинейная зависимость $A(u)$ «не очень сильной». Конечно, сходимость последовательности $u^{(k)}$ к стационарной точке может иметь место и для случаев «сильной» нелинейности $A(u)$.

На каждом шаге ИП (10.7) необходимо обращение симметричной положительно определенной $(m \times m)$-матрицы $P_k^T P_k$. При практической реализации ИП (10.7) матрицу $P_k^T P_k$ оставляют «замороженной» для обслуживания нескольких итераций, а только затем производят ее перерасчет. Это позволяет в ряде случаев значительно сократить общее количество вычислений для получения окончательной оценки.

Еще одной упрощенной схемой ИП (10.7) является ИП градиентного метода:

$$u^{(k+1)} = u^{(k)} + P_k^T (f - A(u^{(k)})), \tag{10.8}$$

не требующий обращения матрицы $P_k^T P_k$.

Часто вместо ИП (10.7) и (10.8) используют, соответственно, ИП

$$u^{(k+1)} = u^{(k)} + \alpha_k (P_k^T P_k)^{-1} P_k^T (f - A(u^{(k)})), \tag{10.9}$$

$$u^{(k+1)} = u^{(k)} + \alpha_k P_k^T (f - A(u^{(k)})), \tag{10.10}$$

где оптимальное значение ускоряющего множителя α_k можно выбирать из условия

$$\alpha_k = \arg \min_{\alpha} Q(u^{(k)} + \alpha v^{(k)}),$$

$$v^{(k)} = (P_k^T P_k)^{-1} P_k^T (f - A(u^{(k)}))$$

для (10.9) и $v^{(k)} = P_k^T(f - A(u^{(k)}))$ для (10.10).

Для ИП (10.9) и (10.10) в качестве подходящих значений α_k можно брать соответственно

$$\alpha_k = \frac{(P_k v^{(k)}, A(u^{(k)}) - f)}{\|P_k v^{(k)}\|^2},$$

$$\alpha_k = -\frac{(u^{(k)}, P_k^T(f - A(u^{(k)})))}{\|P_k^T(f - A(u^{(k)}))\|^2}.$$

Часто основной вычислительной трудностью для реализации ИП Ньютона–Гаусса (10.7) является плохая обусловленность матрицы $P_k^T P_k$ при некоторых значениях $u^{(k)}$. Левенбергом была предложена модификация ИП (10.7), свободная от этого недостатка. ИП Левенберга определяется равенством

$$u^{(k+1)} = u^{(k)} + (P_k^T P_k + \alpha_k I_m)^{-1} P_k^T (f - A(u^{(k)})), \qquad (10.11)$$

где параметр регуляризации $\alpha_k > 0$ выбирается так, чтобы $(m \times m)$-матрица $P_k^T P_k + \alpha_k I_m$ была хорошо обусловлена.

Используется и другая регуляризованная модификация ИП (10.7):

$$u^{(k+1)} = u^{(k)} + (P_k^T P_k + \alpha_k B_k)^{-1} P_k^T (f - A(u^{(k)})), \qquad (10.12)$$

где B_k — симметричные положительно определенные $(m \times m)$-матрицы. В частности, Марквардт предложил в качестве B_k выбирать диагональные матрицы $B_k = \operatorname{diag}((P_k^T P_k)_{11}, \ldots, (P_k^T P_k)_{mm})$, где $(P_k^T P_k)_{jj}$ — диагональные элементы матрицы $P_k^T P_k$.

Еще одной существенной вычислительной трудностью реализации вышеприведенных ИП является выбор начального вектора $u^{(1)}$. Ясно, что если вектор $u^{(1)}$ выбран неудачно, на большом расстоянии от искомой экстремальной точки функционала $Q(u)$, то значительно увеличивается количество итераций, необходимых для получения приемлемого конечного результата. Общих конструктивных рекомендаций выбора $u^{(1)}$ не существует. Для некоторых частных видов нелинейного оператора $A(u)$ такие рекомендации можно дать. Пусть, например, существует такой нелинейный оператор B: $R^n \to R^l$, что $B(A(U)) = Cu$, где C — $(l \times m)$-матрица.

Тогда из равенства $A(u) = f$ следует равенство $Cu = Bf$. Если оператор B найден, то в качестве $u^{(1)}$ можно взять МНК-оценку линейной регрессионной модели $Cu = \varphi$, $u^{(1)} = (C^T C)^{-1} C^T \varphi$, где $\varphi = B(f)$.

Пример 10.1. Рассмотрим задачу восстановления функциональной зависимости вида

$$y = \prod_{k=1}^{l} a_k^{(b_k(x),u)}(x),$$

где $a_k(x) > 0$, $b_k(x)$ — вектор-функции размерности m. Имеем

$$\ln\, y = \sum_{k=1}^{l} (b_k(x), u)\, \ln\, a_k(x) = (u, e(x)),$$

где исходная регрессионная модель линейна по u. Следовательно, можно взять

$$u^{(1)} = (C^T C)^{-1} C^T \varphi,$$

где $C_{ij} = e_j(x_i)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, $e_j(x_i)$ — j-я компонента вектора $e(x_i)$, $\varphi = (\ln\, y_1, \ldots, \ln\, y_n)^T$.

Методы и алгоритмы нелинейного оценивания представлены в работах [11, 27, 56, 101].

Глава 11
МЕТОДЫ ВЫДЕЛЕНИЯ ДЕТЕРМИНИРОВАННЫХ И ХАОТИЧЕСКИХ КОМПОНЕНТ ВРЕМЕННЫХ РЯДОВ

При решении многих практических задач, связанных с математической обработкой неопределенных данных, например задач математической экономики, часто возникает необходимость сглаживания совокупности точек (t_k, y_k), $k = 1, \ldots, n$, фиксирующих функциональную зависимость $y = y(t)$.

Операция сглаживания по существу эквивалентна выделению детерминированных и хаотических компонент из экспериментальной функциональной зависимости, задаваемой совокупностью точек (t_k, y_k), $k = 1, \ldots, n$.

Одним из эффективных методов выделения детерминированной компоненты из экспериментальной функциональной зависимости является представление функциональной зависимости по базисной системе линейно независимых функций $\{\varphi_j(t)\colon j = 1, \ldots, m\}$

$$y = y(t) = \sum_{j=1}^{m} u_j \varphi_j(t) + \varepsilon(t), \tag{11.1}$$

где $\varepsilon(t)$ — хаотическая компонента, $u_1, \ldots, u_m$ — коэффициенты разложения, подлежащие оценке на основе известной совокупности точек (t_k, y_k).

В настоящей и следующей главах мы будем трактовать хаотическую компоненту как случайный шум [37, 63]. Однако существует другой подход, в рамках которого хаотическая компонента трактуется как детерминированный динамический хаос [60, 89].

11.1. Метод сглаживающих ортогональных полиномов

Часто в качестве базисной системы функций в (11.1) берут нормальную систему полиномов. Напомним, что система полиномов $\{\varphi_j(t), j = 1, \ldots, m\}$ называется *нормальной*, если в ней присутствуют все полиномы степени $0, 1, \ldots, m-1$. Не умаляя общности, можно считать, что $\varphi_j(t)$ — полином степени $j-1$.

Наиболее простой нормальной системой полиномов является система полиномов $\varphi_j(t) = t^{j-1}$. Однако если $m \geqslant 7$, то при $\varphi_j(t) = t^{j-1}$ возникают принципиальные трудности вычисления оптимальных оценок коэффициентов u_j по совокупности точек (t_k, y_k), $k = 1, \ldots, n$.

Действительно, МНК-оценка вектора $u = (u_1, \ldots, u_m)^T$ линейной модели

$$y_k = \sum_{j=1}^{m} u_j \varphi_j(t_k) + \varepsilon(t_k), \qquad k = 1, \ldots, n, \tag{11.2}$$

имеет вид (см. § 7.3)

$$\widehat{u} = (A^T \Omega^{-1} A)^{-1} A^T \Omega^{-1} f, \tag{11.3}$$

где $f = (y_1, \ldots, y_n)^T$, $A_{ij} = t_i^{j-1}$ — элементы матрицы A, Ω — ковариационная матрица случайных величин $\varepsilon(t_i)$, $i = 1, \ldots, n$.

Поэтому элементы матрицы $A^T A$ имеют вид

$$(A^T A)_{ij} = \sum_{k=1}^{n} t_k^{i+j-2}.$$

Предположим, не умаляя общности, что $t \in [0, 1]$ и $t_{k+1} - t_k = 1/n$ и $\Omega = \sigma^2 I_n$. Тогда

$$(A^T A)_{ij} = n \sum_{k=1}^{n} \frac{t_k^{i+j-2}}{n} \approx n \int_0^1 t^{i+j-2}\, dt = \frac{n}{i+j-1}, \quad i, j = 1, \ldots, m.$$

Следовательно, матрица $A^T A$ при достаточно большом n близка к известной матрице Гильберта H_m (с точностью до множителя n). Но при увеличении порядка m обусловленность матрицы Гильберта катастрофически ухудшается, что и указывает на неустойчивость МНК-оценки (11.3) параметров линейной модели (11.2).

Одним из наиболее эффективных способов избежать неустойчивости при оценке параметров линейной регрессионной модели, порожденной представлением искомой функциональной зависимости системой полиномов, является использование в качестве базисных полиномов нормальной системы полиномов, ортогональных на множестве точек $\{t_k\}$, $k = 1, \ldots, n$.

Если в модели (11.2) $\Omega = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ и система базисных функций $\{\varphi_j(t)\}$, $j = 1, \ldots, n$, ортонормирована на множестве точек $\{t_k\}$ с весами $\{1/\sigma_k^2\}$, $k = 1, \ldots, n$, то $A^T \Omega^{-1} A$ — единичная матрица и МНК-оценка (11.3) имеет вид

$$\widehat{u}_j = \sum_{k=1}^{n} \frac{1}{\sigma_k^2} \varphi_j(t_k) y_k, \qquad j = 1, \ldots, m. \tag{11.4}$$

Устойчивость оценок (11.4) очевидна.

Для любого множества точек $\{t_k\}$ и весов $\{1/\sigma_k^2\}$, $k = 1, \ldots, n$, можно построить нормальную ортонормированную систему полиномов любого порядка m.

Для построения ортонормированной системы полиномов можно использовать устойчивую схему Форсайта [68], согласно которой полино-

мы системы рассчитываются по рекуррентным формулам

$$a_1 = \sum_{k=1}^{n} \frac{t_k}{\sigma_k^2} \Big(\sum_{k=1}^{n} \frac{1}{\sigma_k^2} \Big)^{-1}, \quad c_1 = \Big(\sum_{k=1}^{n} \frac{(t_k - a_1)^2}{\sigma_k^2} \Big)^{-1/2},$$

$$\varphi_0(t) = \Big(\sum_{k=1}^{n} \frac{1}{\sigma_k^2} \Big)^{-1/2}, \quad \varphi_1(t) = (t - a_1) c_1,$$

$$a_{i+1} = \sum_{k=1}^{n} \frac{t_k \varphi_i^2(t_k)}{\sigma_k^2}, \quad b_{i+1} = \sum_{k=1}^{n} \frac{t_k \varphi_i(t_k) \varphi_{i-1}(t_k)}{\sigma_k^2}, \tag{11.5}$$

$$\widetilde{\varphi}_{i+1}(t) = (t - a_{i+1}) \varphi_i(t) - b_{i+1} \varphi_{i-1}(t),$$

$$c_{i+1} = \Big(\sum_{k=1}^{n} \frac{\widetilde{\varphi}_{i+1}^2(t_k)}{\sigma_k^2} \Big)^{-1/2}, \quad \varphi_{i+1}(t) = c_{i+1} \widetilde{\varphi}_{i+1}(t), \quad i = 1,2,\ldots,m.$$

После формирования ортонормированной нормальной системы полиномов по формулам (11.5) рассчитываем МНК-оценки параметров $\widehat{u} = (\widehat{u}_1, \ldots, \widehat{u}_m)^T$ по формулам (11.4).

Сглаженная детерминированная компонента $y_{\text{дет}}(t)$ исходного временного ряда $\{t_k, y_k\}$, $k = 1, \ldots, n$, определяется равенством

$$y_{\text{дет}}(t) = \sum_{j=0}^{m} \widehat{u}_j \varphi_j(t). \tag{11.6}$$

В частности, значения детерминированной компоненты в точках t_k даются равенствами $y_{\text{дет}}(t_k) = \sum\limits_{j=0}^{m} \widehat{u}_j \varphi_j(t_k)$, $k = 1, \ldots, n$.

Степень сглаживания зависит от числа членов разложения m: чем меньше m, тем больше степень сглаживания.

Подходящее значение m можно находить, используя различные критерии, основанные на анализе остатков (хаотическая компонента)

$$y_{\text{хаот}}(t_k) = y_k - y_{\text{дет}}(t_k), \qquad k = 1, \ldots, n. \tag{11.7}$$

Для выбора подходящего значения m можно использовать, например, показатель Херста [76, 80]. Известно, что показатель Херста для белого шума (хаотической компоненты) равен 0,5. Поэтому для выделения детерминированной компоненты подбирается такое значение m, при котором показатель Херста хаотического временного ряда (11.7) наиболее близок к 0,5. Существуют другие критерии выбора подходящего значения m.

При решении практических задач, связанных с выделением детерминированных компонент временных случайных рядов, в них часто присутствует компонента, соответствующая большим выбросам. В этом случае применение схем выделения детерминированной компоненты, основанных на представлении

$$y_k = y_{\text{дет}}(t_k) + y_{\text{хаот}}(t_k), \tag{11.8}$$

неправомерно и ведет к значительному искажению $y_{\text{дет}}(t_k)$.

При наличии больших выбросов вместо (11.8) необходимо использовать представление

$$y_k = y_{\text{дет}}(t_k) + y_{\text{выб}}(t_k) + y_{\text{хаот}}(t_k), \tag{11.9}$$

где $y_{\text{выб}}(t_k)$ — компонента временного ряда, соответствующая большим выбросам.

Для выделения из временного ряда всех трех компонент представления (11.9) необходимо применять робастные схемы.

Решение практических задач, связанных с нахождением представления (11.9), показало высокую эффективность робастной схемы (8.5) и основанного на ней итерационного процесса (8.15).

Для исследуемой здесь задачи выделения трех компонент в представлении (11.9) схема (8.5) принимает вид

$$\widehat{u}_{\text{роб}j} = \sum_{k=1}^{n} \frac{1}{\sigma_k^2}\, \varphi_j(t_k)\widetilde{y}_k, \qquad j = 1, \ldots, m, \tag{11.10}$$

где

$$\widetilde{y}_k = \begin{cases} y_k, & \text{если } k \in I_0, \\ \sum\limits_{j=1}^{m} \widehat{u}_{\text{роб}j}\varphi_j(t_k) + K\sigma_k, & \text{если } k \in I_+, \\ \sum\limits_{j=1}^{m} \widehat{u}_{\text{роб}j}\varphi_j(t_k) - K\sigma_k, & \text{если } k \in I_-. \end{cases} \tag{11.11}$$

Здесь

$$I_0 = \Big\{k:\ |y_k - \sum_{j=1}^{m} \widehat{u}_{\text{роб}j}\varphi_j(t_k)| \leqslant K\sigma_k\Big\},$$

$$I_+ = \Big\{k:\ y_k - \sum_{j=1}^{m} \widehat{u}_{\text{роб}j}\varphi_j(t_k) > K\sigma_k\Big\},$$

$$I_- = \Big\{k:\ y_k - \sum_{j=1}^{m} \widehat{u}_{\text{роб}j}\varphi_j(t_k) < -K\sigma_k\Big\},$$

K — параметр Хьюбера.

Следовательно, итерационный процесс (8.15) для численного нахождения оценок (11.10), (11.11) принимает вид

$$y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0^{(l)}, \\ \sum\limits_{j=1}^{m} \widehat{u}_j^{(l)} \varphi_j(t_k) + K\sigma_k, & \text{если } k \in I_+^{(l)}, \\ \sum\limits_{j=1}^{m} \widehat{u}_j^{(l)} \varphi_j(t_k) - K\sigma_k, & \text{если } k \in I_-^{(l)}, \end{cases} \tag{11.12}$$

$$u_j^{(l+1)} = \sum_{k=1}^{n} \frac{1}{\sigma_k^2} \varphi_j(t_k) y_k^{(l)}, \qquad j = 1, \ldots, m,$$

где

$$I_0^{(l)} = \Big\{ k : \ |y_k - \sum_{j=1}^{m} \widehat{u}_j^{(l)} \varphi_j(t_k)| \leqslant K\sigma_k \Big\},$$

$$I_+^{(l)} = \Big\{ k : \ y_k - \sum_{j=1}^{m} \widehat{u}_j^{(l)} \varphi_j(t_k) > K\sigma_k \Big\},$$

$$I_-^{(l)} = \Big\{ k : \ y_k - \sum_{j=1}^{m} \widehat{u}_j^{(l)} \varphi_j(t_k) < -K\sigma_k \Big\}.$$

В качестве начального приближения $u_j^{(0)}$, $j = 1, \ldots, m$, итерационного процесса (11.12) можно брать неробастные оценки (11.4), т. е.

$$u_j^{(0)} = \sum_{k=1}^{n} \frac{1}{\sigma_k^2} \varphi_j(t_k) y_k, \qquad j = 1, \ldots, m. \tag{11.13}$$

Предложенный итерационный процесс настолько прост и эффективен, что его можно реализовать на персональных компьютерах даже для задач с размерностью m, достигающей нескольких сотен (при восстановлении функциональных зависимостей и выделении детерминированных компонент число m редко превышает 20–30).

После нахождения робастных оценок $\widehat{u}_{\text{роб}j}$ три компоненты представления (11.9) находятся согласно равенствам

$$\begin{aligned} y_{\text{дет}}(t_k) &= \sum_{j=1}^{m} \widehat{u}_{\text{роб}j} \varphi_j(t_k), \\ y_{\text{выб}}(t_k) &= \begin{cases} 0, & \text{если } k \in I_0, \\ y_k - y_{\text{дет}}(t_k) - K\sigma_k, & \text{если } k \in I_+, \\ y_k - y_{\text{дет}}(t_k) + K\sigma_k, & \text{если } k \in I_-, \end{cases} \\ y_{\text{хаот}}(t_k) &= y_k - y_{\text{дет}}(t_k) - y_{\text{выб}}(t_k), \quad k = 1, \ldots, n. \end{aligned} \tag{11.14}$$

На рис. 11.1 представлены результаты использования сглаживающих (сплошная линия) и робастных сглаживающих (пунктирная линия) ортогональных полиномов для $m = 7, 15$. В качестве временного ряда использованы последние 50 значений индекса РТС за 2002 год.

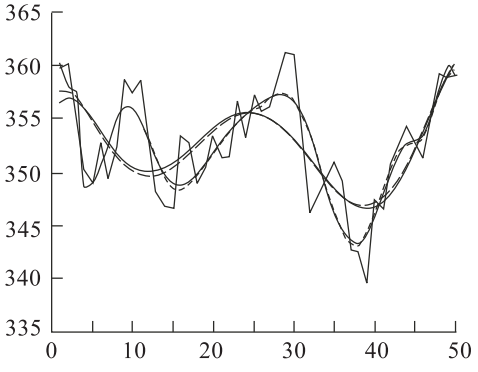

Рис. 11.1

Пример демонстрирует влияние максимальной степени полиномов m на результат сглаживания. При увеличении m сглаживающий график приближается к исходному.

Метод сглаживающих робастных полиномов можно использовать для конструирования алгоритмов, позволяющих оптимальным образом объединять экспериментальные данные при восстановлении функциональных зависимостей [21].

11.2. Метод сглаживающих линейных сплайнов

Проведение операции сглаживания, как правило, априори предполагает, что рассматривается задача восстановления функциональной зависимости $y = y(t)$ с непрерывным изменением аргумента t, т. е. восстановленная функциональная зависимость должна позволять подсчитывать значения функции $y(t)$ для всех значений аргумента t.

Однако часто при рассмотрении временных рядов только дискретная совокупность «экспериментальных» точек (t_k, y_k), $k = 1, \ldots, n$, соответствует реально существующей зависимости, а остальные значения аргумента t и соответствующие им значения y не имеют никакого смысла в рамках рассматриваемой задачи сглаживания.

При рассмотрении такого рода задач по существу речь идет о сглаживании только совокупности «экспериментальных» точек (t_k, y_k), $k = 1, \ldots, n$, и исследователя совершенно не интересуют промежуточные значения аргумента t, принадлежащие открытым интервалам $]t_k, t_{k+1}[$, $k = 1, \ldots, n-1$.

Представленные ниже сглаживающие линейные сплайны дают возможность наиболее экономно и в то же время эффективно произво-

дить операцию сглаживания дискретной совокупности точек (t_k, y_k), $k = 1, \ldots, n$, по сравнению, например, со сглаживающими кубическими сплайнами. Действительно, линейные сглаживающие сплайны имеют наиболее простую структуру для значений $t \in]t_k, t_{k+1}[$, $k = 1, \ldots, n-1$, а именно линейную зависимость от t, что и дает возможность сконструировать более простые и эффективные алгоритмы их численного восстановления, в том числе восстановления робастных линейных сглаживающих сплайнов, по сравнению, например, с кубическими сглаживающими сплайнами (см. § 11.4).

Напомним, что *интерполяционным сплайном первой степени (линейным интерполяционным сплайном)* называется кусочно линейная функция $S(t)$, проходящая через заданные точки (t_k, y_k), $k = 1, \ldots, n$, $t_1 < t_2 < \ldots < t_n$.

