51 31-859

А.В. Крянев Г.В. Лукин Д.К.Удумян

## МЕТРИЧЕСКИЙ АНАЛИЗ И ОБРАБОТКА ДАННЫХ



УДК 519.2+6 ББК 22.17, 22.19 K 85

Крянев А.В., Лукин Г.В., Удумян Д.К. Метрический анализ M.: ФИЗМАТЛИТ, 2012. - 308 c. обработка данных. -ISBN 978-5-9221-1068-6.

Основная цель книги — ознакомить читателя с наиболее эффективными и апробированными классическими и новыми стохастическими и детерминированными методами оценки и прогнозирования, научить использовать эти метозы при решении конкретных задач обработки данных.

Изложены основные понятия параметрической и непараметрической статистики, а также некоторые новые методы робастного оценивания, учета априорной информации и прогнозирования, включая алгоритмы их численной реализации. Представлены основы нового направления обработки данных метрического анализа, позволяющего решать задачи интерполяции, восстановдения и прогнозирования функций одной и многих переменных на основе эффективного использования информации стохастического и детерминированного характеров об исследуемой функциональной зависимости.

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

Предназначена преподавателям, научным работникам, аспирантам, студентам старших курсов различных специальностей, использующих математические методы обработки данных.

Учебное излание

КРЯНЕВ Александр Витальевич ЛУКИН Глеб Владимирович УДУМЯН Давид Каджикович

### метрический анализ и обработка данных

Редактор Н.Б. Бартошевич-Жагель Оригинал-макет: Ю.В. Владимирова Оформление переплета: Н.В. Гришина

Подписано в лечать 12.04.10. Формат 60×90/16. Бумага офсетная. Печать офсетная.

Усл. веч. л. 19.25. Уч.-изд. л. 19.25. Тираж 500 экз. Заказ № К-8216

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 электронных носителей издательства в ГУП «ИПК «Чуващия», 428019 г. Чебоксары, пр-т И.Яковлева. 13



© ФИЗМАТЛИТ, 2012

© А.В. Крянев, Г.В. Лукин, Д.К. Удумян,

ISBN 978-5-9221-1068-6

### ОГЛАВЛЕНИЕ

Предисловие	6
Введение	12
Основные обозначения	16
Глава 1. Элементы математической статистики	19 20 23 29 35
Глава 2. Методы учета априорной информации в рамках	
параметрической статистики  1. Метод Байеса 2. Минимаксный метод учета априорной информации  3. Обобщенный метод максимального правдоподобия учета априорной детерминированной информации  4. Обобщенный метод максимального правдоподобия учета априорной стохастической информации	42 42 48 49 52
Глава 3. <b>Устойчивые методы оценивания параметра положения</b> 1. Минимаксный метод Хьюбера	56 57
2. Робастные М-оценки	65
Глава 4. Методы непараметрической статистики	70
1. Восстановление функции распределения	70 72
3. Восстановление плотности распределения методом Розенблатта-Парзена	74
4. Восстановление плотности распределения проекционными методами	77
ным методом гистограмм 6. Метод корневой оценки плотности распределения	80 83
Глава 5. Проверка гипотез о законе распределения	86
1. Проверка гипотезы о законе распределения в рамках непараметрической статистики	86 88
Глава 6. Численные методы статистического моделирова-	30
ния	92 92
2. Применение метода статистического моделирования для	98