Введем класс функций W_1, заданных на промежутке $[t_1, t_n]$, таких, что:

1) $S(t)$ непрерывна на $[t_1, t_n]$;

2) $S(t_k) = y_k$, $k = 1, \ldots, n$, где y_k — заданные числа;

3) $S(t)$ непрерывно дифференцируема на каждом $]t_k, t_{k+1}[$, $k = 1, \ldots, n-1$;

4) в точках $t = t_k$, $k = 2, \ldots, n-1$, $S'(t)$ может иметь разрывы первого рода.

Обозначим $h_k = t_{k+1} - t_k$, $k = 1, \ldots, n-1$.

Лемма 11.1. *Среди всех функций $f(t) \in W_1$ линейный интерполяционный сплайн $S(t)$ и только он минимизирует функционал*

$$J(f) = \int_{t_1}^{t_n} (f'(t))^2 \, dt. \tag{11.15}$$

Доказательство. Пусть $f(t)$ — произвольная функция из класса W_1. Имеем: $J(S - f) = J(f) - J(S) + Q$, где

$$Q = 2 \int_{t_1}^{t_n} S'(t)(S'(t) - f'(t)) \, dt.$$

Получаем следующую цепочку равенств:

$$Q = 2 \sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} S'(t)(S'(t) - f'(t)) \, dt =$$

$$= 2 \sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} \frac{y_{k+1} - y_k}{h_k} \Big(\frac{y_{k+1} - y_k}{h_k} - f'(t) \Big) \, dt =$$

$$= 2 \sum_{k=1}^{n-1} \Big[\frac{(y_{k+1} - y_k)^2}{h_k} - \frac{y_{k+1} - y_k}{h_k} (y_{k+1} - y_k) \Big] = 0.$$

Следовательно, $J(S) = J(f) - J(S-f)$. Но $J(S-f) \geqslant 0$, поэтому линейный интерполяционный сплайн $S(t)$ доставляет минимум функционалу (11.15).

Предположим, что $f^*(t)$ — еще одна функция из класса W_1, минимизирующая функционал (11.15). Имеем: $J(f^* - S) = 0$, т. е.

$$\int_{t_k}^{t_{k+1}} (f'^*(t) - S'(t))^2\, dt = 0, \quad k = 1, \ldots, n-1.$$

Поскольку $f'^*(t)$ и $S'(t)$ непрерывны на $[t_k, t_{k+1}]$, из последнего равенства получаем: $f'^*(t) = S'(t)$ на любом промежутке $t \in [t_k, t_{k+1}]$. Следовательно, $f^*(t) = S(t) + C_k$ для $t \in [t_k, t_{k+1}]$. Но $f^*(t_k) = S(t_k)$, откуда $C_k = 0$. Лемма доказана.

Рассмотрим экстремальную задачу

$$f^*(t) = \arg\min_{f(t)} J_\alpha(f(t)) =$$

$$= \arg\min_{f(t)} \Big\{ \sum_{k=1}^{n} \frac{(y_k - f(t_k))^2}{\sigma_k^2} + \alpha \int_{t_1}^{t_n} (f'(t))^2\, dt \Big\}, \quad (11.16)$$

где $\alpha > 0$ — фиксировано.

Лемма 11.2. *Минимум функционала $J_\alpha(f(t))$ достигается на линейном интерполяционном сплайне.*

Доказательство. Зафиксировав значения $f^*(t_k)$, $k = 1, \ldots, n$, проведем через точки $(t_k, f^*(t_k))$ линейный интерполяционный сплайн $S_\alpha(t)$. В силу леммы 11.1 имеем неравенство

$$\int_{t_1}^{t_n} (S'_\alpha(t))^2\, dt \leqslant \int_{t_1}^{t_n} (f'^*(t))^2\, dt.$$

Следовательно, $J_\alpha(S_\alpha(t)) \leqslant J_\alpha(f^*(t))$, поскольку

$$\sum_{k=1}^{n} \frac{(y_k - f^*(t_k))^2}{\sigma_k^2} = \sum_{k=1}^{n} \frac{y_k - S_\alpha(t_k))^2}{\sigma_k^2}.$$

Лемма доказана.

Пусть $S_\alpha(t)$ — линейный сглаживающий сплайн, минимизирующий функционал $J_\alpha(f)$. Обозначим $S_k = S_\alpha(t_k)$, $k = 1, \ldots, n$, $\overline{S} = (S_1, \ldots, S_n)^T$. Имеем

$$J_\alpha(S_\alpha(t)) = J_\alpha(\overline{S}) = \sum_{k=1}^{n} \frac{(y_k - S_k)^2}{\sigma_k^2} + \alpha \sum_{k=1}^{n-1} \frac{1}{h_k} (S_{k+1} - S_k)^2.$$

Отсюда

$$\frac{1}{2}\frac{\partial J_\alpha}{\partial S_1} = \frac{S_1 - y_1}{\sigma_1^2} + \frac{\alpha}{h_1}(S_1 - S_2) = 0,$$

$$\frac{1}{2}\frac{\partial J_\alpha}{\partial S_k} = \frac{S_k - y_k}{\sigma_k^2} + \frac{\alpha}{h_{k-1}}(S_k - S_{k-1}) + \frac{\alpha}{h_k}(S_k - S_{k+1}) = 0,$$

$$k = 2, \ldots, n-1,$$

$$\frac{1}{2}\frac{\partial J_\alpha}{\partial S_n} = \frac{S_n - y_n}{\sigma_n^2} + \frac{\alpha}{h_{n-1}}(S_n - S_{n-1}) = 0.$$

Следовательно, для определения искомого вектора $\overline{S}$ получаем систему

$$A\overline{S} = \overline{F}, \tag{11.17}$$

где симметричная трехдиагональная $(n \times n)$-матрица A и вектор $\overline{F}$ задаются равенствами

$$A = \begin{pmatrix} \frac{1}{\sigma_1^2} + \frac{\alpha}{h_1} & -\frac{\alpha}{h_1} & 0 & 0 & \ldots & 0 & 0 & 0 \\ -\frac{\alpha}{h_1} & \frac{1}{\sigma_2^2} + \frac{\alpha}{h_1} + \frac{\alpha}{h_2} & -\frac{\alpha}{h_2} & 0 & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & -\frac{\alpha}{h_{n-1}} & \frac{1}{\sigma_n^2} + \frac{\alpha}{h_{n-1}} \end{pmatrix},$$

$$\overline{F} = \begin{pmatrix} y_1/\sigma_1^2 \\ y_2/\sigma_2^2 \\ \ldots \\ y_n/\sigma_n^2 \end{pmatrix}.$$

Матрица A — положительно определенная с диагональным преобладанием. Следовательно, существует единственное решение $\overline{S}$ системы (11.17), которое, в частности, можно найти с помощью метода прогонки [36].

Построим теперь линейный сглаживающий сплайн с закреплением крайних точек, т.е. предполагается выполнение условий $S_\alpha(t_1) = y_1$, $S_\alpha(t_n) = y_n$, где y_1 и y_n фиксированы.

Определим линейный сглаживающий сплайн как решение условной экстремальной задачи

$$\min \widetilde{J}_\alpha(S(t)), \quad S(t_1) = y_1, \quad S(t_n) = y_n, \tag{11.18}$$

где

$$\widetilde{J}_\alpha(S(t)) = \sum_{k=2}^{n-1} \frac{(y_k - S(t_k))^2}{\sigma_k^2} + \alpha \int_{t_1}^{t_n} (S'(t))^2\, dt.$$

Обозначим $\overline{S}_0 = (S_2, \ldots, S_{n-1})^T$. Имеем

$$\widetilde{J}_\alpha(\overline{S}_0) = \sum_{k=2}^{n-1} \frac{(y_k - S_k)^2}{\sigma_k^2} + \alpha \sum_{k=1}^{n-1} \frac{1}{h_k} (S_{k+1} - S_k)^2.$$

Отсюда

$$\frac{1}{2} \frac{\partial \widetilde{J}_\alpha(\overline{S}_0)}{\partial S_2} = \frac{S_2 - y_2}{\sigma_2^2} + \frac{\alpha}{h_1}(S_2 - y_1) + \frac{\alpha}{h_2}(S_2 - S_3) = 0,$$

$$\frac{1}{2} \frac{\partial \widetilde{J}_\alpha(\overline{S}_0)}{\partial S_k} = \frac{S_k - y_k}{\sigma_k^2} + \frac{\alpha}{h_{k-1}}(S_k - S_{k-1}) + \frac{\alpha}{h_k}(S_k - S_{k+1}) = 0,$$

$$k = 3, \ldots, n-2,$$

$$\frac{1}{2} \frac{\partial \widetilde{J}_\alpha(\overline{S}_0)}{\partial S_{n-1}} =$$

$$= \frac{S_{n-1} - y_{n-1}}{\sigma_{n-1}^2} + \frac{\alpha}{h_{n-2}}(S_{n-1} - S_{n-2}) + \frac{\alpha}{h_{n-1}}(S_{n-1} - y_n) = 0.$$

Следовательно, для определения вектора $\overline{S}_0$ получаем систему

$$A_0 \overline{S}_0 = \overline{F}_0, \tag{11.19}$$

где симметричная трехдиагональная $(n-2) \times (n-2)$-матрица A_0 и вектор $\overline{F}_0$ задаются равенствами

$$A_0 = \begin{pmatrix} \frac{1}{\sigma_2^2} + \frac{\alpha}{h_1} + \frac{\alpha}{h_2} & -\frac{\alpha}{h_2} & 0 & \ldots & 0 & 0 \\ -\frac{\alpha}{h_2} & \frac{1}{\sigma_3^2} + \frac{\alpha}{h_2} + \frac{\alpha}{h_3} & -\frac{\alpha}{h_3} & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \frac{-\alpha}{h_{n-2}} & \frac{1}{\sigma_{n-1}^2} + \frac{\alpha}{h_{n-2}} + \frac{\alpha}{h_{n-1}} \end{pmatrix},$$

$$\overline{F}_0 = \left(\frac{y_2}{\sigma_2^2} + \frac{\alpha}{h_1} y_1, \frac{y_3}{\sigma_3^2}, \ldots, \frac{y_{n-2}}{\sigma_{n-2}^2}, \frac{y_{n-1}}{\sigma_{n-1}^2} + \frac{\alpha}{h_{n-1}} y_n \right)^T.$$

A_0 — положительно определенная матрица с диагональным преобладанием. Следовательно, существует единственное решение системы (11.19), которое, в частности, можно найти с помощью метода прогонки.

Значение линейного сглаживающего сплайна $S_\alpha(t)$ в этом случае определяется равенствами

$$S_\alpha(t_1) = y_1, \quad S_\alpha(t_k) = S_k, \quad k = 2, \ldots, n-1, \quad S_\alpha(t_n) = y_n.$$

11.3. Робастные линейные сглаживающие сплайны

Для построения линейных робастных сглаживающих сплайнов применим схему получения робастных М-оценок.

Рассмотрим робастный функционал

$$J_\alpha(S) = \sum_{k=1}^{n} \rho\Big(\frac{y_k - S(t_k)}{\sigma_k}\Big) + \alpha \int_{t_1}^{t_n} (S'(t))^2\, dt, \tag{11.20}$$

где $\rho(t)$ — четная дифференцируемая функция, неубывающая для $t > 0$, причем $\rho(0) = 0$, $\rho(t) \approx t^2$ для малых $|t|$.

Аналогично лемме 11.2 доказывается, что функция, доставляющая минимум функционалу (11.20) в классе кусочно непрерывных функций, есть линейный сплайн $S_\alpha(t)$.

Обозначим $S_\alpha(t_k) = S_k$, $\overline{S} = (S_1, \ldots, S_n)^T$, $J_\alpha(\overline{S}) = J_\alpha(S_\alpha(t))$.

Возьмем в качестве $\rho(t)$ функцию Хьюбера

$$\rho(t) = \begin{cases} t^2, & |t| \leqslant K, \\ 2K|t| - K^2, & |t| > K. \end{cases}$$

Введем три множества индексов:

$$\begin{aligned} I_0(\overline{S}) &= \{k:\ |y_k - S_k| \leqslant K\sigma_k\}, \\ I_+(\overline{S}) &= \{k:\ y_k - S_k > K\sigma_k\}, \\ I_-(\overline{S}) &= \{k:\ y_k - S_k < -K\sigma_k\}. \end{aligned} \tag{11.21}$$

Каждый индекс $k = 1, \ldots, m$ входит в одно и только в одно из множеств $I_0(\overline{S})$, $I_-(\overline{S})$, $I_+(\overline{S})$.

Имеем

$$\frac{1}{2}\frac{\partial J_\alpha(\overline{S})}{\partial S_1} = \frac{S_1 - \widetilde{y}_1}{\sigma_1^2} + \frac{\alpha}{h_1}(S_1 - S_2) = 0,$$

$$\frac{1}{2}\frac{\partial J_\alpha(\overline{S})}{\partial S_k} = \frac{S_k - \widetilde{y}_k}{\sigma_k^2} + \frac{\alpha}{h_{k-1}}(S_k - S_{k-1}) + \frac{\alpha}{h_k}(S_k - S_{k+1}) = 0,$$

$$k = 2, \ldots, n-1, \tag{11.22}$$

$$\frac{1}{2}\frac{\partial J_\alpha(\overline{S})}{\partial S_n} = \frac{S_n - \widetilde{y}_n}{\sigma_n^2} + \frac{\alpha}{h_{n-1}}, (S_n - S_{n-1}) = 0,$$

где

$$\widetilde{y}_k = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}), \\ S_k + K\sigma_k, & \text{если } k \in I_+(\overline{S}), \\ S_k - K\sigma_k, & \text{если } k \in I_-(\overline{S}). \end{cases}$$

Следовательно, для определения искомого вектора $\overline{S}$ получаем систему

$$A\overline{S} = \widetilde{F}, \tag{11.23}$$

где матрица A определена равенством (11.17), а

$$\widetilde{F} = (\widetilde{y}_1/\sigma_1^2, \ldots, \widetilde{y}_n/\sigma_n^2)^T.$$

Система (11.23) нелинейна, поскольку компоненты вектора $\widetilde{F}$ зависят от искомого решения $\overline{S}$.

Для решения системы (11.23) используем эффективный итерационный процесс

$$A\overline{S}^{(l+1)} = \widetilde{F}^{(l)}, \tag{11.24}$$

где

$$\widetilde{F}^{(l)} = \Big(\frac{y_1^{(l)}}{\sigma_1^2}, \ldots, \frac{y_n^{(l)}}{\sigma_n^2}\Big)^T, \quad y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}^{(l)}), \\ S_k^{(l)} + K\sigma_k, & \text{если } k \in I_+(\overline{S}^{(l)}), \\ S_k^{(l)} - K\sigma_k, & \text{если } k \in I_-(\overline{S}^{(l)}), \end{cases}$$

а множества $I_0(\overline{S}^{(l)})$, $I_+(\overline{S}^{(l)})$, $I_-(\overline{S}^{(l)})$ определены равенствами (11.21).

В качестве начального $\widetilde{F}^{(0)}$ рекомендуется брать $\widetilde{F}^{(0)} = \overline{F} = (y_1/\sigma_1^2, \ldots, y_n/\sigma_n^2)^T$.

Аналогично строится итерационная процедура нахождения линейного робастного сглаживающего сплайна с фиксацией крайних точек $S_\alpha(t_1) = y_1$, $S_\alpha(t_n) = y_n$:

$$A_0\,\overline{S}_0^{(l+1)} = \overline{F}_0^{(l)}, \tag{11.25}$$

где

$$\overline{S}_0^{(l)} = (S_2^{(l)}, \ldots, S_{n-1}^{(l)})^T,$$

$$\overline{F}_0^{(l)} = \Big(\frac{y_2^{(l)}}{\sigma_2^2} + \frac{\alpha}{h_1}\,y_1, \frac{y_3^{(l)}}{\sigma_3^2}, \ldots, \frac{y_{n-2}^{(l)}}{\sigma_{n-2}^2}, \frac{y_{n-1}^{(l)}}{\sigma_{n-1}^2} + \frac{\alpha}{h_{n-1}}\,y_n\Big)^T,$$

$$y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}_0^{(l)}), \\ S_k^{(l)} + K\sigma_k, & \text{если } k \in I_+(\overline{S}_0^{(l)}), \\ S_k^{(l)} - K\sigma_k, & \text{если } k \in I_-(\overline{S}_0^{(l)}), \end{cases} \qquad k = 2, \ldots, n-1.$$

Отметим, что робастные М-оценки, полученные с помощью функции Хьюбера, предполагают симметрию больших выбросов относительно сглаженной робастной кривой. В условиях нарушения симметрии,

когда, например, количество больших выбросов мало, вместо функции Хьюбера можно использовать функцию Тьюки [81, 88]

$$\rho(t) = \begin{cases} t^2, & |t| \leqslant R, \\ R^2, & |t| > R. \end{cases} \tag{11.26}$$

Введем два множества индексов:

$$I_0(\overline{S}) = \{k : \ |y_k - S_k| \leqslant R\sigma_k\}, \quad I_R(\overline{S}) = \{k : \ |y_k - S_k| > R\sigma_k\}, \tag{11.27}$$

и функционал $J_\alpha(\overline{S})$ вида (11.20), где $\rho(t)$ определена равенством (11.26). Каждый индекс $k = 1, \ldots, n$ принадлежит одному и только одному из множеств $I_0(\overline{S})$, $I_R(\overline{S})$.

Для производных $\dfrac{\partial J_\alpha(\overline{S})}{\partial S_k}$, $k = 1, \ldots, n$, имеем систему равенств (11.22), где

$$\widetilde{y}_k = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}), \\ S_k, & \text{если } k \in I_R(\overline{S}), \end{cases} \qquad k = 2, \ldots, n-1. \tag{11.28}$$

Следовательно, для определения вектора $\overline{S}$ получаем систему (11.23), где компоненты вектора $\widetilde{F} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)^T$ даются равенствами (11.28).

Для решения системы (11.23), (11.28) используем эффективный итерационный процесс

$$A\overline{S}^{(l+1)} = \overline{F}^{(l)},$$

где

$$\overline{F}^{(l)} = \Big(\frac{y_1^{(l)}}{\sigma_1^2}, \ldots, \frac{y_n^{(l)}}{\sigma_n^2}\Big)^T, \quad y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}^{(l)}), \\ S_k, & \text{если } k \in I_R(\overline{S}^{(l)}), \end{cases}$$

а множества $I_0(\overline{S}^{(l)})$, $I_R(\overline{S}^{(l)})$ определены равенствами (11.27).

Аналогично строится итерационная процедура нахождения робастного линейного сглаживающего сплайна с нейтрализацией больших выбросов в условиях нарушения симметрии и с фиксацией крайних точек $S_\alpha(t_1) = y_1$, $S_\alpha(t_n) = y_n$:

$$A_0\overline{S}_0^{(l+1)} = \overline{F}_0^{(l)},$$

где

$$\overline{S}_0^{(l)} = (S_2^{(l)}, \ldots, S_{n-1}^{(l)})^T,$$

$$\overline{F}_0^{(l)} = \Big(\frac{y_2^{(l)}}{\sigma_2^2} + \frac{\alpha}{h_1}\, y_1, \frac{y_3^{(l)}}{\sigma_3^2}, \ldots, \frac{y_{n-2}^{(l)}}{\sigma_{n-2}^2}, \frac{y_{n-1}^{(l)}}{\sigma_{n-1}^2} + \frac{\alpha}{h_{n-1}}\, y_n\Big)^T,$$

$$y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}^{(l)}), \\ S_k^{(l)}, & \text{если } k \in I_R(\overline{S}^{(l)}), \end{cases} \qquad k = 2, \ldots, n-1,$$

а множества $I_0(\overline{S})$, $I_R(\overline{S})$ определены равенствами (11.27).

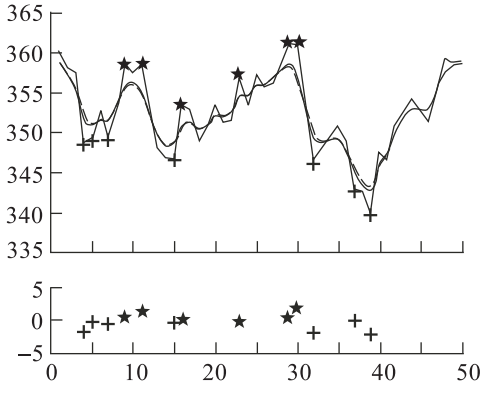

Рис. 11.2

На рис. 11.2 представлены результаты использования методов сглаживания с помощью линейного (сплошная линия) и робастного линейного (пунктирная линия) сплайнов. В качестве временного ряда использованы последние 50 значений индекса РТС за 2002 год. Символами (★, +) отмечены значения с $i \in I_+$ и $i \in I_-$ соответственно. На нижнем графике показаны величины корректировок значений исходного временного ряда в результате применения итерационного процесса (11.24).

Приведем один из примеров использования робастных линейных сглаживающих сплайнов при решении задачи оперативного управления портфелем инвестиций на рынке облигаций.

Пусть на рассматриваемую дату все выпуски облигаций пронумерованы индексом $k = 1, \ldots, n$. Обозначим через P_k, $k = 1, \ldots, n$, цену k-го выпуска (в процентах от номинальной стоимости), имеющего на рассматриваемую дату срок до погашения τ_k дней. Тогда совокупность точек (τ_k, P_k), $k = 1, \ldots, n$, описывает состояние рынка облигаций на рассматриваемую дату.

Согласно определению сглаженная совокупность точек $(\tau_k, P_{\text{сгл}k})$, $k = 1, \ldots, n$, образует текущее равновесное состояние рынка облигаций. Здесь $P_{\text{сгл}k}$ — сглаженное значение цены k-го выпуска на текущую дату.