Глава 7. Метод наименьших квадратов для линейных моде-	
лей с неопределенными данными	105
1. Примеры линейных моделей с неопределенными данными	106
2. Классическая схема МНК	108
3. Обобщения классической схемы МНК	121
4. Прогнозирование с помощью линейных регрессионных	
моделей	128
	120
Глава 8. Робастные методы для линейных моделей	
с неопределенными данными	131
1. Робастные М-оценки параметров линейных моделей	131
2. Численные методы нахождения робастных оценок пара-	
метров линейных моделей	134
Глава 9. Учет априорной информации в линейных моделях	
с неопределенными данными	138
1. Метод Байеса в рамках линейных моделей	140
2. Минимаксный метод учета детерминированной априорной	1 10
информации	142
3. Обобщенный метод максимального правдоподобия учета	
априорной стохастической информации в рамках линей-	
ных моделей	146
4. Обобщенный метод максимального правдоподобия учета	2 2 0
априорной детерминированной информации в рамках ли-	
нейных моделей	161
5. Регуляризованный метод наименьших квадратов	164
Глава 10. Метод наименьших квадратов для нелинейных	171
моделей с неопределенными данными	171
1. МНК-оценки параметров нелинейных моделей и их свой-	171
CTBa	171
2. Численные методы нахождения МНК-оценок параметров	174
нелинейных моделей	1/4
Глава 11. Методы выделения детерминированных и хаоти-	
ческих компонент временных рядов	179
1. Метод сглаживающих ортогональных полиномов	179
2. Метод сглаживающих линейных сплайнов	185
3. Робастные линейные сглаживающие сплайны	190
4. Метод сглаживающих кубических сплайнов	194
5. Метод вейвлетов	206
Глава 12. Методы прогнозирования хаотических времен-	
ных рядов	212
Метопы просиозиворомия	Z 1 Z
1. Методы прогнозирования с использованием априорной информации и ортогом в именто в приорной информации и ортогом в полительной в приорной информации и ортогом в полительной в полительном в полительной в полительном в полительном в полительном в полител	212
формации и ортогональных полиномов	232
формации и линейных сплайнов	215
3. Методы прогнозирования с использованием сингулярно-	210
CHACTRO SLUGGO AND THE	217
Circui parionol O analinaa	

Глава 13. Интерполяция функций одной и многих переменных с помощью схем, основанных на метрическом анализе	232
1. Интерполяция функций одной и многих переменных с по- мощью схем, основанных на метрическом анализе 2. Примеры интерполяции	234 248
Глава 14. Аппроксимация и сглаживание функций одной и многих переменных на основе метрического анализа 1. Схемы аппроксимации и сглаживания методом метрического анализа 2. Оптимальность восстановленных значений функции на основе метрического анализа	252 252 261
Глава 15. Алгоритмы и вычислительные схемы прогнозирования на основе метрического анализа	268
Глава 16. <b>Динамический детерминированный хаос</b>	273 273 276
Глава 17. Планирование оптимальных измерений при восстановлении функциональных зависимостей	288 289 296
Список литературы	303

### Предисловие

Представляемая книга является расширенным вариантом ранее вышедшего издания — Крянев А. В., Лукин Г. В. Метрический анализ и обработка данных. М.: Физматлит, 2010.

При подготовке настоящей книги были внесены существенные дополнения, включающие некоторые последние достижения, в том числе авторов книги, и прежде всего по новому направлению — метрическому анализу, что и определило название первого и настоящего изданий книги.

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

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 информации

мация 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  $u \in R_a$ , — оба из которых используют понятия метрического характера, связанные прежде всего с использованием расстояния в пространстве искомого вектора u. 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 искомои 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,  $\omega^2$ ,  $\chi^2$ .

Глава 6 посвящена численным методам статистического мотелипования. В этой главе представлены способы моделирования случайных величин, в частности датчики равномерно распределенной нормированной случайной величины у, 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 вариационновзвешенных 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 детерминированная информация, задаваемая в виде априорного детерминированного множества, которому заведомо принадлежит искомый вектор u. В п. 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 сингулярноспектральном 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. Значительная часть гл. 12 посвящена нестационарному сингулярно-спектральному анализу.

В гл. 13 рассматриваются основы метрического анализа, позволяющего решать задачи интерполяции, экстраполяции и прогнозирования функций одной и многих переменных. Вводится понятие матрицы метрической неопределенности, с помощью которой в дальнейшем определяются интерполяционные, восстановленные и прогнозируемые значения исследуемой функциональной зависимости. Предлагаются различные схемы интерполяции, основанные на метрическом анализе. Рассматриваются примеры интерполяции функций одной и многих переменных.

В гл. 14 рассматриваются задачи аппроксимации и сглаживания функций одной и многих переменных на основе метрического анализа. Предложены вычислительные схемы аппроксимации и сглаживания функций одной и многих переменных на основе метрического анализа. Рассматриваются примеры аппроксимации и сглаживания функций одной и многих переменных с помощью предложенных схем. Исследуется оптимальность восстановленных значений функции на основе метрического анализа. Показано, что в рамках вычислительных схем метрического анализа можно оптимальным образом учитывать неопределенности как метрического, так и стохастического характеров.