Выпуски, для которых $P_k > P_{\text{сгл}k}$ ($P_k < P_{\text{сгл}k}$) называются *переоцененными* (*недооцененными*) рынком.

Динамическое управление портфелем выпусков облигаций состоит в перевложении средств из переоцененных выпусков облигаций в недо-

оцененные. Аналогично может решаться задача формирования портфеля акций [44, 87].

11.4. Метод сглаживающих кубических сплайнов

В настоящем параграфе рассматривается задача восстановления функциональной зависимости, когда независимая переменная не обязательно является временем.

Пусть между двумя величинами x и y существует неизвестная исследователю функциональная зависимость $y = f(x)$. Для восстановления этой зависимости производятся экспериментальные измерения y_i в точках x_i, $i = 1, \ldots, n$. Задача осложняется тем, что измерения y_i производятся с ошибками ξ_i, имеющими случайный характер, т. е. $y_i = y_{i\tau} + \xi_i$, где $y_{i\tau}$ — неизвестное точное значение $y = f(x)$ в точке $x = x_i$.

Задача восстановления функциональной зависимости состоит в получении оценки $\widehat{y} = \widehat{f}(x)$, причем не только в узлах измерения $x = x_i$, но и между ними. Более того, часто наряду с восстановлением самой функции $f(x)$ требуется найти оценку ее производной $f'(x)$.

Задача восстановления функциональной зависимости с непрерывным изменением аргумента является одной из основных задач математической обработки и интерпретации результатов экспериментов.

В настоящее время одним из эффективных способов восстановления функциональной зависимости с непрерывным изменением аргумента по экспериментальным данным является применение кубических сплайн-аппроксимаций, в том числе применение кубических сглаживающих сплайнов при наличии ошибок в измерениях y_i.

Сглаживающие кубические сплайны тесно связаны с интерполяционными кубическими сплайнами. Поэтому предварительно дадим краткое введение в теорию кубических интерполяционных сплайнов.

В дальнейшем, не умаляя общности, предполагаем, что узлы $\{x_i,\ i = 1, \ldots, n\}$, упорядочены, т. е.

$$a = x_1 < x_2 < \ldots < x_n = b. \tag{11.29}$$

Функция $S_n(x)$ называется *кубическим сплайном с узлами* (11.29), если: 1° на каждом интервале $[x_i, x_{i+1}]$ $S_n(x)$ — кубический полином, т. е.

$$S_n(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, \tag{11.30}$$

2° $S_n(x)$ — дважды непрерывно дифференцируема на $[a, b]$.

Кубический сплайн называется *интерполяционным*, если

$$S_n(x_i) = y_i, \qquad i = 1, \ldots, n. \tag{11.31}$$

Из (11.30) следует, что $y_i = a_i$, $i = 1, \ldots, n-1$.

Для построения кубического интерполяционного сплайна необходимо знать $4(n-1)$ коэффициентов $\{a_i, b_i, c_i, d_i,\ i = 1, \ldots, n-1\}$. Из условий непрерывности $S_n(x)$, $S_n'(x)$, $S_n''(x)$ во всех внутренних

узлах x_i, $i = 2, \ldots, n-1$, получаем $3(n-2)$ равенств. Кроме того, имеем n равенств интерполяции (11.31). Итого получаем $4n-6$ равенств с вхождением в них неизвестных коэффициентов. Недостающие два равенства задают дополнительно в крайних узлах x_1, x_n и называют *краевыми условиями*. Обычно используют следующие типы краевых условий:

1° $S_n'(x_1) = m_1$, $S_n'(x_n) = m_2$;
2° $S_n''(x_1) = m_1$, $S_n''(x_n) = m_2$;
3° $S_n'''(x_1) = m_1$, $S_n'''(x_2) = m_2$;
4° $S_n^{(k)}(x_1) = S_n^{(k)}(x_n)$, $k = 0, 1$.

Часто используют краевое условие 2° с $m_1 = m_2 = 0$.

Ниже приводится удобный и компактный алгоритм построения интерполяционного кубического сплайна $S_n(x)$, в котором сначала находится вектор значений вторых производных в узлах (11.29) $M_i = S_n''(x_i)$, $i = 1, \ldots, n$, а затем восстанавливаются коэффициенты b_i, c_i, d_i, $i = 1, \ldots, n-1$, по формулам

$$b_i = \frac{y_{i+1} - y_i}{h_i} - h_i \frac{M_{i+1} + 2M_i}{6}, \quad c_i = \frac{M_i}{2}, \quad d_i = \frac{M_{i+1} - M_i}{6h_i}, \quad (11.32)$$

где $h_i = x_{i+1} - x_i$, $i = 1, \ldots, n-1$.

Для нахождения неизвестных значений M_i, $i = 1, \ldots, n$, используем условия непрерывности первой производной сплайна во внутренних узлах x_i, $i = 2, \ldots, n-1$, с учетом условий интерполяции (11.31). Получаем $n-2$ равенства:

$$\frac{h_{i-1}}{6} M_{i-1} + \frac{h_{i-1} + h_i}{3} M_i + \frac{h_i}{6} M_{i+1} = \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}, \quad (11.33)$$

$$i = 2, \ldots, n-1.$$

Два недостающих уравнения получаем из краевых условий, которые относительно M_i, $i = 1, \ldots, n$, имеют вид

$$u_1 M_1 + u_2 M_2 = D_1, \quad v_1 M_{n-1} + v_2 M_n = D_2, \quad (11.34)$$

где $u_1, u_2, v_1, v_2, D_1, D_2$ в зависимости от типа краевых условий даются следующими равенствами:

1) для краевых условий типа 1° $u_1 = h_1/3$, $u_2 = h_1/6$, $v_1 = h_{n-1}/6$, $v_2 = h_{n-1}/3$, $D_1 = (y_2 - y_1)/h_1 - m_1$, $D_2 = (y_{n-1} - y_n)/h_{n-1} + m_2$;

2) для краевых условий типа 2° $u_1 = 1$, $u_2 = 0$, $v_1 = 0$, $v_2 = 1$, $D_1 = m_1$, $D_2 = m_2$;

3) для краевых условий типа 3° $u_1 = 1$, $u_2 = -1$, $v_1 = -1$, $v_2 = 1$, $D_1 = -m_1 h_1$, $D_2 = m_2 h_{n-1}$.

Совокупность n равенств (11.33), (11.34) — квадратная система n линейных алгебраических неоднородных уравнений с n неизвестными M_i, $i = 1, \ldots, n$.

Систему (11.33), (11.34) относительно искомого n-мерного вектора $M = (M_1, \ldots, M_n)^T$ запишем в векторно-матричной форме:

$$AM = Hy + f, \quad (11.35)$$

где $y = (y_1, \ldots, y_n)^T$ — n-мерный вектор, $f = (f_1, 0, \ldots, 0, f_n)^T$ имеет только две ненулевые компоненты f_1 и f_n, зависящие от типа краевых условий. Для краевых условий типа 1° $f_1 = -m_1$, $f_n = -m_2$, для остальных краевых условий $f_1 = D_1$, $f_n = D_2$. Матрица A представима в виде

$$A = \begin{pmatrix} A_1 \\ A_0 \\ A_2 \end{pmatrix},$$

где

$$A_0 = \begin{pmatrix} \frac{h_1}{6} & \frac{h_1+h_2}{3} & \frac{h_2}{6} & 0 \ldots & 0 & 0 & 0 \\ 0 & \frac{h_2}{6} & \frac{h_2+h_3}{3} & \frac{h_3}{6} \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 \ldots & \frac{h_{n-2}}{6} & \frac{h_{n-2}+h_{n-1}}{3} & \frac{h_{n-1}}{6} \end{pmatrix}$$

— $((n-2) \times n)$-матрица, $A_1 = (u_1, u_2, 0, \ldots, 0)$ — матрица-строка, $A_2 = (0,0, \ldots, 0, v_1, v_2)$ — матрица-строка.

Матрица H представима в виде

$$H = \begin{pmatrix} H_1 \\ H_0 \\ H_2 \end{pmatrix},$$

где

$$H_0 = \begin{pmatrix} \frac{1}{h_1} & -\left(\frac{1}{h_1}+\frac{1}{h_2}\right) & \frac{1}{h_2} & 0 \ldots & 0 & 0 & 0 \\ 0 & \frac{1}{h_2} & -\left(\frac{1}{h_2}+\frac{1}{h_3}\right) & \frac{1}{h_3} \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 \ldots & \frac{1}{h_{n-2}} & -\left(\frac{1}{h_{n-2}}+\frac{1}{h_{n-1}}\right) & \frac{1}{h_{n-1}} \end{pmatrix}$$

— $((n-2) \times n)$-матрица, H_1, H_2 — матрицы-строки, элементы которых зависят от типа краевых условий. Для краевых условий типа 1° $H_1 = (-\frac{1}{h_1}, \frac{1}{h_1}, 0, \ldots, 0)$, $H_2 = (0, \ldots, 0, \frac{1}{h_{n-1}}, -\frac{1}{h_{n-1}})$, для остальных типов краевых условий H_1, H_2 — нулевые матрицы.

Поскольку матрица A трехдиагональная, система (11.35) эффективно решается методом прогонки, требующим только $\sim 16n$ арифметических операций. Устойчивость метода прогонки обеспечена в силу выполнения неравенства $(h_{i-1} + h_i)/3 > h_{i-1}/6 + h_i/6$, которое означает, что A — матрица с диагональным преобладанием. В частности, диагональное преобладание обеспечивает невырожденность матрицы A и тем самым единственность решения системы (11.35). Итак, кубический интерполяционный сплайн единственен.

Для вычисления значений кубического интерполяционного сплайна $S_n(x)$ и его производной $S_n'(x)$ в точках, не совпадающих с узла-

ми (11.29), наиболее компактной является схема Горнера:

$$\begin{aligned} S_n(x) &= y_i + t(b_i + t(c_i + td_i)), \\ S_n'(x) &= b_i + t(2c_i + 3td_i), \end{aligned} \qquad x_i \leqslant x \leqslant x_{i+1}, \quad t = x - x_i. \tag{11.36}$$

Для подсчета $S_n(x)$ и $S_n'(x)$ по формулам (11.36) необходимо предварительно подсчитать $\{b_i, c_i, d_i,\ i = 1, \ldots, n-1\}$ по формулам (11.32).

Если число вычисляемых значений кубического интерполяционного сплайна меньше $1.1n$, то целесообразно вместо формул (11.36) использовать формулы

$$S_n(x) =$$

$$= y_i + t\Big\{\frac{y_{i+1} - y_i}{h_i} - (x_{i+1} - x)\frac{(x_{i+1} - x + h_i)M_i + (t + h_i)M_{i+1}}{6}\Big\},$$

$$S_n'(x) =$$

$$= \frac{y_{i+1} - y_i}{h_i} + (t + x - x_{i+1})\frac{(x_{i+1} - x + h_i)M_i + t(t + h_i)M_{i+1}}{6} +$$

$$+ t(x - x_{i+1})\frac{(2t + h_i)M_{i+1} - M_i}{6}, \tag{11.37}$$

$$x_i \leqslant x \leqslant x_{i+1}.$$

Интерполяционные кубические сплайны обладают следующим свойством сходимости. Обозначим $h_{\max} = \max_i h_i$. Пусть $f(x)$ имеет на отрезке $[a, b]$ непрерывные производные до 3-го порядка включительно. Тогда $S_n^{(k)}(x)$ равномерно по $x \in [a, b]$ стремится к $f^{(k)}(x)$, $k = 0, 1, 2$, при $h_{\max} \to 0$. Более того, имеют место неравенства

$$\sup_{x \in [a,b]} |f^{(k)}(x) - S_n^{(k)}(x)| \leqslant C_k h_{\max}^{3-k}, \quad k = 0, 1, 2, \tag{11.38}$$

где C_k не зависят от узлов (11.29), а определяются только свойствами $f(x)$ и ее первых трех производных на промежутке $[a, b]$.

Замечательным качеством интерполяционных кубических сплайнов является их экстремальное свойство. Рассмотрим квадратичный функционал

$$J(\varphi) = \int_a^b (\varphi''(x))^2\, dx. \tag{11.39}$$

Поставим задачу на минимум функционала (11.39) на классе дважды непрерывно дифференцируемых на $[a, b]$ функций, удовлетворяющих условиям интерполяции $\varphi(x_i) = y_i$, $i = 1, \ldots, m$, и одному из краевых условий 1°–4°. Обозначим класс таких интерполяционных функ-

ций через $V_{2k}(x)$, $k = 1, 2, 3, 4$, в зависимости от выполнения краевых условий K°.

Теорема 11.1. $S_n(x) = \arg\min_{\varphi(x)\in V_{2k}(x)} J(\varphi)$, *где* $S_n(x)$ — *интерполяционный кубический сплайн, удовлетворяющий краевым условиям* K°.

В частности, интерполяционный кубический сплайн $S_n(x)$, удовлетворяющий нулевым краевым условиям 2° ($m_1 = m_2 = 0$), доставляет минимум функционалу $J(\varphi)$ в классе дважды непрерывно дифференцируемых интерполяционных функций.

Функционал (11.39) определяет меру кривизны функции $\varphi(x)$ на $[a, b]$. Поэтому экстремальное свойство интерполяционных кубических сплайнов означает, что они имеют минимальное значение меры кривизны среди всех дважды непрерывно дифференцируемых интерполяционных функций.

Рассмотрим механическую задачу об изгибании тонкого упругого стержня, который должен проходить через точки y_i, $i = 1, \ldots, n$, с шарнирным закреплением в этих точках.

Тогда функционал (11.39) будет пропорционален упругой энергии стержня, имеющего форму $\varphi(x)$, а экстремальное свойство интерполяционного кубического сплайна $S_n(x)$ означает, что стержень примет форму $S_n(x)$ (предполагается, что на краях $x = a$, $x = b$ стержень закреплен в соответствии с выбранными краевыми условиями).

Интерполяционные сплайны применяют только в том случае, когда ошибки в измерении y_i малы, а расстояние между соседними узлами x_i достаточно большое. В противном случае интерполяционный сплайн, проходя через точку y_i, будет реагировать даже на сравнительно небольшие ошибки в измерении y_i и в нем будут присутствовать высокочастотные колебания, имеющие случайный характер. Присутствие высокочастотных ложных флуктуаций доставляет особые неприятности в случае необходимости дифференцирования сплайна.

Поэтому при наличии ошибок в измерениях y_i для восстановления функциональных зависимостей вместо интерполяционных сплайнов применяют сглаживающие сплайны, во многом свободные от вышеуказанных недостатков интерполяционных сплайнов.

Пусть случайные величины ξ_i (ошибки в измерении y_i), $i = 1, \ldots, n$, независимы и $\mathbf{M}\,\xi = 0$, $\mathbf{D}\,\xi = \sigma_i^2$. Существует несколько приемов введения сглаживающих кубических сплайнов. Один из приемов описан ниже.

Рассмотрим функционал

$$J_\alpha(\varphi) = \sum_{i=1}^{n} \frac{(y_i - \varphi(x_i))^2}{\sigma_i^2} + \alpha \int_a^b (\varphi''(x))^2\, dx, \qquad (11.40)$$

где $\alpha > 0$ — параметр сглаживания.

Поставим задачу на минимум функционала (11.40) в классе W_k дважды непрерывно дифференцируемых функций, удовлетворяющих условиям K°. Подчеркнем, что в отличие от экстремальной задачи на

минимум функционала (11.39) не требуется выполнения интерполяционных равенств $\varphi(x_i) = y_i$, $i = 1, \ldots, n$.

Теорема 11.2. *Решением задачи на минимум функционала* (11.40) *в классе* W_k *является кубический сплайн, удовлетворяющий краевым условиям* K°.

Доказательство. Пусть функция $\varphi^*(x) \in W_k$ доставляет минимум функционалу (11.40). Обозначим $y_i^* = \varphi^*(x_i)$, $i = 1, \ldots, n$, и построим интерполяционный кубический сплайн $S_n^*(x)$, удовлетворяющий интерполяционным условиям $S_n^*(x_i) = y_i^*$, $i = 1, \ldots, n$, и краевым условиям K°. Из теоремы 11.1 следует, что $\int\limits_a^b (S_n^*(x))^2\,dx \leqslant \int\limits_a^b (\varphi^{*\prime\prime}(x))^2 dx$. Но

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2}(y_i - \varphi^*(x_i)^a)^2 = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(y_i^a - S_n^*(x_i))^2.$$

Поэтому $J_\alpha(S_n^*) \leqslant J_\alpha(\varphi)$. Поскольку функционал (11.40) в классе функций W_k сильно выпуклый, то решение задачи на минимум функционала (11.40) единственно. Следовательно, $S_n^*(x) = \varphi^*(x)$. Теорема доказана.

В дальнейшем $S_n^*(x)$ будем называть *сглаживающим кубическим сплайном* и обозначать $S_{n\alpha}(x)$.

На первый взгляд задача нахождения сглаживающего сплайна кажется более сложной, чем для интерполяционного сплайна, поскольку априори отсутствует информация о координатах $S_{n\alpha}(x_i)$. Но эта задача решается с использованием экстремального свойства сглаживающих кубических сплайнов, зафиксированного в теореме 11.2.

Решение задачи на минимум функционала (11.40) приводит к следующему замечательному результату относительно искомого сглаживающего сплайна $S_{n\alpha}(x)$ [32].

Теорема 11.3. *Кубический сплайн* $S_{n\alpha}(x)$ *является сглаживающим тогда и только тогда, когда выполнена система равенств*

$$S_{n\alpha}(x_i) + \alpha\sigma_i^2 Z_i = y_i, \qquad i = 1, \ldots, n, \tag{11.41}$$

где $Z_1 = S_{n\alpha}'''(a)$, $Z_i = S_{n\alpha}'''(x_i + 0) - S_{n\alpha}'''(x_i - 0)$, $i = 1, 2, \ldots, n-1$, $Z_n = S_{n\alpha}'''(b)$.

Система равенств (11.41) дает возможность построить эффективный численный алгоритм нахождения сглаживающего кубического сплайна. Обозначим

$$Z = (Z_1, \ldots, Z_n)^T, \quad y = (y_1, \ldots, y_n)^T,$$

$$S_\alpha = (S_{\alpha 1}, \ldots, S_{\alpha n})^T, \quad M = (M_1, \ldots, M_n)^T,$$

где $S_{\alpha i} = S_{n\alpha}(x_i)$, $M_i = S_{n\alpha}''(x_i)$, $i = 1, \ldots, n$. Введем матрицу $K_\xi = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.

Тогда систему уравнений (11.41) можно записать в векторно-матричной форме:

$$S_\alpha + \alpha K_\xi Z = y. \tag{11.42}$$

Векторы S_α и Z можно выразить через вектор M. Действительно, для кубических сплайнов справедливы равенства

$$S_n'''(x_i + 0) = \frac{M_{i+1} - M_i}{h_i}, \quad i = 1, \ldots, n-1,$$

$$S_n'''(x_i - 0) = \frac{M_i - M_{i-1}}{h_{i-1}}, \quad i = 2, \ldots, n,$$

из которых следует представление

$$Z_1 = \frac{M_2 - M_1}{h_1}, \quad Z_i = \frac{M_{i+1} - M_i}{h_i} - \frac{M_i - M_{i-1}}{h_{i-1}}, \quad i = 2, \ldots, n-1,$$

$$Z_n = \frac{M_{n-1} - M_n}{h_{n-1}}.$$

Тогда систему предыдущих равенств можно записать в векторно-матричной форме:

$$Z = H^T M, \tag{11.43}$$

где матрица H определена в формуле (11.35).

Из системы (11.35), справедливой для любого кубического сплайна, для сплайна $S_{n\alpha}(x)$ следует соотношение

$$H S_\alpha = AM - f. \tag{11.44}$$

Подействуем на обе части системы (11.42) матрицей H и выразим Z и $H S_\alpha$ через M согласно равенствам (11.43) и (11.44). Получаем систему линейных алгебраических уравнений относительно неизвестного вектора M:

$$(A + \alpha H K_\xi H^T) M = Hy + f. \tag{11.45}$$

Поскольку матрица A трехдиагональна, а матрица $H K_\xi H^T$ пятидиагональна, матрица $A + \alpha H K_\xi H^T$ пятидиагональна.

Для всех типов 1°–4° краевых условий матрицы A и $H K_\xi H^T$ положительно определены. Поэтому матрица $A + \alpha H K_\xi H^T$ положительно определена для любого $\alpha \geqslant 0$.

Итак, система (11.45) для любого $\alpha \geqslant 0$ имеет единственное решение.

Поскольку матрица $A + \alpha H K_\xi H^T$ пятидиагональна, система (11.45) эффективно решается методом обобщенной прогонки, требующим для своей реализации $\sim 30n$ арифметических операций [36].

После решения системы (11.45) значения сглаживающего кубического сплайна $S_{n\alpha}(x)$ в узлах x_i можно подсчитать по формуле

$$S_\alpha = y - \alpha K_\xi H^T M. \tag{11.46}$$

Если требуется вычислить значения $S'_{n\alpha}(x)$ и $S_{n\alpha}(x)$ в произвольных точках $x \in [a, b]$, то можно воспользоваться формулами Горнера (11.36) с заменой в этих формулах y_i на $S_{n\alpha}(x_i)$, $i = 1, \ldots, n$.

До сих пор предполагалось, что параметр сглаживания фиксирован. Если $\alpha \to 0$, то сглаживающий сплайн $S_{n\alpha}(x)$ стремится к интерполяционному сплайну $S_n(x)$ с условиями интерполяции $S_{n\alpha}(x_i) = y_i$, $i = 1, \ldots, n$. Поэтому при малых значениях $\alpha > 0$ у сглаживающего сплайна могут появиться случайные высокочастотные флуктуации из-за ошибок в измерениях y_i, $i = 1, \ldots, n$. Напротив, если $\alpha \to +\infty$, то сглаживающий сплайн стремится к линейной функции, для которой $\varphi''(x) = 0$. Поэтому при больших α происходит слишком большое сглаживание экспериментальных точек (x_i, y_i), $i = 1, \ldots, n$. Ясно, что подходящее значение параметра сглаживания должно быть согласовано с уровнем ошибок в измерениях y_i, $i = 1, \ldots, n$. В частности, если уровень ошибок стремится к нулю, то α должен также стремиться к нулю согласованно с уровнем ошибок.

Чаще всего для выбора подходящего значения параметра сглаживания используют способ невязки. Введем функционал $l_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(S_{n\alpha}(x_i) - y_i)^2 \frac{1}{\sigma_i^2}$, называемый *невязкой*. Тогда, согласно способу невязки [54], выбирают такое значение α_0, для которого

$$l_n(\alpha_0) = \beta, \tag{11.47}$$

где β рекомендуется брать из промежутка [0,8; 1].

Фактическое нахождение подходящего значения α_0 согласно способу невязки сводится к последовательному нахождению $S_{n\alpha}(x)$ для различных α и выбору из них такого $S_{n\alpha_0}(x)$, для которого наиболее точно выполняется равенство (11.47).

В частности, если все измерения равноточны, т. е. $\sigma_i^2 = \sigma^2$, $i = 1, \ldots, n$, то среднеарифметическая

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2,$$

где $f(x_i)$ — значения искомой функции в узлах, является состоятельной оценкой σ^2 (если бы σ^2 была неизвестной) и при достаточно большом n $\widehat{\sigma}^2$ близка к σ^2. Если предположить, что $S_{n\alpha}(x_i)$ достаточно близки к $f(x)$, и в $\widehat{\sigma}^2$ заменить $f(x_i)$ на $S_{n\alpha}(x_i)$, то

$$\widehat{\sigma}^2 \approx \frac{1}{n}\sum_{i=1}^{n}(S_{n\alpha}(x_i) - y_i)^2.$$

Приравняв $\widehat{\sigma}^2$ известной σ^2, получаем равенство для определения подходящего значения α:

$$\frac{1}{n}\sum_{i=1}^{n}(S_{n\alpha}(x_i) - y_i)^2 = \sigma^2 \tag{11.48}$$

(сравните с формулой (11.47)).

Существуют другие методы определения подходящего значения параметра сглаживания, в том числе для случая, когда дисперсия ошибок σ^2 неизвестна [22].

Пусть σ^2 неизвестна. Тогда для определения подходящего значения α_0 используют способ перекрестной значимости, согласно которому

$$\alpha_0 = \arg \min_{\alpha} M\|y - S_\alpha\|,$$

где усреднение производится по различным измерениям значений искомой функции в узлах x_i, $i = 1, \ldots, n$.

Поскольку обычно исследователю известна лишь одна реализация измерений y, то функционал $M\|y - S_\alpha\|$ заменяют на

$$u(\alpha) = \sum_{k=1}^{n} (y_k - S_k(x_k))^2, \tag{11.49}$$

где $S_k(x)$ — сглаживающий кубический сплайн, построенный по $n-1$ измерению $y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_n$ в узлах $x_1 < x_2 < \ldots < x_{k-1} < x_{k+1} < \ldots < x_n$, $k = 1, \ldots, n$.

За подходящее значение параметра сглаживания берут то, которое минимизирует функционал перекрестной значимости (11.49). Численная реализация способа перекрестной значимости требует построения семейства сглаживающих кубических сплайнов $\{S_k(x),\ k = 1, \ldots, n\}$ для различных значений α.

В заключение параграфа приведем схему восстановления функциональной зависимости с помощью робастных сглаживающих кубических сплайнов.

Вместо функционала (11.40) рассмотрим робастный сглаживающий функционал

$$J_{\text{роб}} = \sum_{i=1}^{n} \rho\Big(\frac{y_i - \varphi(x_i)}{\sigma_i}\Big) + \alpha \int_a^b (\varphi''(x))^2\, dx, \tag{11.50}$$

где $\rho(s)$ — функция Хьюбера.

Следующая теорема доказывается аналогично теореме 11.2.

Теорема 11.4. *Решением задачи на минимум функционала* (11.50) *в классе* w_k *является кубический сплайн* $S_n^{**}(x)$, *удовлетворяющий краевым условиям* K°.

В дальнейшем $S_n^{**}(x)$ называется *робастным сглаживающим кубическим сплайном* и обозначается $S_{\text{роб}}(x)$.

Обозначим $S_{\text{роб}} = (S_{\text{роб}}(x_1), \ldots, S_{\text{роб}}(x_n))^T$, $M = (M_1, \ldots, M_n)^T$, $M_i = S''_{\text{роб}}(x_i)$.

Можно показать, что вектор $S_{\text{роб}}$ связан с M соотношением

$$S_{\text{роб}} = y' - \alpha K_\xi H^T M, \tag{11.51}$$

а вектор вторых производных находится из системы нелинейных уравнений

$$(A + \alpha\, H\, K_\xi H^T)M = Hy' + f, \tag{11.52}$$

где матрицы A, H, K_ξ и векторы y, f такие же, как и в (11.45), а вектор $y' = (y'_1, \ldots, y'_n)^T$ определяется равенствами

$$y'_i = \begin{cases} y_i, & \text{если } i \in I_0, \\ S_{\text{роб}}(x_i) + K\sigma_i, & \text{если } i \in I_+, \\ S_{\text{роб}}(x_i) - K\sigma_i, & \text{если } i \in I_-, \end{cases}$$

$$I_0 = \{i:\ |y_i - S_{\text{роб}}(x_i)| \leqslant K\sigma_i\},$$

$$I_+ = \{i:\ y_i - S_{\text{роб}}(x_i) > K\sigma_i\}, \quad I_- = \{i:\ y_i - S_{\text{роб}}(x_i) < -K\sigma_i\}.$$

Для решения системы нелинейных уравнений (11.52) можно использовать нижеследующий итерационный процесс:

$$(A + \alpha H K_\xi H^T) M^{(l+1)} = H\, y^{(l)} + f, \tag{11.53}$$

где

$$y_i^{(l)} = \begin{cases} y_i, & \text{если } i \in I_0^{(l)}, \\ S_i^{(l)} + K\sigma_i, & \text{если } i \in I_+^{(l)}, \\ S_i^{(l)} - K\sigma_i, & \text{если } i \in I_-^{(l)}, \end{cases}$$

$$S^{(l)} = (S_1^{(l)}, \ldots, S_n^{(l)})^T = y^{(l-1)} - \alpha\, K_\xi H^T M^{(l)},$$

$$I_0^{(l)} = \{i:\ |y_i - S_i^{(l)}| \leqslant K\sigma_i\},$$

$$I_+^{(l)} = \{i:\ y_i - S_i^{(l)} > K\sigma_i\}, \quad I_-^{(l)} = \{i:\ y_i - S_i^{(l)} < -K\sigma_i\}.$$

В качестве начального приближения $S^{(0)}$ можно взять сглаживающий кубический сплайн, задаваемый равенствами (11.45), (11.46).

Итерационный процесс (11.53) сходится за конечное число шагов к решению системы (11.51), (11.52).

Заметим, что итерационный процесс (11.53) обладает высокой вычислительной эффективностью, в частности из-за того, что на каждом шаге процесса решается система (11.53) с одной и той же пятидиагональной матрицей $A + \alpha H K_\varphi (H)^T$, что дает возможность, обратив ее однажды, «обслуживать» все дальнейшие итерации.

На рис. 11.3 представлены результаты использования кубического сглаживающего (сплошная линия) и робастного кубического сглаживающего (пунктирная линия) сплайнов. Символами ($\bigstar$, $+$) отмечены значения с $i \in I_+$ и $i \in I_-$. На нижнем графике показаны величины корректировок значений исходного временного ряда в результате применения итерационного процесса минимизации робастного сглаживающего функционала (11.50).

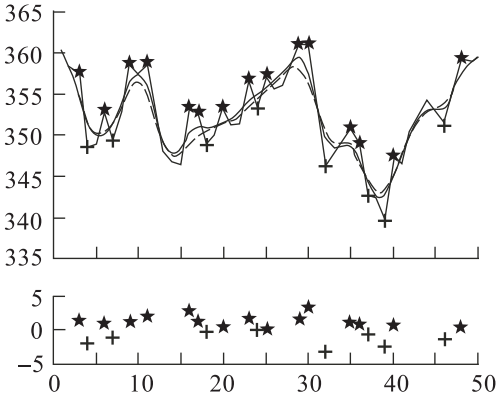

Рис. 11.3

Робастные сглаживающие кубические сплайны были успешно использованы для решения ряда некорректных задач при обработке данных физических экспериментов [7, 8].

11.5. Метод вейвлетов

В последние годы для исследования неопределенных данных был разработан новый инструмент — *вейвлет-преобразование*.

Ниже дано краткое введение в теорию вейвлет-преобразования и его применения при обработке неопределенных данных.

Термин «wavelet» был впервые введен в 1909 г. Альфредом Хааром (Alfred Haar). В дальнейшем, особенно в последние 15 лет, вейвлет-анализ развивался в нескольких направлениях усилиями многих ученых (Jean Morlet, Y. Meyer, Stephane Mallat, Ingrid Daubechies) [95, 99, 102–104, 107].

В определенной степени вейвлет-преобразование является обобщением преобразования Фурье.

Напомним, что любая функция $f(t)$ из класса $L_2(-\pi, \pi)$ может быть представлена в виде ряда Фурье:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{ikt}, \tag{11.54}$$

где $c_k = \dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} f(t) e^{-ikt}\, dt.$

Функции $\psi_k(t) = e^{ikt}$ образуют ортогональный базис пространства $L_2(-\pi, \pi)$ и построены с помощью масштабного преобразования базисной функции $\psi(t) = e^{it}$.

Для всех $f(t) \in L_2(-\infty, \infty)$ верно $\int\limits_{-\infty}^{\infty} |f(t)|^2\, dt < \infty$ и справедливо разложение в интеграл Фурье:

$$f(t) = \int\limits_{-\infty}^{\infty} c(\omega) e^{i\omega t}\, d\omega, \quad c(\omega) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} f(t) e^{-i\omega t}\, dt.$$

В то же время, поскольку $e^{ikt} \not\to 0$ при $t \to \infty$, e^{ikt} не может быть счетным базисом в пространстве $L_2(-\infty, \infty)$.

Поставим задачу о сохранении представления типа (11.54) для $f(t) \in L_2(-\infty, \infty)$:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \psi_k(t), \quad c_k = \frac{1}{C} \int\limits_{-\infty}^{\infty} f(t) \widetilde{\psi}_k(t)\, dt, \tag{11.55}$$

где $\psi_k(t)$ образуют ортогональный базис пространства $L_2(-\infty, \infty)$.

Оказывается, можно построить ортогональный базис в $L_2(-\infty, \infty)$ на основе одной базисной функции $\psi(t)$, которая и называется (*базисным*) *вейвлетом*, но с помощью двух операций: масштабного преобразования и параллельного сдвига, например $\psi_{nk}(t) = 2^{n/2}\psi(2^n t - k)$.

Для того чтобы $\psi(t) \in L_2(-\infty, \infty)$ была базисным вейвлетом, необходимо выполнение двух условий:

1° $\psi(t) = 0$ за пределами некоторого конечного промежутка либо $\psi(t) \to 0$ при $t \to \infty$;

2° $\int\limits_{-\infty}^{\infty} \psi(t)\, dt = 0.$

Кроме того, поскольку $\psi(t) \in L_2(-\infty, \infty)$, то должно выполняться неравенство $\int\limits_{-\infty}^{\infty} |\psi(t)|^2\, dt < \infty$.

Иногда требуют равенства нулю всех моментов до определенного порядка, т. е. $\int\limits_{-\infty}^{\infty} t^k \psi(t)\, dt = 0$, $k = 1, \ldots, m$.

Ниже даны примеры базисных вейвлетов, используемых в вейвлет-анализе.

Пример 11.1 (Вейвлет Хаара).

$$\psi(t) = \begin{cases} 1, & 0 \leqslant t \leqslant 0{,}5, \\ -1, & 0{,}5 < t \leqslant 1, \\ 0, & t \notin [0, 1]. \end{cases} \tag{11.56}$$

Базисный вейвлет Хаара имеет конечный носитель $t \in [0, 1]$.

Функции $\psi_{nk}(t) = 2^{n/2}\psi(2^n t - k)$, $n, k = 0, \pm 1, \pm 2, \ldots$, образуют ортонормированный базис пространства $L_2(-\infty, \infty)$.

Определим семейство функций

$$\widetilde{\psi}_{nk}(t) = \sum_{j=-\infty}^{\infty} 2^{n/2}\psi(2^n(t+j)-k),$$
$$n = 0,1,2,\ldots, \quad k = 0,1,\ldots,2^n-1. \tag{11.57}$$

Можно показать, что $\widetilde{\psi}_{nk}(t)$ — периодические функции с периодом $T=1$. Их называют *периодическими вейвлетами Хаара.*

Периодические вейвлеты (11.57) образуют ортогональный базис пространства $L_2(0,1)$.

Пример 11.2 (Французская шляпа — FHAT-wavelet).

$$\psi(t) = \begin{cases} -1/2, & t \in [-1,-1/3), \\ 1, & t \in [-1/3,1/3], \\ 1/2, & t \in (1/3,1], \\ 0, & t \notin [-1,1]. \end{cases} \tag{11.58}$$

Этот базисный вейвлет имеет конечный носитель $t \in [-1,1]$.

Нижеследующие базисные вейвлеты, в отличие от предыдущих, непрерывны на всей числовой оси.

Пример 11.3. Семейство базисных вейвлетов $\psi(t;\ m)$ определяется равенством

$$\psi(t;\ m) = (-1)^m \frac{d^m}{dt^m} e^{-t^2/2}, \qquad m = 1,2,\ldots \tag{11.59}$$

Вейвлет $\psi(t;\ 1)$ называют *волновым вейвлетом* (wave-wavelet), вейвлет $\psi(t;\ 2)$ — *мексиканской шляпой.*

Базисные вейвлеты семейства $\psi(t;\ m)$ имеют бесконечный носитель $t \in (-\infty,\infty)$, но очень быстро стремятся к нулю при $t \to \infty$.

Пример 11.4 (Вейвлеты Добеши — Daubechies wavelets). Базисные вейвлеты $\psi(t)$ вейвлетов Добеши непрерывно дифференцируемы, но, в отличие от базисных вейвлетов семейства $\psi(t;\ m)$, имеют конечный носитель.

Семейство базисных вейвлетов Добеши принято снабжать символом Daub$N$, $N = 2,3,\ldots$ Наиболее часто используется вейвлет Daub4 четвертого порядка. Для ψ_{Daub4} справедливо равенство $\int\limits_{-\infty}^{\infty} t\psi_{\mathrm{Daub4}}(t)\,dt = 0$ и его носитель $t \in [0,3]$.

Базисные вейвлеты $\psi_{\mathrm{Daub}N}(t)$ могут быть получены из равенства

$$\psi_{\mathrm{Daub}N}(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} (-1)^k C_{1-k}(N)\Phi(N,2t-k), \tag{11.60}$$

где $C_i(N)$ — коэффициенты, зависящие от порядка N базисного вейвлета Добеши; для каждого N число $C_i(N)$, отличных от нуля, конечно. Например, для базисного вейвлета $\psi_{\mathrm{Daub4}}(t)$

$$C_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \quad C_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \quad C_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \quad C_3 = \frac{1-\sqrt{3}}{4\sqrt{2}},$$

а остальные C_i равны нулю.

Функция $\Phi(N,t)$ называется *сканирующей* для базисного вейвлета ψ_{Daub4} и определяется как предел

$$\Phi(N,t) = \lim_{n\to\infty} \Phi_n(N,t), \tag{11.61}$$

где $\Phi_0(N,t) = \eta(x) - \eta(x-1)$, $\eta(x)$ — функция Хевисайда, $\Phi_n(N,t) = \sqrt{2} \sum\limits_{k=-\infty}^{\infty} C_k(N)\Phi_{n-1}(N, 2t-k)$.

Все функции $\Phi(N,t)$, $N = 2,3,\ldots$, непрерывны и имеют носитель $t \in [0,3]$. Из равенства (11.60) следует, что базисные вейвлеты $\psi_{\mathrm{Daub}N}$ также непрерывны и имеют носитель $t \in [0,3]$.

Аналогично (11.57) определим семейство функций

$$\widetilde{\psi}_{nk}(N,t) = \sum_{j=-\infty}^{\infty} 2^{n/2}\psi_{\mathrm{Daub}N}(2^n(t+j)-k), \quad n = 0,1,2,\ldots, \quad k = 0,1,\ldots,2^{n-1}. \tag{11.62}$$

Можно показать, что $\widetilde{\psi}_{nk}(N,t)$ — периодические функции с периодом $T = 1$. Периодические вейвлеты Добеши (11.62) образуют ортонормированный базис пространства $L_2(0,1)$. Для любой $f(t) \in L_2(0,1)$ справедливо разложение

$$f(t) = c_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \alpha_{nk}(N)\widetilde{\psi}_{nk}(N,t), \tag{11.63}$$

где

$$c_0 = \int_0^1 f(t)\,dt, \quad \alpha_{nk}(N) = \int_0^1 f(t)\widetilde{\psi}_{nk}(N,t)\,dt.$$

С вейвлетами Добеши тесно связаны еще два класса вейвлетов: *кофлеты* (coiflets) и *симлеты* (symlets), которые так же, как и вейвлеты Добеши, имеют конечный носитель. Например, базисные кофлеты порядка N, сконструированные Добеши и Кофманом (R. Coifman), имеют $2n$ моментов, равных нулю, и носитель $t \in [-N, 2N-1]$. Так же как и вейвлеты Добеши, базисные вейвлеты семейства кофлетов могут быть представлены в виде разложения (11.60) через сканирующую функцию (разную для разных кофлетов) с конечным числом отличных от нуля коэффициентов C_i.

Пусть базисный вейвлет $\psi(t)$ удовлетворяет условию нормировки $\int\limits_{-\infty}^{\infty} |\psi(t)|^2\,dt = 1$. Покажем, что все вейвлеты семейства

$$\psi_{nk}(t) = 2^{n/2}\psi(2^n t - k), \quad n, k = 0, \pm 1, \ldots,$$

также нормированы. Действительно,

$$\int\limits_{-\infty}^{\infty} \psi_{nk}^2(t)\,dt = 2^n \int\limits_{-\infty}^{\infty} \psi^2(2^n t - k)\,dt = \int\limits_{-\infty}^{\infty} \psi^2(t')\,dt' = 1,$$

где $t' = 2^n t - k$, $dt = 2^{-n}dt'$. Таким образом, для всех вышеперечисленных семейств вейвлетов $\psi_{nk}(t)$, $n, k = 0, \pm 1, \ldots$, выполнено условие ортонормированности

$$\int\limits_{-\infty}^{\infty} \psi_{nk}(t)\psi_{n'k'}(t)\,dt = \delta_{nn'}\delta_{kk'},$$

где $\delta_{nn'}$ — символ Кронекера.

Одной из наиболее значительных прикладных областей вейвлетов является сжатие информации, содержащейся в сигналах. Дело в том, что полезный сигнал, подлежащий сжатию, представляет, как правило, суперпозицию локальных образований солитонного типа, а образования такого рода с высокой точностью (в отличие от разложений по синусам и косинусам) могут быть представлены в виде разложения по ортонормированной системе вейвлетов с небольшим числом отличных от нуля коэффициентов.

Другое направление применения вейвлетов связано со сглаживанием хаотических временных процессов, т. е. выделением детерминированных и хаотических компонент.

Более подробную информацию о свойствах вейвлетов, вейвлет-анализе и его применении можно найти в [102, 104].

Глава 12
МЕТОДЫ ПРОГНОЗИРОВАНИЯ ХАОТИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ

Прогнозирование временных процессов — одно из основных направлений исследований во многих прикладных областях и в науке в целом. Особую значимость прогнозирование имеет при исследовании экономических процессов, в частности процессов в финансовой области. Уметь прогнозировать с приемлемой точностью развитие исследуемого временного экономического процесса — базовая предпосылка для успешного управления процессом и реализации максимальной эффективности работы управляемой экономической системы.

К настоящему времени разработано много методов и основанных на них алгоритмов прогнозирования временных процессов [1, 4, 5, 37, 50, 60, 61, 75, 98]. Ниже рассмотрены сравнительно новые методы прогнозирования, два из которых используют априорную информацию экспертного характера, а третий основан на выделении так называемых *главных компонент* исследуемого временного ряда.

12.1. Методы прогнозирования с использованием априорной информации и ортогональных полиномов

При прогнозировании временных рядов часто в распоряжении исследователей имеется априорная информация о значениях временного ряда в будущие моменты времени, для которых и производится прогноз. Ясно, что при решении практических задач априорная информация в абсолютном большинстве случаев носит неопределенный характер. В таких ситуациях задача прогнозирования состоит в оптимальном учете априорной информации при ее «объединении» в процессе прогнозирования с информацией, содержащейся в уже реализованной части временного ряда.

Ниже предполагается, что априорная информация задана в виде априорных экспертных оценок значений временного ряда

$$y_1, \ldots, y_L \tag{12.1}$$

для будущих временных дат

$$t_1, \ldots, t_L$$

и задаются уровни погрешности этих оценок в виде числовых значений дисперсий:

$$\sigma_1^2, \ldots, \sigma_L^2. \tag{12.2}$$

Предполагается также, что в распоряжении исследователя имеется реализованная часть исследуемого временного ряда для прошедших временных дат $t_{-N}, t_{-N+1}, \ldots, t_{-1}$ и настоящей даты t_0.