Задачи экстраполяции и прогнозирования на основе метрического анализа рассматриваются в гл. 15. Показано, что задача прогнозирования на основе метрического анализа сводится к задаче восстановления значений функции с помощью метрического анализа в рамках нелинейной авторегрессионной модели. Это позволяет получать эффективные прогнозные значения. Приведены конкретные примеры прогнозирования с помощью

метрического анализа.

Глава 16 посвящена интенсивно развиваемому в последние годы детерминированному хаосу. В главе приведены широко интируемые в научной литературе, посвященной детерминированному хаосу, примеры процессов с детерминированным хаосом. Приведены наиболее значимые известные к настоящему времени свойства детерминированных хаотических процессов, использующие понятия и инструменты метрического характера. Читателям, желающим более обстоятельно ознакомиться с содержанием теории детерминированного хаоса, можно рекомендовать книгу С.П. Кузнецова [55] и книги [61, 62].

В гл. 17 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 [75], Ж. Закса [38], М. Кендалла и А. Стюарта [43], С.Р. Рао [78], энциклопедию «Вероятность и математическая статистика» [22]. Теория матриц достаточно полно изложена в монографии Ф. Р. Гантмахера [26]. С численными методами, используемыми при обработке неопределенных данных, можно ознакомиться в книrax [23, 41, 63, 68, 83, 90].

Авторы благодарят заведующего кафедрой «Прикладная математика» МИФИ, д.ф.-м.н., профессора Н.А. Кудряшова, д.ф.-м.н., профессора этой же кафедры Т.И. Савелову 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 адресу: avkryanev@mephi.ru. Материалы, использованные в данной монографии, а также информацию о направлениях деятельности авторов можно найти на сайте www.kryanev.ru.

Москва, апрель 2011 г. А.В. Крянев

А.В. Крянев Г.В. Лукин Д.К. Удумян

### 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

1.

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 потребностей 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 7-12 книги представлены методы и алгоритмы восстановления регрессионных зависимостей и их применения для прогнозирования.

Задачи восстановления регрессионных зависимостей встречаются во всех областях естествознания при обработке неопреде-

ленных 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 решення 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 приемлемога приближенного решения этих систем естественно применять статистические методы.

Применение статистических подходов и методов для решения задач восстановления регрессионных зависимостей имеет ряд особенностей, 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.

В главах 13-15 книги представлены новые методы восстановления функциональных зависимостей с помощью схем, основанных на метрическом анализе, которые позволяют получать эффективные оценки неизвестных значений функции в условиях неопределенности.

В частности, в главе 15 на ряде примеров показана эффективность экстраполяции и прогнозирования значений функций одной переменной и временных рядов с помощью метрического анализа и проведено сравнение с результатами экстраполяции и прогнозирования с помощью схем, основанных на сингулярно-спектральном анализе

Часто исследователь не является лишь пассивным обработчиком неопределенных данных, но и может повлиять на процесс их получения. В таких условиях естественно добиваться такого Введение

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

 $X_1, \dots, X_n$  — элементы выборки

p(x) — плотность вероятностей

F(x) — функция распределения вероятностей

 $p_{\xi}(x)$  — плотность вероятностей случайной величины  $\xi$ 

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  $\widehat{u}$ 

 $P_{\text{mos}}$  — доверительная вероятность

ок — начальный момент k-го порядка

μ<sub>k</sub> — центральный момент k-го порядка

 $\sigma^{z}(\xi), \mathbf{D}(\xi)$  — дисперсия случайной величины  $\xi$ 

 $s^2$  — эмпирическая дисперсия

$$\Phi(t)=rac{1}{\sqrt{2\pi}}\int\limits_0^t e^{-rac{s^2}{2}}\,ds$$
 — интеграл вероятностей

 $\Gamma(t)$  — гамма-функция

 $\chi_n^2$  — случайная величина, распределенная по закону хи-квадрат с n степенями своболы

 $I_n, I_n(u)$  — информация Фишера или информационная матрица Фишера

 $L_n, L_n(u)$  — функция правдоподобия

 $l_n, l_n(u)$  — логарифмическая функция правдоподобия

 $K_{\xi}$  — ковариационная матрица случайного вектора  $\xi$ 

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$
 — средняя арифметическая

 $\mathcal{N}(m,\sigma^2)$  — нормальный закон распределения с математическим ожиданием m и дисперсией  $\sigma^2$ 