Для осуществления прогноза значений временного ряда на будущие даты $t_1, t_2, \ldots, t_L$, объединяющего информацию в реализованной части временного ряда и экспертную информацию, в этом параграфе мы используем схему выделенной детерминированной компоненты с помощью разложения временного ряда по базисной системе полиномов, ортогональных на множестве значений аргумента $t_{-N}, t_{-N+1}, \ldots, t_0, t_1, \ldots, t_L$ (12.2) с весами

$$w_k = 1/\sigma_k^2, \qquad k = -N, -N+1, \ldots, 0, 1, \ldots, L. \tag{12.3}$$

Согласно результатам, представленным в § 11.1, детерминированная компонента временного ряда дается равенством

$$y_{\text{дет}}(t_k) = \sum_{j=1}^{m} \widehat{u}_j \, \varphi_j(t_k), \qquad k = -N, -N+1, \ldots, 0, 1, 2, \ldots, L, \tag{12.4}$$

где

$$\widehat{u}_j = \sum_{k=-N}^{L} w_k \varphi_j(t_k) y_k, \tag{12.5}$$

$y_{-N}, y_{-N+1}, \ldots, y_0$ — значения реализованной части исследуемого временного ряда, $y_1, \ldots, y_L$ — экспертные оценки (12.1).

Окончательно прогнозные оценки $y_{\text{прог}}(t_k)$ значений временного ряда на будущие даты $t_1, t_2, \ldots, t_L$ даются равенствами

$$y_{\text{прог}}(t_k) = y_{\text{дет}}(t_k), \qquad k = 1, \ldots, L, \tag{12.6}$$

где $y_{\text{дет}}(t_k)$ определяются из равенств (12.4).

Замечание 12.1. Точность прогноза (12.6) во многом определяется точностью экспертных оценок (12.1), задаваемой дисперсиями (12.2). В частности, при $\sigma_k^2 \to 0$, $y_{\text{прог}}(t_k) \to y_k$, $k = 1, \ldots, L$ и информация в реализованной части временного ряда становится ненужной.

Замечание 12.2. Если экспертные оценки (12.1) отсутствуют, то по умолчанию будем полагать

$$y_1 = y_2, \ldots, y_L = y_0. \tag{12.7}$$

В этом случае достаточно надежный прогноз можно получить лишь для небольшого числа временных шагов вперед (в предположении выполнения равенства $h_k = t_{k+1} - t_k = \text{const}$).

Замечание 12.3. Если дисперсии $\sigma_{-N}^2, \sigma_{-N+1}^2, \ldots, \sigma_0^2$ априори неизвестны, то положим $\sigma_{-N}^2 = \sigma_{-N+1}^2 = \ldots = \sigma_0^2 = \sigma^2$, где σ^2 предварительно определяется из равенства

$$\sigma^2 = \frac{1}{N} \sum_{k=-N}^{0} (y_k - \widehat{y}_k)^2, \tag{12.8}$$

где $\widehat{y}_k = \sum\limits_{j=1}^{m} u'_j \varphi'_j(t_k)$, $k = -N, -N+1, \ldots, 0$,

$$u'_j = \sum_{k=-N}^{0} \varphi'_j(t_k) y_k, \tag{12.9}$$

а $\{\varphi'_j(t)\}$ — система полиномов, ортогональных на множестве значений аргумента $t_{-N}, t_{-N+1}, \ldots, t_0$ с весом, равным единице.

Замечание 12.4. При отсутствии экспертных оценок и использовании равенств (12.7) можно положить $\sigma_k^2 = \sigma_0^2(1+\beta k)$, $k = 1, \ldots, L$, $\beta \geqslant 0$.

Замечание 12.5. Все используемые в представленной выше схеме прогнозирования параметры (m, N, L) следует определять на этапе обучения схемы прогнозирования с использованием реализованной части исследуемого временного ряда. Критерием отбора подходящих значений параметров в процессе обучения прогнозируемой схемы должно являться качество прогноза, например сумма квадратов невязок между прогнозируемыми и реализованными значениями исследуемого временного ряда.

Если в реализованной части $y_{-N}, \ldots, y_0$ временного ряда имеются большие выбросы, то вместо предложенной выше схемы выделения детерминированной компоненты должна применяться робастная схема.

В этом случае, согласно 11.1, детерминированная компонента временного ряда дается равенством

$$y_{\text{дет}}(t_k) = \sum_{j=1}^{m} u_{\text{робj}} \varphi_j(t_k), \quad k = -N, -N+1, \ldots, 0, 1, \ldots, L, \tag{12.10}$$

$$u_{\text{робj}} = \sum_{k=-N}^{L} w_k \varphi_j(t_k) \widetilde{y}_k, \qquad j = 1, \ldots, m, \tag{12.11}$$

где

$$\widetilde{y}_k = \begin{cases} y_k, & \text{если } k \in I_0, \\ \sum\limits_{j=1}^{m} u_{\text{робj}} \varphi_j(t_k) + K\sigma_k, & \text{если } k \in I_+, \\ \sum\limits_{j=1}^{m} u_{\text{робj}} \varphi_j(t_k) - K\sigma_k, & \text{если } k \in I_-, \end{cases}$$

$$I_0 = \{k : \ |y_k - \sum_{j=1}^{m} u_{\text{робj}} \varphi_j(t_k)| \leqslant K\sigma_k\},$$

$$I_+ = \{k : \ y_k - \sum_{j=1}^{m} u_{\text{робj}} \varphi_j(t_k) > K\sigma_k\},$$

$$I_- = \{k : \ y_k - \sum_{j=1}^{m} u_{\text{робj}} \varphi_j(t_k) < -K\sigma_k\}.$$

Нелинейная робастная оценка (12.11) может быть найдена с помощью итерационной процедуры (11.12), (11.13).

Окончательные прогнозные оценки $y_{\text{прог}}(t_k)$ значений временного ряда даются равенствами (12.6), где $y_{\text{дет}}(t_k)$ определяются из равенства (12.10).

12.2. Методы прогнозирования с использованием априорной информации и линейных сплайнов

Так же как и в предыдущем параграфе, предполагается, что априорная информация задана в виде экспертных оценок (12.1) и погрешностей этих оценок (12.2).

В этом параграфе для выделения детерминированной компоненты мы используем разложение по системе линейных сплайнов, основанной на минимизации сглаживающего функционала

$$\sum_{k=-N}^{L} w_k(y_k - y(t_k))^2 + \alpha \sum_{k=-N}^{L-1} \frac{1}{h_k}\,(y(t_{k+1}) - y(t_k))^2, \qquad (12.12)$$

где w_k определяются равенством (12.3), $h_k = t_{k+1} - t_k$, $\alpha > 0$ — параметр сглаживания.

Тем самым, согласно § 11.2, вектор детерминированной компоненты $Y_{\text{дет}} = (y_{\text{дет}}(t_{-N}), \ldots, y_{\text{дет}}(t_L))^T$ определяется из системы линейных алгебраических уравнений

$$A\,Y_{\text{дет}} = F, \qquad (12.13)$$

где симметричная трехдиагональная матрица A и вектор F даются равенствами

$$A = \begin{pmatrix} w_{-N} + \dfrac{\alpha}{h_{-N+1}} & -\dfrac{\alpha}{h_{-N+1}} & 0 & 0 & \ldots \\ -\dfrac{\alpha}{h_{-N+1}} & w_{-N+1} + \dfrac{\alpha}{h_{-N+1}} + \dfrac{\alpha}{h_{-N}} & -\dfrac{\alpha}{h_{-N+1}} & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots \end{pmatrix}$$

$$\begin{pmatrix} \ldots & 0 & 0 & & \\ \ldots & 0 & & 0 & \\ \ldots & \ldots & \ldots & & \\ \ldots & 0 & -\dfrac{\alpha}{h_{L-1}} & w_L + \dfrac{\alpha}{h_{L-1}} & \end{pmatrix},$$

$$F = (y_{-N}w_{-N}, \ldots, y_L w_L)^T.$$

Поскольку A — матрица с диагональным преобладанием, то для любых наборов исходных неопределенных данных (11.1)–(11.3) существует единственное решение $Y_{\text{дет}}$ системы (12.13).

Окончательно прогнозные значения даются последними L компонентами вектора $Y_{\text{дет}}$.

Для предлагаемой в этом параграфе схемы остаются справедливыми все замечания предыдущего параграфа с заменой в замечании 12.5 параметров (m, N, L) на параметры (α, N, L).

Если в реализованной части $y_{-N}, \ldots, y_0$ временного ряда присутствуют большие выбросы, то для выделения детерминированной компоненты временного ряда $y_{-N}, y_{-N+1}, \ldots, y_0, y_1, \ldots, y_L$ вместо сглаживающих линейных сплайнов должны использоваться робастные линейные сглаживающие сплайны.

Согласно § 11.3 в этом случае вместо сглаживающего функционала (12.12) используем робастный сглаживающий функционал

$$\sum_{k=-N}^{L} \rho(\sqrt{w_k}\,(y_k - y(t_k)) + \alpha \sum_{k=-N}^{L-1} \frac{1}{h_k}\,(y(t_{k+1}) - y(t_k))^2, \qquad (12.14)$$

где $\rho(s)$ — четная дифференцируемая функция, $\rho(0) = 0$, $\rho(s) \sim s^2$ для малых $|s|$, $\rho(s)/s^2 \to 0$ при $s \to \infty$.

Если в качестве $\rho(s)$ взять функцию Хьюбера, то детерминированная компонента $Y_{\text{дет}} = (y_{\text{дет}}(t_{-N}), \ldots, y_{\text{дет}}(t_L))^T$ временного ряда $y_{-N}, \ldots, y_L$, определяемая как вектор, минимизирующий робастный сглаживающий функционал (12.14), является решением системы нелинейных уравнений

$$A\,Y_{\text{дет}} = F', \qquad (12.15)$$

где

$$F' = (w_{-N}y'_{-N}, \ldots, w_L y'_L)^T, \quad y'_k = \begin{cases} y_k, & \text{если } k \in I_0, \\ y_{\text{дет}} + K\sigma_k, & \text{если } k \in I_+, \\ y_{\text{дет}} - K\sigma_k, & \text{если } k \in I_-, \end{cases}$$

$I_0 = \{k\colon |y_k - y_{\text{дет}}(t_k)| \leqslant K\sigma_k\}$, $I_+ = \{k\colon y_k - y_{\text{дет}}(t_k) > K\sigma_k\}$, $I_- = \{k\colon y_k - y_{\text{дет}}(t_k) < -K\sigma_k\}$.

Для фактического нахождения $Y_{\text{дет}}$ можно использовать итерационную процедуру, изложенную в § 11.3.

Если в качестве $\rho(s)$ взять функцию Тьюки

$$\rho(s) = \begin{cases} s^2, & |s| \leqslant R, \\ R^2, & |s| > R, \end{cases}$$

то $Y_{\text{дет}}$ как решение задачи на минимум функционала (12.14) является решением системы нелинейных уравнений (12.15), где

$$y'_k = \begin{cases} y_k, & \text{если } k \in I_0, \\ y_{\text{дет}}(t_k), & \text{если } k \in I_R, \end{cases} \qquad (12.16)$$

$$I_0 = \{k : |y_k - y_{\text{дет}}(t_k)| \leqslant R\sigma_k\}, \quad I_R = \{k : |y_k - y_{\text{дет}}(t_k)| > R\sigma_k\}.$$

Для фактического нахождения $Y_{\text{дет}}$ как решения нелинейной системы (12.16) можно использовать описанную в § 11.3 итерационную процедуру, сходящуюся к искомому решению за конечное число шагов.

Как и прежде, прогнозные значения для дат $t_1, \ldots, t_L$ определяются равенствами

$$y_{\text{прог}}(t_k) = y_{\text{дет}}(t_k), \qquad k = 1, \ldots, L.$$

12.3. Методы прогнозирования с использованием сингулярно-спектрального анализа

Методы исследования временных рядов, основанные на сингулярно-спектральном анализе (Singular-Spectrum Analysis — SSA), были развиты в основном на протяжении последних 16 лет [24, 92–94, 97, 98, 100, 110], хотя основы этого метода были заложены в более ранних работах [12, 39]. В настоящее время SSA успешно используется для выявления скрытых периодичностей и выделения детерминированных компонент, включая сглаживание исследуемых временных рядов. В последние годы разрабатываются алгоритмы прогнозирования временных рядов, использующих SSA [24, 98].

Пусть задан исходный временной ряд

$$y_1, \ldots, y_n. \tag{12.17}$$

Зададим два целых числа M и L, называемых *шириной окна* и *числом шагов* соответственно, и сформируем $(L \times M)$-матрицу пошагового движения:

$$Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_M \\ y_2 & y_3 & \cdots & y_{M+1} \\ \cdots & \cdots & \cdots & \cdots \\ y_L & y_{L+1} & \cdots & y_{L+M-1} \end{pmatrix}. \tag{12.18}$$

Рассмотрим задачу на собственные значения симметричной неотрицательной $(M \times M)$-матрицы $Y^T Y$:

$$Y^T Y \psi = \mu \psi. \tag{12.19}$$

Собственные значения матрицы $Y^T Y$ называются *сингулярными числами матрицы* Y (что и дало название SSA).

Произведем ранжирование сингулярных чисел в порядке убывания:

$$\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_M \geqslant 0, \tag{12.20}$$

обозначив через $\psi_j = (\psi_{j1}, \ldots, \psi_{jM})^T$, $j = 1, \ldots, M$, соответствующую им ортонормированную систему собственных векторов.

Сформируем ортогональную матрицу $\Omega = (\psi_1, \ldots, \psi_M)$ и определим $(L \times M)$-матрицу главных компонент V равенством

$$V = Y\,\Omega = (V_1, \ldots, V_M), \tag{12.21}$$

где V_j, $j = 1, \ldots, M$, суть L-мерные векторы, называемые *главными компонентами*.

Из равенств (12.21) следует соотношение

$$Y = V\,\Omega^T = \sum_{j=1}^{M} V_j \psi_j^T, \tag{12.22}$$

представляющее собой разложение исходного временного ряда по главным компонентам V_j, $j = 1, \ldots, M$.

Числовые значения сингулярных чисел μ_j определяют уровень «информационного вклада» компоненты V_j во временной ряд, причем компонентам с большим индексом j соответствуют изменения временного ряда большей частоты.

Поэтому для осуществления операции сглаживания необходимо отбросить несколько последних по индексу главных компонент, соответствующих шумовой хаотической компоненте, оставив первые m из них, $V_1, \ldots, V_m$, и вместо (12.22) использовать равенство

$$Y' = \sum_{j=1}^{m} V_j \psi_j^T. \tag{12.23}$$

Сглаженные детерминированные значения $y_{\text{дет}}(t_k)$ исходного временного ряда получаются путем усреднения значений элементов на побочных диагоналях матрицы $Y' = \{y'_{ij}\}$, $i = 1, \ldots, L$, $j = 1, \ldots, M$: $y_{\text{дет}}(t_1) = y'_{11}$, $y_{\text{дет}}(t_2) = (y'_{12} + y'_{21})/2$, $y_{\text{дет}}(t_3) = (y'_{13} + y'_{22} + y'_{31})/3$, ...

Одна из возможных схем прогнозирования с помощью SSA основана на рассмотрении системы линейных алгебраических уравнений

$$\begin{aligned} \sum_{i=1}^{m} x_i\, \psi_{1i} &= y_{n-M+2}, \\ &\cdots\cdots\cdots\cdots \\ \sum_{i=1}^{m} x_i\, \psi_{M-1,i} &= y_n, \end{aligned} \tag{12.24}$$

где ψ_{ij} — i-я компонента вектора ψ_j, y_i — члены исходного или сглаженного рядов (12.17).

В системе (12.24) m неизвестных $x_1, \ldots, x_m$ и $M - 1$ уравнение.

Если система (12.24) совместна, то прогнозируемое значение y_{n+1} может быть представлено в виде

$$y_{n+1} = \sum_{i=1}^{m} x_i \psi_{Mi}, \tag{12.25}$$

где $x_1, \ldots, x_m$ — решение системы (12.24).

Пусть исходный временной ряд (12.17) порожден однородным конечно-разностным уравнением порядка m с постоянными коэффици-

ентами:

$$y_{m+i} + a_1 y_{m+i-1} + \ldots + a_m y_i = 0, \qquad i = 1,2,\ldots \tag{12.26}$$

Тогда при $M \geqslant m+1$ система (12.24) совместна и имеет единственное решение.

В этом случае можно продолжить прогноз на последующие временные даты, используя систему

$$\begin{gathered} \sum_{i=1}^{m} x_i^{(k)} \psi_{1i} = y_{n-M+2+k}, \\ \ldots\ldots\ldots\ldots \\ \sum_{i=1}^{m} x_i^{(k)} \psi_{M-1,i} = y_{n+k}, \end{gathered} \tag{12.27}$$

и прогноз

$$y_{n+k+1} = \sum_{i=1}^{m} x_i^{(k)} \psi_{Mi}, \qquad k = 1,2,\ldots \tag{12.28}$$

Для $k = 0$ система (12.27) и прогноз (12.28) соответствуют (12.24) и (12.25).

Системы (12.27) совместны или несовместны одновременно для всех $k = 0,1,2,\ldots$

Системы (12.27) несовместны (при $M > m$) тогда и только тогда, когда исходный ряд не порожден уравнением типа (12.26).

Систему (12.24) запишем в векторно-матричном виде:

$$A\,X = F - \varepsilon, \tag{12.29}$$

где $F = (y_{n-M+2}, \ldots, y_n)^T$ — вектор правых частей уравнений системы (12.24), ε — вектор «погрешностей» по отношению к неизвестному вектору F^*, для которого система $A\,X = F^*$ совместна. [1])

В условиях несовместности системы (12.24) определим вектор X^* как МНК-оценку вектора X линейной регрессионной модели (12.29).

Имеем

$$X^* = (A^T A)^{-1} A^T F, \tag{12.30}$$

где $X^* = (x_1^*, \ldots, x_m^*)^T$.

Прогнозируемое значение y_{n+1} определяется равенством

$$y_{n+1}^* = \sum_{i=1}^{m} x_i^* \psi_{Mi}. \tag{12.31}$$

Определим вектор $Y^* = (y_{n-M+2}^*, \ldots, y_n^*, y_{n+1}^*)^T$, где y_{n+1}^* дается равенством (12.31), $(y_{n-M+2}^*, \ldots, y_n^*)^T = A\,X^*$.

[1]) Если система (12.24) совместна, то $\varepsilon = 0$.

Вместо рекуррентных систем (12.27) рассмотрим рекуррентные системы

$$\begin{gathered}\sum_{i=1}^{m} x_i^{(k)}\psi_{1i} = y^*_{n-M+2+k},\\ \ldots\ldots\ldots\ldots\\ \sum_{i=1}^{m} x_i^{(k)}\psi_{M-1,i} = y^*_{n+k},\ k = 0,1,2,\ldots\end{gathered} \tag{12.32}$$

Все системы (12.32) совместны и каждая из них имеет единственное решение $X^{(k)*} = (x_1^{(k)*}, \ldots, x_m^{(k)*})^T$.

Определим прогноз для последующих временных дат равенством

$$y^*_{n+k+1} = \sum_{i=1}^{m} x_i^{(k)*}\psi_{Mi}, \qquad k = 1,2,\ldots \tag{12.33}$$

При реализации схем прогнозирования, определяемых формулами (12.25), (12.28), (12.31), (12.33), требуется задание ширины окна M, числа шагов L и числа выделяемых главных компонент m. Число m определяется размерностью (рангом) исследуемого временного ряда и может быть определено из анализа соотношений между числовыми значениями сингулярных чисел μ_j. Ширина окна M должна быть несколько больше числа m. И, наконец, число L должно быть больше M (иногда в несколько раз).

Оптимальные значения параметров (m, M, L) могут быть найдены, например, исходя из минимизации длины вектора невязок реализованных и прогнозных значений исследуемого временного ряда на отрезке временного ряда, соответствующего этапу обучения прогнозной модели.

В другой возможной схеме прогнозирования с помощью SSA рассматривается совместная система (12.24), где $y_i = y_{\text{дет}}(t_i)$.

Заметим, что конечно-разностное уравнение (12.26) можно записать в виде

$$Y_i = u_0 + \sum_{j=1}^{m-1} u_j X_{ji}, \tag{12.34}$$

где $Y_i = y_{m+i}/y_{m+i-1}$, $X_{ji} = y_{m+i+j-1}/y_{m+i-1}$, $u_0 = -a_1$, $u_j = -a_{j+1}$.

Если рассматривается хаотический временной ряд, детерминированная компонента которого удовлетворяет уравнению (12.26), то для исходного временного ряда можно рассмотреть линейную регрессионную модель

$$y_{m+i} + a_1 y_{m+i-1} + \ldots + a_m y_i + \varepsilon_i = 0,$$

которая в обозначениях уравнения (12.34) запишется в виде

$$Y_i = u_0 + \sum_{j=1}^{m-1} u_j X_{ji} + \varepsilon_i, \qquad i = 1,2,\ldots \tag{12.35}$$

Система равенств (12.35) полностью соответствует многофакторной линейной регрессионной модели (7.30) и тем самым между SSA и линейным регрессионным анализом существует очевидная взаимосвязь.

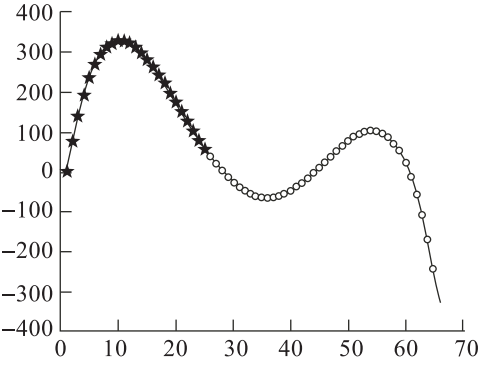

Рис. 12.1

Рисунок (12.1) демонстрирует применение метода сингулярно-спектрального анализа для прогнозирования ряда. Для примера использован полином 4-й степени (сплошная линия). Символом (★) обозначены значения, использованные при обучении, а символом (O) — спрогнозированные значения.

Глава 13

ПЛАНИРОВАНИЕ ОПТИМАЛЬНЫХ ИЗМЕРЕНИЙ ПРИ ВОССТАНОВЛЕНИИ ФУНКЦИОНАЛЬНЫХ ЗАВИСИМОСТЕЙ

В предыдущих главах оценка искомого вектора u проводилась на основе уже полученных измерений («исторических» данных). Но, как показали вышеприведенные исследования, точность оценки во многом зависит не только от количества измерений, но и от их расположения.

Пусть, например, рассматривается задача восстановления функциональной зависимости

$$y = \sum_{j=1}^{m} u_j \varphi_j(x),$$

где $\varphi_j(x)$, $j = 1, \ldots, m$, — заданная система функций, и пусть выполнены предположения классической схемы МНК, но с наличием корреляций между компонентами вектора случайных погрешностей, так что $K_\varepsilon = \sigma^2 \Omega$. Тогда, как мы знаем, оптимальной в классе всех несмещенных линейных оценок является взвешенная МНК-оценка (7.22), где $(n \times m)$-матрица A определяется равенством

$$A^T = (\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_n)),$$

$\varphi(x_i) = (\varphi_1(x_i), \ldots, \varphi_m(x_i))^T$ — вектор размерности m, а качество оценки определяется ее ковариационной матрицей, даваемой равенством $K_{\widehat{u}} = \sigma^2 (A^T \Omega^{-1} A)^{-1}$. Поскольку матрица A зависит от координат x_i, $i = 1, \ldots, n$, фиксации независимой переменной x, то $K_{\widehat{u}}$ также зависит от x_i, $i = 1, \ldots, n$. Таким образом, в зависимости от того, при каких значениях x_i, $i = 1, \ldots, n$, проведены измерения, будем получать различные по точности МНК-оценки. Из вышесказанного естественным образом возникает следующая задача. Пусть исследователь имеет возможность разместить x_i, $i = 1, \ldots, n$, по своему усмотрению. Спрашивается, как разместить x_i, чтобы получить наилучшую по точности (в том или ином смысле, который уточняется ниже) МНК-оценку? Сформулированная здесь задача и составляет основу теории планирования оптимальных экспериментов. Аналогичные задачи оптимизации экспериментов можно поставить при восстановлении линейной регрессии

$$\sum_{j=1}^{m} u_j x_j,$$

где $x_1, \ldots, x_m$ — регрессоры, и для других линейных и нелинейных задач восстановления.

13.1. Постановка задач планирования оптимальных измерений

В главе 13 мы ограничимся решением задач планирования оптимальных экспериментов при восстановлении функциональных зависимостей вида

$$y = \sum_{j=1}^{n} u_j \varphi_j(x) \equiv \varphi^T(x)\, u.$$

Пусть $K_\varepsilon = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Обозначим $W_i = 1/\sigma_i^2$, $i = 1, \ldots, n$. Тогда $K_\varepsilon^{-1} = \operatorname{diag}(W_1, \ldots, W_n)$. Числа W_i называют *весами*. Для рассматриваемого случая МНК-оценку (7.22) можно представить в виде

$$\widehat{u}_{\text{МНК}} = M^{-1} Y,$$

где

$$M = A^T K_\varepsilon^{-1} A = \sum_{i=1}^{n} W_i \varphi(x_i) \varphi^T(x_i) = \sum_{i=1}^{n} M_i,$$

$$Y = A^T K_\varepsilon^{-1} f = \sum_{i=1}^{n} W_i y_i \varphi(x_i)$$

— вектор размерности m. В принятых обозначениях ковариационная матрица МНК-оценки определяется равенством $K_{\widehat{u}} = M^{-1}$.

Напомним, что наилучшей линейной несмещенной оценкой функции $y = \varphi^T(x)\, u$ является МНК-оценка $\widehat{y}_{\text{МНК}} = \varphi^T(x) \widehat{u}_{\text{МНК}}$ с наименьшей дисперсией $\sigma^2(\widehat{y}_{\text{МНК}}) = \varphi^T(x) M^{-1} \varphi(x) \equiv d(x)$. Функцию $\sqrt{d(x)}$ называют *коридором ошибок*.

Часто в одной точке $x = x_i$ проводится несколько независимых измерений $y_{i1}, \ldots, y_{ir_i}$ функции y, каждое с дисперсией σ_i^2. Тогда всю совокупность измерений y_{ik}, $k = 1, \ldots, r_i$, можно заменить одним

$$y_i = \frac{1}{r_i} \sum_{k=1}^{r_i} y_{ik}$$

с весом $W_i = r_i/\sigma_i^2$.

В большинстве прикладных задач дисперсия σ_i^2 измерения y_i зависит от местоположения фиксации $x = x_i$. Поскольку при решении задач планирования оптимальных экспериментов при восстановлении функциональных зависимостей значения x_i априори неизвестны, а выбираются в процессе решения самой задачи оптимизации, необходимо знать значения дисперсии $\sigma^2(x)$ погрешностей измерений функции для любых значений независимой переменной x из промежутка его доступного измерения. Функцию $\lambda(x) = 1/\sigma^2(x)$ назовем *функцией эффективности измерений*. Если $\sigma^2(x)$ задана с точностью до неизвестной константы σ_0^2, т. е. $\sigma^2(x) = \sigma_0^2 \widetilde{\sigma}^2(x)$, то функцию $\widetilde{\lambda}(x) = 1/\widetilde{\sigma}^2(x)$ назовем *функцией относительной эффективности измерений*.

Очевидно, что $\widehat{u}_{\text{МНК}} = M^{-1}Y$, где

$$M = \sum_{i=1}^{n} \widetilde{\lambda}(x_i)\varphi(x_i)\varphi^T(x_i), \quad Y = \sum_{i=1}^{n} \widetilde{\lambda}(x_i)y_i\varphi(x_i).$$

Статистика

$$s_0^2 = \frac{1}{n-m}\sum_{i=1}^{n} \widetilde{\lambda}(x_i)(y_i - \varphi^T(x_i)\widehat{u}_{\text{МНК}})^2$$

является несмещенной и состоятельной оценкой неизвестной σ_0^2.

Пусть измерения проводятся при значениях $x_1, \ldots, x_n$, причем теперь в точке x_i проводится r_i измерений. Тогда $\sum\limits_{i=1}^{n} r_i = N$ — общее количество измерений.

Совокупность величин $\{x_1, \ldots, x_n;\ r_1, \ldots, r_n\}$ называют *планом* (*эксперимента*) и обозначают символом $\mathcal{E}(N)$ или $\mathcal{E}$.

Совокупность величин $\{x_1, \ldots, x_n;\ p_1, \ldots, p_n\}$, где $p_i = r_i/N$, $\sum\limits_{i=1}^{n} p_i = 1$, называют *нормированным планом* (*эксперимента*) и обозначают символом $\varepsilon(N)$.

Совокупность точек $x_1, \ldots, x_n$ называется *спектром плана эксперимента*.

Поскольку информационная матрица M зависит от плана эксперимента $\mathcal{E}$ (или ε), в дальнейшем часто будем использовать явную запись этой зависимости $M = M(\mathcal{E})$ (или $M(\varepsilon)$).

Имеем: $M(\varepsilon) = N\sum\limits_{i=1}^{n} p_i\widetilde{\lambda}(x_i)\varphi(x_i)\varphi^T(x_i)$. В дальнейшем матрицу M^{-1}, обратную информационной, будем обозначать $\mathcal{D}(\mathcal{E})$ (или $\mathcal{D}(\varepsilon)$).

Два плана

$$\mathcal{E}_1 = \{x_1^{(1)}, \ldots, x_{n_1}^{(1)};\quad r_1^{(1)}, \ldots, r_{n_1}^{(1)}\},$$

$$\mathcal{E}_2 = \{x_1^{(2)}, \ldots, x_{n_2}^{(2)};\quad r_1^{(2)}, \ldots, r_{n_2}^{(2)}\}$$

называют *одинаковыми* тогда и только тогда, когда $n_1 = n_2 = n$ и

$$x_{(i)}^{(1)} = x_{(i)}^{(2)}, \quad r_{(i)}^{(1)} = r_{(i)}^{(2)}, \quad i = 1, \ldots, n,$$

где $x_{(1)}^{(1)} < x_{(2)}^{(1)} < \ldots < x_{(n)}^{(1)}$ и $x_{(1)}^{(2)} < x_{(2)}^{(2)} < \ldots < x_{(n)}^{(2)}$ — ранжированные совокупности $x_1^{(1)}, \ldots, x_n^{(1)}, x_1^{(2)}, \ldots, x_n^{(2)}$ соответственно. В противном случае планы $\mathcal{E}_1$ и $\mathcal{E}_2$ называют *различными*.

В дальнейшем, как правило, будут рассматриваться различные планы, для которых выполнены равенства

$$n_1 = n_2 = n, \quad \sum_{i=1}^{n} r_i^{(1)} = \sum_{i=1}^{n} r_i^{(2)} = N.$$

Аналогично вводится понятие различия двух нормированных планов.

Различные планы можно сравнивать друг с другом с точки зрения качества МНК-оценок искомого вектора u, получаемых на основе каждого из этих планов. Поскольку в основу определения качества (точности) оценок могут быть положены различные критерии (в зависимости от поставленных приоритетных целей), то на практике используют различные критерии сравнения планов экспериментов.

Перечислим наиболее часто используемые критерии сравнения планов экспериментов.

Говорят, что план $\mathcal{E}_1$ *предпочтительней* (*лучше*) по выбранному критерию сравнения плана $\mathcal{E}_2$, и записывают этот факт $\mathcal{E}_1 \succ \mathcal{E}_2$, если:

$|\mathcal{D}(\mathcal{E}_1)| < |\mathcal{D}(\mathcal{E}_2)|$ — первый критерий сравнения;

$\operatorname{Sp}\mathcal{D}(\mathcal{E}_1) < \operatorname{Sp}\mathcal{D}(\mathcal{E}_2)$ — второй критерий сравнения;

$\max_{j=1,\dots,m} \mathcal{D}_{jj}(\mathcal{E}_1) < \max_{j=1,\dots,m} \mathcal{D}_{jj}(\mathcal{E}_2)$ — третий критерий сравнения;

$\max_{x\in X} d(x;\ \mathcal{E}_1) < \max_{x\in X} d(x;\ \mathcal{E}_2)$ — четвертый критерий сравнения;

$\int\limits_X d(x;\ \mathcal{E}_1)\,dx < \int\limits_X d(x;\ \mathcal{E}_2)\,dx$ — пятый критерий сравнения.

Здесь $d(x;\ \mathcal{E})$ — дисперсия $\sigma^2(\widehat{y}_{\text{МНК}}(x))$, вычисленная при реализации плана $\mathcal{E}$.

В четвертом и пятом критериях сравнения множество X задается априори и может быть частью всей области изменениям при выборе фиксаций x_i или вообще не совпадать с этой областью. В последнем случае говорят, что рассматривается *задача экстраполяции*.

В общем случае план $\mathcal{E}_1$ может быть предпочтительней плана $\mathcal{E}_2$ в смысле одного или нескольких критериев, а по остальным критериям $\mathcal{E}_2$ предпочтительней $\mathcal{E}_1$.

Пример 13.1. Пусть восстанавливается функциональная зависимость $y = u_1 + u_2x + u_3x^2$ и имеются два плана эксперимента $\mathcal{E}_1 = \{-1{,}0;\ 4, 4, 4\}$, $\mathcal{E}_2 = \{-1{,}0;\ 3, 6, 3\}$. Имеем

$$\mathcal{D}(\mathcal{E}_1) = \frac{1}{12}\begin{pmatrix} 3 & 0 & -3 \\ 0 & 3/2 & 0 \\ -3 & 0 & 9/2 \end{pmatrix}, \quad \mathcal{D}(\mathcal{E}_2) = \frac{1}{12}\begin{pmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{pmatrix},$$

$|\mathcal{D}(\mathcal{E}_1)| = 6{,}75 \cdot 12^{-3}$, $|\mathcal{D}(\mathcal{E}_2)| = 8 \cdot 12^{-3}$, $\operatorname{Sp}\mathcal{D}(\mathcal{E}_1) = 3/4$, $\operatorname{Sp}\mathcal{D}(\mathcal{E}_2) = 2/3$.

Следовательно, $\mathcal{E}_1 \succ \mathcal{E}_2$ по первому критерию, но $\mathcal{E}_2 \succ \mathcal{E}_1$ по второму критерию.

При реализации плана эксперимента $\mathcal{E}$ производятся затраты $\tau(\mathcal{E})$ на его проведение. Положительный функционал затрат $\tau(\mathcal{E})$ может иметь различный физический смысл в зависимости от условий проведения эксперимента. Например, если необходимо провести эксперимент в короткие сроки, не считаясь с материальными затратами на его проведение, в качестве $\tau(\mathcal{E})$ необходимо взять суммарное время $\tau_B(\mathcal{E})$ на проведение всех измерений, соответствующих плану эксперимента $\mathcal{E}$.

Наоборот, если более важны материальные затраты, а время проведения эксперимента не играет существенной роли, то в качестве $\tau(\mathcal{E})$ необходимо взять суммарные материальные затраты $\tau_M(\mathcal{E})$ на проведение всех измерений, соответствующих плану $\mathcal{E}$. Часто необходимо найти компромисс между временем проведения эксперимента и материальными затратами на его проведение. Тогда в качестве $\tau(\mathcal{E})$ берут сумму $k_B\tau_B(\mathcal{E}) + k_M\tau_M(\mathcal{E})$, где постоянные множители k_B, $k_M > 0$ определяют степень этого компромисса.

Одной из основных задач планирования оптимальных экспериментов является задача выбора компромисса между затратами на проведение эксперимента и погрешностью получаемой МНК-оценки искомого вектора u.

Погрешность оценки будем определять положительным функционалом $L[\mathcal{D}(\mathcal{E})]$. В соответствии с вышеприведенными критериями сравнения планов обычно используют один из нижеследующих пяти видов функционала погрешности оценки:

$$L[\mathcal{D}(\mathcal{E})] = |\mathcal{D}(\mathcal{E})|, \qquad L[\mathcal{D}(\mathcal{E})] = \operatorname{Sp}\mathcal{D}(\mathcal{E}),$$

$$L[\mathcal{D}(\mathcal{E})] = \max_{j=1,\dots,m} \mathcal{D}_{jj}(\mathcal{E}),$$

$$L[\mathcal{D}(\mathcal{E})] = \max_{x\in X} d(x;\ \mathcal{E}) = \max_{x\in X} \varphi^T(x)\mathcal{D}\varphi(x),$$

$$L[\mathcal{D}(\mathcal{E})] = \int_X d(x;\ \mathcal{E})\,dx.$$

Поставим две основные задачи планирования оптимальных экспериментов.

Задача А. Пусть затраты лимитированы. Требуется при выполнении условия лимита затрат найти план эксперимента, минимизирующий функционал погрешности МНК-оценки искомого вектора u.

Математически задача A эквивалентна нахождению оптимального плана $\mathcal{E}^*$, являющегося решением экстремальной задачи с ограничением

$$\mathcal{E}^* = \arg \min_{\tau(\mathcal{E})\leqslant T} L[\mathcal{D}(\mathcal{E})], \tag{13.1}$$

где T задано.

Задача Б. Пусть задана необходимая точность МНК-оценки искомого вектора u. Требуется при выполнении условия достижения необходимой точности оценки найти план эксперимента, минимизирующий затраты на проведение эксперимента.

Математически задача Б эквивалентна нахождению плана $\mathcal{E}^*$, являющегося решением экстремальной задачи с ограничением

$$\mathcal{E}^* = \arg \min_{L[\mathcal{D}(\mathcal{E})] \leqslant L_0} \tau(\mathcal{E}), \tag{13.2}$$

где L_0 задано.

Функционал затрат часто представим в виде $\tau(\mathcal{E}) = \sum_{i=1}^{r} c_i r_i$, где $c_i > 0$ — стоимость одного измерения в точке x_i. В частности, если стоимость измерений не зависит от x_i, то $c_i = c$, $i = 1, \ldots, n$, и $\tau(\mathcal{E}) = cN$.

Экстремальные задачи с ограничением (13.1), (13.2) эквивалентны экстремальной задаче без ограничения

$$\mathcal{E}^* = \arg \min R(\mathcal{E}), \tag{13.3}$$

где $R(\mathcal{E}) = \tau(\mathcal{E}) + k_L L[\mathcal{D}(\mathcal{E})]$ называется *функционалом потерь*, $k_L > 0$ — множитель Лагранжа.

Для задачи А величина k_L находится из равенства $\tau(\mathcal{E}^*) = T$. Для задачи Б величина k_L находится из равенства $L[\mathcal{D}(\mathcal{E})] = L_0$.

Если N фиксировано и $\tau(\mathcal{E}) = cN$, то $\tau(\mathcal{E})$ не зависит от $\mathcal{E}$. Поэтому для данного случая задача (13.1) эквивалентна задаче без ограничения

$$\mathcal{E}^* = \arg \min L[\mathcal{D}(\mathcal{E})]. \tag{13.4}$$

План $\mathcal{E}^*$ из (13.4) называется *$\mathcal{D}$-оптимальным*, если $L[\mathcal{D}(\mathcal{E})] = |\mathcal{D}(\mathcal{E})|$; *A-оптимальным*, если $L[\mathcal{D}(\mathcal{E})] = \operatorname{Sp}\mathcal{D}(\mathcal{E})$; *минимаксным в пространстве параметров*, если $L[\mathcal{D}(\mathcal{E})] = \max_{j=1,\ldots,m} \mathcal{D}_{jj}(\mathcal{E})$; *минимаксным на множестве X*, если $L[\mathcal{D}(\mathcal{E})] = \max_{x \in X} d(x;\ \mathcal{E})$; *Q-оптимальным на множестве X*, если $L[\mathcal{D}(\mathcal{E})] = \int\limits_X d(x;\ \mathcal{E})\, dx$.

Заметим, что поскольку $|M(\mathcal{E})| = 1/|\mathcal{D}(\mathcal{E})|$, то $\mathcal{D}^*$ — оптимальный план $\mathcal{E}^*$ — можно определить равенством $\mathcal{E}^* = \arg\max |M(\mathcal{E})|$.

Пример 13.2. Пусть $y = u_1 + u_2 x$, $x \in [-1;\ 1]$, $N = 3$, $\lambda(x) = \text{const}$, $x_1 \leqslant x_2 \leqslant x_3$. Имеем

$$M(\mathcal{E}) = \frac{1}{3} \begin{pmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{pmatrix},$$

$$|M(\mathcal{E})| = x_1^2 + x_2^2 + x_3^2 - \frac{1}{3}(x_1 + x_2 + x_3)^2.$$

Существует два $\mathcal{D}$-оптимальных плана $\mathcal{E}_1^*$, $\mathcal{E}_2^*$, максимизирующих $|M(\mathcal{E})|$ при ограничениях $-1 \leqslant x_i \leqslant 1$, $i = 1, 2, 3$, $\mathcal{E}_1^* = \{-1{,}1;\ 2{,}1\}$ и $\mathcal{E}_2^* = \{-1{,}1;\ 1{,}2\}$.

Если априори задана приемлемая погрешность оценки L_0, а N не фиксируется, то решение задачи (13.4) состоит в нахождении такого N^*

и оптимальных планов $\mathcal{E}^*(N^*)$, $\mathcal{E}^*(N^*+1)$, при которых выполнены неравенства $L[\mathcal{D}(\mathcal{E}^*(N^*))] > L_0 \geqslant L[\mathcal{D}(\mathcal{E}^*(N^*+1))]$. Тогда эксперимент проводится по плану $\mathcal{E}^*(N^*+1)$.

Наиболее полные результаты по оптимальному планированию экспериментов удается получить в том случае, когда N настолько велико, что можно считать функционал потерь функцией непрерывного параметра N.

Непрерывным нормированным планом ε называется совокупность величин $\{x_1, \ldots, x_n;\ p_1, \ldots, p_n\}$, где $0 \leqslant p_i \leqslant 1$, $\sum\limits_{i=1}^{n} p_i = 1$.

Если $N \gg n$, то $\varepsilon(N) = \varepsilon$, поэтому

$$M(\mathcal{E}) = N \sum_{i=1}^{n} p_i \widetilde{\lambda}(x_i)\varphi(x_i)\varphi^T(x_i) \approx NM(\mathcal{E}).$$

Следовательно,

$$\mathcal{D}(\mathcal{E}) = M^{-1}(\mathcal{E}) \approx \frac{1}{N}\mathcal{D}(\varepsilon).$$

Имеем:

1° $|\mathcal{D}(\mathcal{E})| \approx \dfrac{1}{N^m}\,|\mathcal{D}(\varepsilon)|$;

2° $\operatorname{Sp}\mathcal{D}(\mathcal{E}) \approx \dfrac{1}{N}\operatorname{Sp}\mathcal{D}(\varepsilon)$;

3° $\max_{j=1,\ldots,m} \mathcal{D}_{jj}(\mathcal{E}) \approx \dfrac{1}{N}\max_{j=1,\ldots,m} \mathcal{D}_{jj}(\varepsilon)$;

4° $\max_{x\in X} d(x;\ \mathcal{E}) \approx \dfrac{1}{N}\max_{x\in X} d(x;\ \varepsilon)$;

5° $\int\limits_X d(x;\ \mathcal{E})\,dx \approx \dfrac{1}{N}\int\limits_X d(x;\ \varepsilon)\,dx$.