и
Б — оценка Байеса параметра и

 $K_a$  — ковариационная матрица априорного распределения

 $u_a$  — центр априорного распределения параметра u

 $p_{\mathsf{anp}}(u)$  — плотность априорного распределения

 $U_a, R_a$  — априорное множество

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

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

 $\Delta u$  — смещение оценки u

 $\displaystyle \operatorname*{arg\ min}_{u}(max)f(u)$  — значение u, для которого целевая функция

f(u) достигает минимума (максимума)

К — параметр Хьюбера

ho(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

 $\mathfrak{O}_{K_n(u)}$  — функция распределения Масси

 $\delta(x)$  — обобщенная функция Дирака

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

 $L_k^{lpha}(x)$  — полином степени k Чебышева-Лагерра

L(c) — функция Лагранжа

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

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

 $\widehat{u}_{\mathsf{MHK}}$  — 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

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

 $\widehat{u}_{\mathsf{MM}}$  — оценка минимаксного метода параметра u

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

||u|| — евклидова норма вектора u

 $\psi(t)$  — базисный вейвлет

 $S_n(x)$  — кубический интерполяционный сплайн

 $S_{nlpha}(x)$  — кубический сглаживающий сплайн

 $S_{lpha}(t)$  — линейный сглаживающий сплайн

MA H. 3 BAYMAHA

1094636

 $\mathcal{E}(N)$  — план эксперимента

 $\varepsilon(N)$  — нормированный план эксперимента

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

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

W - матрица метрической неопределенности



# 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 [75]. 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  $\xi$ , для которой известна (из эксперимента) совокупность ее случайных реализаций

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

Требуется с максимально возможной точностью определить закон распределения  $\xi$ , 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 случай-

ных 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  $\xi$ , т. е.

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

с неизвестными параметрами  $m, \sigma^2$ .

В рамках параметрического подхода задача восстановления плотности вероятностей сводится к задаче восстановления значений тех неизвестных параметров и, 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. Параметрическая статистика. Свойства оценок

Интунтивно 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 веро-

ятностей) 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  $p_{\xi}$  будем обозначать симво-

лами  $\widehat{u}$  и  $\widehat{p}_{\mathcal{E}}$  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  $u_T-u$ скомое 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  $\widehat{u}=\widehat{u}(X_1,\dots,X_n)$  называется состоятельной по вероятности (в среднем квадратичном), если  $\widehat{u}(X_1,\dots,X_n)$  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

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

где S — оператор из конечномерного пространства элементов  $(X_1, \dots, X_n)$  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:  $\mathbf{M}\,\widehat{u}(X_1,\dots,X_n)\to u_T$  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 — uнтервальный. При интервальном оценивании находят две оценки скалярного параметра u  $u_{\rm H}, u_{\rm B}$  такие, что определена вероятность  ${\bf P}\left\{u\in(u_{\rm H},u_{\rm B})\right\}=p(u)$ . Величину  $p_{\rm дов}=inf_up(u)$  называют доверительным уровнем, а  $1-p_{\rm дов}$  уровень риска. В приложениях часто доверительный уровень и уровень риска выражают в процентах, а границы  $u_{\rm H},u_{\rm B}$  подбирают так, чтобы доверительный уровень равнялся одному из стандартных значений: 0,9; 0,95; 0,99; 0,999. Например, если  $p_{\rm дов}=0,95$ , 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

$$\int\limits_{\widehat{u}-\delta_1}^{\widehat{u}}p_{\widehat{u}}(u)\,du=rac{p_{\mathtt{MOB}}}{2}, \quad \int\limits_{\widehat{u}}^{\widehat{u}+\delta_2}p_{\widehat{u}}(u)\,du=rac{p_{\mathtt{MOB}}}{2}.$$

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

$$\int\limits_{\widehat{u}-\delta}^{\widehat{u}+\delta}p_{\widehat{u}}(u)\,du=p_{ exttt{ iny AOB}}.$$

С методами математической статистики можно ознакомиться в книгах [2-5, 38, 43, 46, 58, 78].

### 2. Метод моментов

Пусть (1.1) — повторная выборка скалярной случайной величины  $\xi$ , 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  $\mathbf{M} \left( \boldsymbol{\xi} - \mathbf{M} \, \boldsymbol{\xi} \right)^k$  возьмем статистику

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

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

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

Оценки (1.3) несмещены, так как  $\mathbf{M} \, \widehat{\alpha}_k = \frac{1}{n} \sum_{i=1}^n \mathbf{M} \, X_i^k = \mathbf{M} \, \xi^k$ , 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)

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

где  $\mu_4$  — центральный момент 4-го порядка случайной величины  $\xi$ .