Таким образом, для случая $\tau(\mathcal{E}) = cN$, $N \gg n$ функционал потерь $R(\mathcal{E})$ имеет вид

$$R(\mathcal{E}) \approx Nc + \frac{k_L}{N^\beta}\,L[\mathcal{D}(\varepsilon)], \tag{13.5}$$

где $\beta = m$ для критерия $L[\mathcal{D}(\varepsilon)] = |\mathcal{D}(\varepsilon)|$ и $\beta = 1$ для остальных четырех критериев.

Поскольку $L[\mathcal{D}(\varepsilon)]$ не зависит от N, решение экстремальной задачи (13.3), когда $R(\mathcal{E})$ дается равенством (13.5), можно разбить на два этапа: на первом этапе находится оптимальный непрерывный нормированный план $\varepsilon^* = \arg\min_\varepsilon L[\mathcal{D}(\varepsilon)] = \{x_1^*, \ldots, x_n^*;\ p_1^*, \ldots, p_n^*\}$; на втором этапе при фиксированном $\varepsilon = \varepsilon^*$ находится N^*, при котором достигается минимум $Nc + k_L N^{-\beta} L[\mathcal{D}(\varepsilon^*)]$, из соответствующего условия находится k_L и полагается $\mathcal{E}^* = \{x_1^*, \ldots, x_n^*;\ r_1^*, \ldots, r_n^*\}$, где $r_i^* = [p_i^* N^*]$ — округление до ближайшего целого.

Очевидно, $N^* = [(\beta k_L c^{-1} L[\mathcal{D}(\varepsilon^*)])^{1/(\beta+1)}]$ — округление до ближайшего целого.

13.2. Методы решения задач планирования оптимальных измерений

В дальнейшем рассматриваются только непрерывные нормированные планы ε. Заметим, что информационная матрица $M(\varepsilon)$ вырождена, если спектр плана ε содержит менее m точек. Поэтому количество точек n спектра любого разумного плана ε должно состоять не менее чем из m точек.

Естественно конструировать планы с наименьшим числом точек спектра.

Теорема 13.1 (Де Ла Гарза). *Для любого плана ε с числом точек спектра $n \geqslant m$ всегда найдется такой план $\widetilde{\varepsilon}$ с числом точек спектра $m+1$, что $M(\widetilde{\varepsilon}) = M(\varepsilon)$.*

Из теоремы 13.1 следует, в частности, что оптимальные планы ε^* можно искать в классе планов ε с числом точек спектра, равным $m+1$. Ниже будет показано, что для определенного класса задач число точек спектра планов, среди которых существуют оптимальные, можно понизить до m — минимально возможного.

Непрерывный нормированный план ε называется *невырожденным*, если матрица плана $M(\varepsilon)$ невырожденная.

Теорема 13.2. *Пусть $\varepsilon = \{x_1, \ldots, x_n;\ p_1, \ldots, p_n\}$ — невырожденный план. Тогда* 1° $\sum\limits_{i=1}^{n} p_i \widetilde{\lambda}(x_i) d(x_i;\ \varepsilon) = m$; 2° $\max_{x \in X} \widetilde{\lambda}(x) d(x;\ \varepsilon) \geqslant m$, *где X — вся область изменения независимой переменной x, где фиксируется спектр планов.*

Наиболее широко используют $\mathcal{D}$-оптимальность. $\mathcal{D}$-оптимальные непрерывные планы обладают рядом замечательных свойств.

Теорема 13.3 (Вольфовиц–Кифер). *Следующие утверждения эквивалентны: ε^* — $\mathcal{D}$ - оптимальный план; ε^* минимизирует* $\max_{x \in X} \widetilde{\lambda}(x) d(x;\ \varepsilon)$; $max_{x \in X} \widetilde{\lambda}(x) d(x;\ \varepsilon) = m$.

Теорему 13.3 принято называть *теоремой эквивалентности*. Из теоремы эквивалентности, в частности, следует, что при $\widetilde{\lambda}(x) = \text{const}$ $\mathcal{D}$-оптимальный непрерывный план одновременно является минимаксным непрерывным планом, и наоборот.

Теорема 13.4. *Пусть $z = Z(x)$ — взаимно однозначное преобразование и $\varepsilon_z^* = \{z_1, \ldots, z_n;\ p_1, \ldots, p_n\}$ — $\mathcal{D}$-оптимальный непрерывный план восстановления функциональной зависимости, $y = f^T(z)u$, где $f(z) \equiv \varphi(Z^{-1}(z))$. Тогда непрерывный план $\varepsilon_x^* = \{x_1, \ldots, x_n;\ p_1, \ldots, p_n\}$, где $x_i = Z^{-1}(z_i)$, $i = 1, \ldots, n$, $\mathcal{D}$-оптимален при восстановлении функциональной зависимости $y = \varphi^T(x)\, u$.*

Теорема 13.4 часто дает возможность с помощью замены независимой переменной переходить от сложных вектор-функций $\varphi(x)$ к более простым $f(x)$ и искать $\mathcal{D}$-оптимальные планы относительно новой независимой переменной z.

Пример 13.3. Пусть $y = u_1 + u_2 e^{-\alpha x}$, $x \geqslant 0$, $\widetilde{\lambda}(x) = e^{-\alpha x}$, $\alpha > 0$. Найдем $\mathcal{D}$-оптимальный план ε_x^*. Сделаем замену переменной $z = e^{-\alpha x}$. Имеем: $y = u_1 + u_2 x$, $\widetilde{\lambda}(z) = z$, $0 \leqslant z \leqslant 1$. Рассмотрим класс планов вида $\varepsilon_z = \{1, z_0;\ 1-p, p\}$, где $0 < z_0 \leqslant 1$, $0 < p < 1$. Имеем

$$M(\varepsilon_z) = \begin{pmatrix} (1-p) + z_0 p & (1-p) + z_0^2 p \\ (1-p) + z_0^2 p & (1-p) + z_0^3 p \end{pmatrix},$$

$$|M(\varepsilon_z)| = p(1-p)(z_0^3 - 2z_0^2 + z_0).$$

Следовательно, максимум $|M(\varepsilon_z)|$ достигается при $p = 1/2$, $z_0 = 1/3$. Положим $\varepsilon_z^* = \{1{,}1/3;\ 1/2{,}1/2\}$. Имеем: $\max \widetilde{\lambda}(z) d(z;\ \varepsilon_z^*) = 2$. Из теоремы эквивалентности следует, что ε_z^* — $\mathcal{D}$-оптимальный непрерывный план для $y = u_1 + u_2 z$, $\widetilde{\lambda}(z) = z$, $0 < z \leqslant 1$. Тогда из теоремы 13.4 следует, что $\varepsilon_x^* = \{0, \frac{1}{\alpha} \ln 3;\ 1/2{,}1/2\}$ — $\mathcal{D}$-оптимальный непрерывный план для восстановления функциональной зависимости $y = u_1 + u_1 e^{-\alpha x}$, $\widetilde{\lambda}(x) = e^{-\alpha x}$, $x \geqslant 0$. Таким образом, если половину измерений делать при $x = 0$, а вторую половину — при $x = \frac{1}{\alpha} \ln 3$, то план $\mathcal{E}$ будет $\mathcal{D}$-оптимальным.

Поиск $\mathcal{D}$-оптимального непрерывного плана состоит в максимизации определителя $|M(\varepsilon)| = |\sum_{i=1}^{n} p_i \widetilde{\lambda}(x_i) \varphi(x_i) \varphi^T(x_i)|$ по переменным $x_1, \ldots, x_n, p_1, \ldots, p_n$, $\sum_{i=1}^{n} p_i = 1$, $1 \geqslant p_i \geqslant 0$.

Можно показать, что при $n = m$ справедливо равенство $|M(\varepsilon)| = \prod_{i=1}^{m} p_i \widetilde{\lambda}(x_i) |A|^2$, где $A = (\varphi(x_1), \ldots, \varphi(x_m))$ — $(m \times m)$-матрица. Из последнего равенства следует, что максимальное значение $|M(\varepsilon)|$ по $p_1, \ldots, p_m$ достигается при $p_i = 1/m$, $i = 1, \ldots, m$, независимо от $x_1, \ldots, x_m$. Таким образом, если спектр $\mathcal{D}$-оптимального плана состоит из m точек, то количество измерений должно быть размещено равномерно по точкам спектра, а задача поиска $\mathcal{D}$-оптимального плана сводится к максимизации по $x_1, \ldots, x_m$ функционала $|A|^2 \prod_{i=1}^{m} \widetilde{\lambda}(x_i)$.

Пусть $\varphi_j(x) = x^{j-1}$, $j = 1, \ldots, m$, т. е. рассматривается задача восстановления полиномиальной функциональной зависимости $y = \sum_{j=1}^{m} u_j x^{j-1}$.

Чтобы для полиномиальной зависимости существовал $\mathcal{D}$-оптимальный план с минимально возможным количеством точек спектра $n = m$, необходимо наложить некоторые условия на функцию относительной эффективности измерений $\widetilde{\lambda}(x)$.

Система непрерывных функций $f_i(x)$, $i = 1, \ldots, m$, определенная на промежутке X, называется *чебышевской*, если любая линейная комбинация этих функций $\sum\limits_{i=1}^{m} c_i f_i(x)$ (c_i действительные, $\sum\limits_{i=1}^{m} c_i > 0$) имеет не более $m-1$ корней на промежутке X.

Очевидно, что система $f_i(x) = x^{i-1}$, $i = 1, \ldots, m$, — чебышевская.

Теорема 13.5. *Пусть* $1, \widetilde{\lambda}(x), \ldots, \widetilde{\lambda}(x)x^{2(m-1)}$ *— чебышевская система на промежутке* X. *Тогда при восстановлении полиномиальной зависимости* $y = \sum\limits_{j=1}^{m} u_j x^{j-1}$ *существует* $\mathcal{D}$*-оптимальный план с числом точек спектра* $n = m$.

Рассмотрим в качестве промежутка X один из трех стандартных промежутков $[-1, 1]$, $[0, \infty)$, $(-\infty, \infty)$. С помощью линейной замены независимой переменной любой промежуток можно свести к одному из этих трех, а при линейной замене полиномиальная зависимость остается полиномиальной (с сохранением степени полинома).

Теорема 13.6. *Пусть рассматривается задача восстановления полиномиальной зависимости* $y = \sum\limits_{j=1}^{m} u_j x^{j-1}$ *для следующих четырех случаев*:

1° $\widetilde{\lambda}(x) = \text{const}$, $x \in [-1, 1]$;

2° $\widetilde{\lambda}(x) = (1-x)^{\alpha}(1+x)^{\beta}$, $\alpha, \beta > -1$, $x \in [-1, 1]$;

3° $\widetilde{\lambda}(x) = x^{\alpha+1}e^{-x}$, $\alpha > -1$, $x \geqslant 0$;

4° $\widetilde{\lambda}(x) = e^{-x^2}$, $x \in (-\infty; \infty)$.

Тогда для каждого из этих четырех случаев существует $\mathcal{D}$*-оптимальный непрерывный план* $\varepsilon^* = \{x_1, \ldots, x_m;\ 1/m, \ldots, 1/m\}$, *единственный в классе планов с* $n = m$. *Точки спектра* $\mathcal{D}$*-оптимальных планов являются корнями полиномов для случаев*: 1. $(1-x^2)\dfrac{dP_{m-1}}{dx}$, *где* P_{m-1} *— полином Лежандра степени* $m-1$; 2. $P_m^{(\alpha,\beta)}(x)$ *— полином Якоби степени* m *с параметрами* α, β; 3. $L_m^{(\alpha)}(x)$ *— полином Чебышева–Лагерра с параметром* α; 4. $H_m(x)$ *— полином Чебышева–Эрмита.*

Доказательство всех четырех случаев теоремы 13.6 аналогично. Рассмотрим, например, первый случай.

Имеем: $|M(\varepsilon)| = \dfrac{1}{m}\, C|A|^2$, где $|A|^2 = \prod\limits_{k \neq l,\ k,l=1}^{m} (x_k - x_l)^2$, $\ln |A|^2 = \sum\limits_{k \neq l,\ k,l=1}^{m} \ln (x_k - x_l)^2$. Поэтому задача максимизации $|M(\varepsilon)|$ эквивалентна задаче максимизации $\sum\limits_{k \neq l,\ k,l=1}^{m} \ln (x_k - x_l)^2$. Для нахождения

точек спектра имеем систему уравнений $\dfrac{\partial \ln |A|^2}{\partial x_k}$, $k = 1, \ldots, m$, которую запишем в виде

$$\frac{1}{x_k - x_1} + \ldots + \frac{1}{x_k - x_{k-1}} + \frac{1}{x_k - x_{k+1}} + \ldots + \frac{1}{x_k - x_m} = 0,$$

$$k = 1, \ldots, m.$$

Обозначим $f(x) = (x - x_1) \cdot \ldots \cdot (x - x_m)$. Имеем

$$\frac{df(x)}{dx} = (x - x_2)(x - x_3) \cdot \ldots \cdot (x - x_m) + \ldots$$

$$\ldots + (x - x_1)(x - x_2) \cdot \ldots \cdot (x - x_{m-1}),$$

$$\frac{df(x_k)}{dx} = (x_k - x_1) \cdot \ldots \cdot (x_k - x_{k-1})(x_k - x_{k+1}) \cdot \ldots \cdot (x_k - x_m),$$

$$\frac{d^2 f(x)}{dx^2} = (x - x_3) \cdot \ldots \cdot (x - x_m) + \ldots + (x - x_1) \cdot \ldots \cdot (x - x_{m-2}),$$

$$\frac{d^2 f(x_k)}{dx^2} = (x_k - x_2) \cdot \ldots \cdot (x_k - x_{k-1})(x_k - x_{k+1}) \cdot \ldots \cdot (x_k - x_m) +$$

$$+ (x_k - x_1)(x_k - x_3) \cdot \ldots \cdot (x_k - x_{k-1})(x_k - x_{k+1}) \cdot \ldots$$

$$\ldots \cdot (x_k - x_m) + \ldots + (x_k - x_1)(x_k - x_2) \cdot \ldots$$

$$\ldots \cdot (x_k - x_{k-1})(x_k - x_{k+1}) \cdot \ldots \cdot (x_k - x_{m-1}).$$

Поэтому

$$\frac{d^2 f(x_k)}{dx^2} \Big/ \frac{df(x_k)}{dx} = \frac{1}{x_k - x_1} + \frac{1}{x_k - x_2} + \ldots + \frac{1}{x_k - x_{k-1}} +$$

$$+ \frac{1}{x_k - x_{k+1}} + \ldots + \frac{1}{x_k - x_m} = 0.$$

Из последнего равенства следует, что $\dfrac{d^2 f(x)}{dx^2}(1 - x^2) = af(x)$, где a — постоянная. Приравнивая в обеих частях последнего равенства коэффициенты при x^m, получаем $a = -m(m+1)/2$. Следовательно,

$$(1 - x^2)\frac{d^2 f(x)}{dx^2} + \frac{m(m+1)}{2} f = 0.$$

Но решением последнего уравнения является функция

$$f(x) = (1 - x^2)\frac{dP_{m-1}}{dx}.$$

Теорема 13.6 для рассматриваемого случая доказана.

При решении практических задач часто возникает необходимость восстановления тригонометрической функциональной зависимости

$$y = u_1 + \sum_{k=1}^{m} (u_{k+1} \cos kx + u_{m+1+k} \sin kx),$$

где u_j, $j = 1, \ldots, 2m+1$, неизвестны.

Пусть $X = [0;\ 2\pi]$, $\widetilde{\lambda}(x) = \text{const}$. Тогда любой непрерывный план с $n \geqslant 2m+1$, с равномерным спектром и равными весами $p_i = 1/n$, $i = 1, \ldots, n$, является $\mathcal{D}$-оптимальным. Более того, любой непрерывный план, $\mathcal{D}$-оптимальный для тригонометрической зависимости с $m = m_0$, будет $\mathcal{D}$-оптимальным для тригонометрической зависимости с $m < m_0$.

Выше уже отмечалось, что при $\widetilde{\lambda}(x) = \text{const}$ $\mathcal{D}$-оптимальный план совпадает с минимаксным планом. Ниже приведены условия для совпадения планов, оптимальных по разным критериям.

Теорема 13.7. *План ε^* является одновременно Q- и A-оптимальным, если $\varphi_i(x)$, $i = 1, \ldots, m$, — ортонормированная система на промежутке X, т. е.* $\int\limits_X \varphi(x)\varphi^T(x)\,dx = I_m$.

Для полиномиальной зависимости с $\widetilde{\lambda}(x) = \text{const}$ доказано, что спектры $\mathcal{D}$- и Q-оптимальных планов совпадают ($n = m$), т. е. измерения необходимо проводить в нулях полинома $(1-x^2)\dfrac{dP_{m-1}}{dx}$, но веса для Q-оптимальных планов определяются равенствами

$$p_i = |P_m(x_i)|^{-1} \Big/ \sum_{j=1}^{m} |P_m(x_j)|^{-1}, \quad j = 1, \ldots, m,$$

$P_m(x)$ — полином Лежандра.

Для тригонометрической зависимости $\mathcal{D}$- и A-оптимальные планы совпадают.

Из вышесказанного следует, что нахождение точного $\mathcal{D}$-оптимального непрерывного плана аналитическими методами возможно только для специальных систем функций $\varphi_j(x)$, $j = 1, \ldots, m$, $\lambda(x)$.

В общем случае, когда найти точный $\mathcal{D}$-оптимальный непрерывный план аналитическими методами не удается, его можно найти численными методами. Ниже приводится итерационный процесс, сходящийся к $\mathcal{D}$-оптимальному плану. Основой ИП является теорема эквивалентности (см. выше).

Зададим невырожденный план $\varepsilon_0 = \{x_1, \ldots, x_n;\ p_1, \ldots, p_n\}$, $n \geqslant m$. На каждом шаге ИП осуществляются следующие три этапа.

1. Подсчитываются информационная матрица

$$M(\varepsilon_0) = \sum_{i=1}^{n} p_i \widetilde{\lambda}(x_i)\varphi(x_i)\varphi^T(x_i)$$

и матрица $\mathcal{D}(\varepsilon_0) = M^{-1}(\varepsilon_0)$.

2. Находится точка

$$x_0 = \arg \max_{x \in X} \widetilde{\lambda}(x)\, d(x;\, \varepsilon_0),$$

где $d(x;\, \varepsilon_0) = \varphi^T(x)\mathcal{D}(\varepsilon_0)\varphi(x)$;

3. Формируется новый план

$$\varepsilon_1 = (1 - \alpha_0)\varepsilon_0 + \alpha_0 \varepsilon(x_0),$$

где

$$\varepsilon(x_0) = \{x_0;\, 1\},$$

$$\varepsilon_1 = \{x_1, \ldots, x_n, x_0;\, (1 - \alpha_0)p_1, \ldots, (1 - \alpha_0)p_n, \alpha_0\},$$

а «шаг» α_0 выбирается так, чтобы увеличение $|M(\varepsilon_1)|$ было максимальным.

Для такого α_0 справедливо равенство $\alpha_0 = \delta_0/((\delta_0 + m - 1)m)$, где $\delta_0 = \widetilde{\lambda}(x_0)d(x_0;\, \varepsilon_0) - m > 0$. После осуществления третьего этапа α_0 заменяют на α_1 и возвращаются к первому этапу.

Доказано, что последовательность планов ε_k сходится к $\mathcal{D}$-оптимальному плану ε^*.

Правилом останова ИП может быть выполнение неравенства $\alpha_k < \mu$, где $\mu > 0$ — заданное малое число.

После осуществления k_0 шагов ИП к первоначальным n точкам спектра добавляются k_0 новых точек. Чтобы количество точек спектра не возрастало, используют два правила: 1° точки спектра с малыми весами, изолированные от других точек спектра, отбрасываются (с равномерным распределением веса отбрасываемых точек по множеству оставшихся, так чтобы выполнялось равенство $\sum_{i=1}^{n_k} p_i = 1$, где n_k — количество остающихся точек спектра); 2° точки спектра $x_{l1}, \ldots, x_{lr}$, близкие друг к другу, объединяются в одну точку по правилу

$$x_l = \sum_{i=1}^{r} x_{li}\, \frac{p_{li}}{p_l}, \quad p_l = \sum_{i=1}^{r} p_{li}.$$

Более подробные сведения о решении задач планирования оптимальных экспериментов можно найти в [29, 78].

Список литературы

1. *Айвазян С.А.* Основы эконометрики. — М.: ЮНИТИ, 2001.
2. *Айвазян С.А.*, *Енюков И.С.*, *Мешалкин Л.Д.* Прикладная статистика: исследование зависимостей. Справочное издание под ред. С.А. Айвазяна. — М.: Финансы и статистика, 1985.
3. *Айвазян С.А.*, *Енюков И.С.*, *Мешалкин Л.Д.* Прикладная статистика: основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983.
4. *Айвазян С.А.*, *Мхитарян В.С.* Прикладная статистика в задачах и упражнениях. — М.: ЮНИТИ, 2001.
5. *Айвазян С.А.*, *Мхитарян В.С.* Теория вероятностей и математическая статистика, 2-е изд. — М.: ЮНИТИ-ДАНА, 2001.
6. Алгоритмы и программы восстановления зависимостей. Под ред. В.Н. Вапника. — М.: Наука, 1984.
7. *Арсенин В.Я.*, *Крянев А.В.* Применение статистических методов решения некорректных задач для обработки результатов физических экспериментов. В кн.: Автоматизация научных исследований в экспериментальной физике. — М.: Энергоатомиздат, 1987, С. 19–30.
8. *Арсенин В.Я̇*, *Крянев А.В.*, *Цупко-Ситников М.В.* Применение робастных методов при решении некорректных задач // ЖВММФ. 1989. Т. 29. № 5. С. 653–661.
9. *Арсенин В.Я.*, *Крянев А.В.* Обобщенный метод максимального правдоподобия решения конечномерных некорректных задач // ЖВММФ. 1991. Т. 31. № 5. С. 643–653.
10. *Баласанов Ю.Г.*, *Дойников А.Н.*, *Королев М.Ф.*, *Юровский А.Ю.* Прикладной анализ временных рядов с программой ЭВРИСТА. — М.: Центр СП «Диалог» МГУ, 1991.
11. *Бард Я.* Нелинейное оценивание параметров. — М.: Статистика, 1979.
12. *Белонин М.Д.*, *Татаринов И.В.*, *Калинин О.М.*, *Шиманский В.К.*, *Бескровная О.В.*, *Гранский В.В.*, *Похитонова Т.Е.* Факторный анализ в нефтяной геологии. — М.: ВИЭМС, 1971.
13. *Богданов Ю.И.* Основная задача статистического анализа данных: корневой подход. — М.: Изд-во МИЭТ, 2002.
14. *Большев Л.Н.*, *Смирнов Н.В.* Таблицы математической статистики. — М.: Наука, 1983.

15. *Вапник В.Н.* Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979.

16. *Вапник В.Н., Стефанюк А.Р.* Непараметрические методы восстановления плотности вероятностей. — АиТ. 1978. Т. 8. С. 38–52.

17. *Васильев Ф.П.* Методы оптимизации. — М.: Факториал Пресс, 2002.

18. *Васильев Ф.П.* Методы решения экстремальных задач. — М.: Наука, 1981.

19. Вероятность и математическая статистика. Энциклопедия. Гл. ред. Ю.В. Прохоров. — М.: Изд-во Большая Российская Энциклопедия, 1999.

20. *Воеводин В.В.* Вычислительные основы линейной алгебры. — М.: Наука, 1977.

21. *Волков Н.Г., Дорогов В.И., Коньшин В.А., Крянев А.В.* Комплексная программа восстановления регрессии и проведение оценки σ_f, σ_t, α и ν_f для U^{235}. — Минск, Препринт ИЯЭ АН БССР, 1986.

22. *Воскобойников Ю . Е., Преображенский Н.Г., Седельников А.И.* Математическая обработка эксперимента в молекулярной газодинамике. — Новосибирск: Наука, 1984.

23. *Гантмахер Ф.Р.* Теория матриц. — М.: Наука, 1967.

24. Главные компоненты временных рядов: метод «Гусеница». В сб. под ред. Д. Данилова и А.А. Жиглявского. — СПб.: СПбГУ, 1997.

25. *Девroй Л., Дьерфи Л.* Непараметрическое оценивание плотности. L_1-подход. — М.: Мир, 1988.

26. *Де Гроот М.* Оптимальные статистические решения. — М.: Мир, 1974.

27. *Демиденко Е.З.* Линейная и нелинейная регрессии. — М.: Финансы и статистика, 1981.

28. *Джонсон Н., Лион Ф.* Статистика и планирование эксперимента в технике и науке. — М.: Мир, 1980, Т. 1; 1981, Т. 2.

29. *Ермаков С.М., Жиглявский А.А.* Математическая теория оптимального эксперимента. — М.: Наука, 1987.

30. *Жуковский Е.Л., Мелешко В.И.* О помехоустойчивых решениях в линейной условной алгебре // ДАН СССР. 1978. Т. 241. № 5. С. 1009–1012.

31. *Жуковский Е.Л., Морозов В.А.* О последовательной байесовской регуляризации алгебраических систем уравнений // ЖВММФ. 1972. Т. 12. № 2. С. 464–465.

32. *Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л.* Методы сплайн-функций. — М.: Наука, 1980.

33. *Закс Ш.* Теория статистических выводов. — М.: Мир, 1975.

34. *Зрелов П.В., Иванов В.В.* Интегральные непараметрические статистики $\omega_n^k = n^{\frac{k}{2}} \int\limits_{-\infty}^{\infty} |S_n(x) - P(x)|^k dx$ и их основные свойства. Алгебраический вид, функции распределения и критерии согласия. — Сообщения ОИЯТ, Р10-92-461, Дубна, 1992.
35. *Иванов В.К., Васин В.В., Танана В.П.* Теория линейных некорректных задач и ее приложения. — М.: Наука, 1978.
36. *Калиткин Н.Н.* Численные методы. — М.: Наука, 1978.
37. *Кендалл М.* Временные ряды. — М.: Финансы и статистика, 1981.
38. *Кендалл М. Дж., Стюарт А.* Статистические выводы и связи. — М.: Наука, 1973.
39. *Кислицын М.М.* Многомерная статистика временных рядов наблюдений в авиационной эргономике. — Вопросы кибернетики, 1978, т. 51.
40. *Крамер Г.* Математические методы статистики. — М.: Мир, 1975.
41. *Крянев А.В.* Статистическая форма регуляризованного метода наименьших квадратов А.Н. Тихонова. — Докл. АН СССР. 1986. Т. 291. № 4. С. 780–785.
42. *Крянев А.В.* Применение современных методов параметрической и непараметрической статистики при обработке данных экспериментов на ЭВМ. — М.: МИФИ, 1987.
43. *Крянев А.В.* Применение современных методов математической статистики при восстановлении регрессионных зависимостей на ЭВМ. — М.: МИФИ, 1988.
44. *Крянев А.В., Лукин Г.В.* О постановке и решении задач оптимизации инвестиционных портфелей. Препринт 006–2001. — М.: МИФИ, 2001.
45. *Крянев А.В., Черный А.И.* Робастные линейные сглаживающие сплайны и их применения. Препринт 006–97. — М.: МИФИ, 1997.
46. *Лаврентьев М.М., Романов В.Г., Шишатский С.П.* Некорректные задачи математической физики и анализа. — М.: Наука, 1980.
47. *Лаврентьев М.М., Федотов А.М.* О постановке некоторых некорректных задач математической физики со случайными исходными данными // ЖВММФ. 1982. Т. 22. № 1. С. 133–143.
48. *Леман Э.* Теория точечного оценивания. — М.: Наука, 1991.
49. *Лисковец Д.А.* Вариационные методы решения неустойчивых задач. — Минск: Наука и техника, 1981.
50. *Лукашин Ю.П.* Адаптивные методы краткосрочного прогнозирования. — М.: Статистика, 1979.
51. *Марчук Г.И.* Методы вычислительной математики. — М.: Наука, 1989.
52. *Марчук Г.И., Атанбаев С.А.* Некоторые вопросы глобальной регуляризации // ДАН СССР. 1970. Т. 190. № 3.

53. *Марчук Г.И.*, *Васильев В.Г.* О приближенном решении операторных уравнений первого рода // ДАН СССР. 1970. Т. 199. № 4.

54. *Морозов В.А.* Регулярные методы решения некорректных задач. — М.: Наука, 1987.

55. *Муравьева М.В.* Об оптимальных и предельных свойствах Байесовского решения системы линейных алгебраических уравнений // ЖВММФ. 1973. Т.13. № 7. С. 819–828.

56. *Мэйндоналд Дж.* Вычислительные алгоритмы в прикладной статистике. — М.: Финансы и статистика, 1988.

57. *Надарая Э.А.* Непараметрическое оценивание плотности вероятностей и кривой регрессии. — Тбилиси: ТГУ, 1983.

58. *Пергамент А.Х.* Регуляризованные алгоритмы статистического оценивания функций при наличии корреляционной помехи / Препринт ИПМ им. М.В. Келдыша АН СССР, № 21, 1986.

59. *Пергамент А.Х.*, *Тельковская О.В.* Метод регуляризации и метод максимального правдоподобия при решении интегральных уравнений 1-го рода / Препринт ИПМ им. М.В. Келдыша АН СССР, № 111, 1979.

60. Пределы предсказуемости / Под ред. Ю.А. Кравцова. — М.: ЦентрКом, 1997.

61. *Пригожин И.Р.* Конец определенности: Время, хаос и новые законы природы. — Ижевск: ИРТ, 1999.

62. *Пугачев В.С.* Теория вероятностей и математическая статистика. — М.: Физматлит, 2002.

63. *Пугачев В.С.*, *Синицын И.Н.* Теория стохастических систем. — М.: Логос, 2000.

64. *Пытьев Ю.П.* Математические методы интерпретации эксперимента. — М.: Высшая школа, 1989.

65. *Рао С.Р.* Линейные стохастические методы и их применения. — М.: Наука, 1968.

66. *Рунион Р.* Справочник по непараметрической статистике. Современный подход. — М.: Финансы и статистика, 1982.

67. *Савелова Т.И.* Примеры решения некорректно поставленных задач. — М.: МИФИ, 1999.

68. *Себер Дж.* Линейный регрессионный анализ. — М.: Мир, 1980.

69. *Смоляк С.А.*, *Титаренко Б.П.* Устойчивые методы оценивания. — М.: Статистика, 1980.

70. *Соболь И.М.* Численные методы Монте-Карло. — М.: Наука, 1973.

71. *Тихонов А.Н.*, *Арсенин В.Я.* Методы решения некорректных задач. — М.: Наука, 1979.

72. *Тихонов А.Н.*, *Гончарский А.В.*, *Степанов В.В.*, *Ягола А.Г.* Регуляризующие алгоритмы и априорная информация. — М.: Наука, 1985.

73. *Тихонов А.Н.*, *Леонов А.С.*, *Ягола А.Г.* Нелинейные некорректно поставленные задачи. — М.: Наука, 1995.

74. *Турчин В.Ф.*, *Козлов В.П.*, *Малкевич М.С.* Использование методов математической статистики при решении некорректных задач // Успехи физ. наук. 1970. Т. 102. № 3. С. 345–386.

75. *Тюрин Ю.Н.*, *Макаров А.А.* Статистический анализ данных на компьютере / Под. ред. Фигурнова В.Э. — М.: ИНФРА-М, 1998.

76. *Федер Е.* Фракталы. — М.: Мир, 1991.

77. *Федоренко Р.П.* Введение в вычислительную физику. — М.: Изд-во Физтеха, 1994.

78. *Федоров В.В.* Теория оптимального эксперимента. — М: Наука, 1971.

79. *Федотов А.М.* Линейные некорректные задачи со случайными ошибками в данных. — Новосибирск: Наука, 1982.

80. Фракталы в физике. — М.: Мир, 1989.

81. *Хампель Ф.*, *Рончетти Э.*, *Рауссеу П.*, *Штаэль В.* Робастность в статистике. Подход на основе функций влияния. — М.: Мир, 1989.

82. *Холдендер М.*, *Вулф Д.* Непараметрические методы статистики, — М.: Финансы и статистика, 1983.

83. *Хъюбер П.* Робастность в статистике. — М.: Мир, 1984.

84. *Цыпкин Я.З.* Основы информационной теории идентификации. — М., 1984.

85. *Ченцов Н.Н.* Почему L_1-подход и что за горизонтом? Дополнение 1 к книге Деврой Л., Дьерфи Л. Непараметрическое оценивание плотности. — М.: Мир, 1988.

86. *Ченцов Н.Н.* Статистические решающие правила и оптимальные выводы. — М.: Наука, 1972.

87. *Шарп У.Ф.*, *Александер Г.Дж.*, *Бейли Д.В.* Инвестиции. — М.: Инфра-М, 2000.

88. *Шурыгин А.М.* Прикладная стохастика: робастность, оценивание, прогноз. — М.: Финансы и статистика, 2000.

89. *Шустер Г.* Детерминированный хаос. — М.: Мир, 1988.

90. *Abarbanel H. D. I.* Analysis of Observed Chaotic Data. — N. Y.: Springer, 1996.

91. *Arsenin V. Ya.*, *Krianev A.V.* Generalized maximum likelihood method and its application for solving ill-posed problems. Некорректно-поставленные задачи в естественных науках / Под редакцией А.Н. Тихонова. — М.: ТВП, 1992.

92. *Broomhead D. S.*, *Jones R.*, *King G. P*, *Pike E. R.* Singular system analysis with application to dynamical systems. — In: Chaos, Noise and Fractals, ed. by E. R. Pike and L. A. Lugaito. — Bristol: IOP Publishing, 1987.

93. *Broomhead D. S., King G. P.* On the qualitative analysis of experimental dynamical systems. — In: Nonlinear Phenomena and Chaos, ed. by S. Sarkar. — Bristol: Adam Hilder, 1986, p. 113–144.
94. *Broomhead D. S., King G. P.* Extracting qualitative dynamics from experimental data // Physica D. 1986. V. 20. P. 217–236.
95. *Daubechies I.* Ten lectures on wavelets. — SIAM, 1992.
96. *Draper R. N., Smith H.* Applied Regression Analysis. — N. Y.: Wiley, 1998.
97. *Elsner J. B., Tsonis A. A.* Singular Spectrum Analysis. New Tool in Time Series Analysis. — N. Y.: Plenum Press, 1996.
98. *Golyandina N., Nekrutkin V., Zhigljavsky A.* Analysis of Time Series Structure. SSA and Related Techniques. — Chapman & Hall / CRS, 2001.
99. *Hernander E., Weiss G.* A First Course of Wavelets. — CRS Press, 1996.
100. *Jolliffe I. T.* Principal Component Analysis. Springer Series in Statistics. — N. Y.: Springer-Verlag, 1986.
101. *Kantz H., Schreiber T.* Nonlinear Time Series Analysis. — Cambridge: Cambridge University Press, 1997.
102. *Keizer G.* A Friendly Guide to Wavelets. — Birk, 1995.
103. *Lokenath Debnath.* Wavelet Transforms and Their Applications. — Hardcover Edition, 2002.
104. *Misiti M., Misiti Y., Oppenheim G., Poggy J.-M.* Wavelet Toolbox for Use with MATLAB, 1988.
105. *Mosteller F., Tukey J. W.* Data Analysis and Regression: Second Course in Statistics. Reading. — MA: Addison-Wesley, 1977.
106. *Parzen E.* On estimation of a probabilities for sums of bounded random variables // Annals of Mathematical Statistics. 1962. V. 33. P. 1065–1076.
107. *Persival D. B., Walden A. T.* Wavelet Methods for Time Series. — Hardcover Edition, 2002.
108. *Rosenblatt M.* Remarks on some nonparametric estimates of a density function // Annals of Mathematical Statistics. 1956. V. 27. P. 832–835.
109. *Tikhonov A.N., Leonov A.S., Yagola A.G.* Nonlinear Ill-Posed Problems. V. 1, 2. — London: Chapman Hall, 1998.
110. *Vautard R., Yiou P., Ghil M.* Singular-spectrum analysis: A toolkit for short noisy chaotic signals // Physica D. 1992. V. 58. P. 95–126.

Предметный указатель

Вейвлет 177
— FHAT 179
— базисный 178
— Добеши 179
— Хаара 178
Выборка априорная 44, 126
Выбросы большие 48, 57, 112

Гипотеза о законе распределения 74, 76
— — совпадении законов распределения 77

Дисперсия эмпирическая 19
Доля засорения 54, 56

Задача минимаксная 51
— некорректно поставленная 69
— экстремальная 29, 36, 51
Задачи восстановления некорректно поставленные 117
— линейного регрессионного анализа 90
— прогнозирования 109
— экономические 109

Информация априорная 35, 37, 40, 41, 43, 46, 48, 117
— — детерминированная 137
— Фишера 26, 51
Итерационный процесс Левенберга 151
— — Ньютона–Гаусса 150

Компонента детерминированная 153
— хаотическая 153
Коридор ошибок 193
Кофлеты 180
Критерий ω^2 74
— Колмогорова 74
— Пирсона 75
— хи-квадрат 75

Матрица Гильберта 154
— информационная обобщенная 128
— — Фишера 27, 30, 34, 45, 46
— плохо обусловленная 108
— сопряженности 98
— точности 39
Медиана выборки 54
Мера точности 37
Метод Байеса 35, 38, 40, 118
— вариационно-взвешенных квадратических приближений 59, 115
— выделения главной части 87
— вычетов Лемера 81
— гистограмм 62, 70
— квадратного корня 43, 108
— корневой оценки плотности 71
— максимального правдоподобия 29, 44, 72
— — — обобщенный 35, 42, 45, 126
— минимаксный 40, 118, 121
— — Хьюбера 54
— моментов 20
— Монте-Карло 79
— наименьших квадратов 90
— — — итеративный 58, 114
— Неймана 81
— Ньютона–Гаусса 149
— ортогональных проекций 141
— — — регуляризованный 144
— оценивания устойчивый 49
— регуляризации Тихонова 124
— Розенблатта–Парзена 63
— сглаживающих линейных сплайнов 158

— — ортогональных полиномов 153
— — — — робастных 158
— сопряженных градиентов 42, 70
— Форсайта 154
— Холецкого 43
— Ченцова 71
Методы восстановления робастные 111
— непараметрические 16, 60
— оценивания робастные 111
— проекционные 66
— статистического моделирования 79
Минимаксный подход 49
Множество априорное 40–42, 44, 47, 121
Моделирование случайных величин 79
Модель регрессионная 118
— — многофакторная 109
— смеси 113

Неравенство Фишера–Крамера–Рао 26, 28
— — обобщенное 134
Нижняя граница Фишера–Крамера–Рао 30, 33

Оценка L_p 58
— Андрюса 58
— асимптотически нормальная 34
— Байеса 36–40, 118
— взвешенная МНК Гаусса–Маркова 104
— Винзора 56
— Грама–Шарлье 67
— метода максимального правдоподобия 29–33, 41, 44, 45, 49
— — наименьших квадратов 93, 112, 154
— — ортогональных проекций 141
— — — — регулиризованная 144
— минимаксная 40, 41, 43, 121
— — Хьюбера 56
— обобщенного метода максимального правдоподобия 42, 44–46
— плотности ОММП 70
— Рамсея 58
— робастная 17, 49, 53, 57, 58, 112
— устойчивая 55
— Хьюбера 54–56, 112
— — робастная 113
— Ченцова 69
— эффективная 17, 26, 28, 48
Оценки совместно эффективные 28, 134

Параметр регуляризации 63, 67
— сдвига 49, 51, 56
— Хьюбера 52, 112
План эксперимента 194
— — непрерывный нормированный 198
— — нормированный 194
— — — непрерывный оптимальный 198
Планирование оптимальных экспериментов 192
Плотность вероятностей 15
— — апостериорная 36
— — гипотетическая 48
— — реперная 52, 66
— — робастная 51
— — устойчивая 51, 52, 54
— — — относительно класса 49
Показатель Херста 155
Полином Чебышева–Лагерра 68
— Чебышева–Эрмита 67
Последовательность асимптотически нормальная 33
Прогнозирование временных процессов 182

Распределение апостериорное 37
— — совместное 38, 39
— априорное 38
— — совместное 38, 39
— реперное 67
— условное 39
Риск апостериорный 36

Симлеты 180
Сингулярно-спектральный анализ 187
Система уравнений правдоподобия 29
Сопряженное семейство 37
Сплайн кубический интерполяционный 167
— — сглаживающий 167

— — — робастный 175
— линейный интерполяционный 159
— — сглаживающий 158
— — — робастный 163
Способ невязки 71
Статистическое моделирование 79

Теорема Гаусса–Маркова 94
— Гливенко–Кантелли 60
— Де Ла Гарза 199
— Неймана 81
— Пирсона 76
— Фишера 76
— Фишера–Крамера–Рао 24
— Хьюбера 52
— Эйкера 97

Уравнение интегральное 1-го рода 69, 91
— линейного регрессионного анализа 93
— правдоподобия 53
Уравнения обобщенные правдоподобия 45
Уровень доверия 23
— риска 22

Формула Байеса 36, 118
Функционал затрат 197
— потерь 197
Функция Андрюса 112
— относительной эффективности измерений 193
— потерь 36, 37
— правдоподобия 24, 29, 42, 44
— — априорная 126
— — логарифмическая 24, 44, 54
— — — совместная 45
— — обобщенная 45, 126
— — совместная 45
— Рамсея 112
— Тьюки 165
— Хьюбера 57, 112, 163
— эффективности измерений 193

Числа псевдослучайные 81

Учебное издание

КРЯНЕВ Александр Витальевич
ЛУКИН Глеб Владимирович

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ НЕОПРЕДЕЛЕННЫХ ДАННЫХ

Редактор *Н.Б. Бартошевич-Жагель*
Оригинал-макет: *В.В. Худяков*
Оформление переплета: *А.А. Логунов*

Подписано в печать 19.06.06. Формат 60×90/16. Бумага офсетная.
Печать офсетная. Усл. печ. л. 13,5. Уч.-изд. л. 14,5. Тираж 2000 экз.
Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117997, Москва, ул. Профсоюзная, 90
E-mail: fizmat@maik.ru, fmlsale@maik.ru;
http://www.fml.ru

Отпечатано с готовых диапозитивов
в ОАО «Ивановская областная типография»
153008, г. Иваново, ул. Типографская, 6
E-mail: 091-018@adminet.ivanovo.ru

ISBN 5-9221-0724-0