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

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

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  $\xi$  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  $\xi$  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

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

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

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

B 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.  $p_{\xi}(x;u)=\left\{ egin{array}{ll} ue^{-ux}, & x\geqslant 0, \\ 0, & x<0, \ {
m rge}\ u>0. \end{array} 
ight.$  Возьмем  $a_1(x)=x.$  Имеем:  ${\bf M}\,a_1(x)={\bf M}\,\xi=1/u.$  Поэтому

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

$$\widehat{u} = n / \sum_{i=1}^{n} X_i.$$

Замечание 1.1. Если интегралы

$$\int_{-\infty}^{\infty} p_{\xi}(x; u) a_{j}(x) dx$$

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  $\xi$  — 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  $\xi$  можно взять

жно взять
$$\hat{m} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \left( \sum_{i=1}^{n} X_{i1}, \dots, \sum_{i=1}^{n} X_{ir} \right)^T = (\overline{X}_1, \dots, \overline{X}_r)^T.$$

Статистики

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

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

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

Следовательно, состоятельной и несмещенной оценкой ковариационной матрицы  $K_{\xi}$  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

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

где  $X - (n \times r)$ -матрица.

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

Рассмотрим сначала случай, когда m неизвестен,  $\sigma^2$  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

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

где 
$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int\limits_0^t e^{-s^2/2} \, ds$$
 — 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

$$\widehat{m} - t(p_{\text{MOB}}) \frac{\sigma}{\sqrt{n}} \leqslant m \leqslant \widehat{m} + t(p_{\text{MOB}}) \frac{\sigma}{\sqrt{n}}.$$
 (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  $\sigma^2$  оба неизвестны. Тогда в качестве точечных оценок этих параметров возьмем статистики (1.4) и (1.7). При построении доверительных интервалов для m и  $\sigma^2$  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  $\chi^2_{n-1}$ .

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

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

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  $z_n=\frac{1}{\sqrt{n}}\sum_{i=1}^n y_i$ . 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 изменя-

ются. Обозначим  $m_Y = \frac{1}{n} \sum_{i=1}^n Y_i$ . 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  $s^2(n-1)=\sum_{i=1}^{n-1}Z_i^2$  следует, что величина  $\widehat{m}=m+Z_n/\sqrt{n}$  не зависит от  $s^2$ , а  $s^2$  распределена по закону  $\chi^2_{n-1}$ . Теорема доказана.

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

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

Имеем: 
$$\mathbf{P}\left\{ rac{|\widehat{m}-m|}{s/\sqrt{n}} < t 
ight\} = p_{\mathtt{ДОВ}},$$
 где  $t=t(n-1,p_{\mathtt{ДОВ}})$  находит-

ся из равенства  $2\int\limits_{0}^{\infty}p_{T_{n-1}}(s)\,ds=p_{{{\tt дов}}}.$ 

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  $\sigma^2$  является своеобразной «платой» за незнание  $\sigma^2$ .

Построение доверительного интервала для  $\sigma^2$  также основано

на результате теоремы 1.1. Действительно,

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

где

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

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

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

гласно правилу

$$\int\limits_{0}^{u_{2}}p_{\chi_{n-1}^{2}}(u)\,du=\frac{1-p_{\text{\tiny ДOB}}}{2},\quad \int\limits_{u_{1}}^{+\infty}p_{\chi_{n-1}^{2}}(u)\,du=\frac{1-p_{\text{\tiny ДOB}}}{2}.$$

Таким образом, доверительный интервал для  $\sigma^2$  имеет вид

$$z_1(n-1, p_{\text{doB}})s^2 < \sigma^2 < z_2(n-1, p_{\text{doB}})s^2,$$
 (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.

### 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 логарифмическию функцию правдоподобия

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

Для повторной выборки

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

rae  $l_1(X_i; u) = \ln p_c(X_i; u)$ .

Как мы уже знаем, существует много несмещенных и состоятельных оценок параметра и. 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  $\mathbf{u}_T$  дается дисперсией оценки  $\mathbf{D}\,\widehat{u} = \mathbf{M}\,(\widehat{u} - u_T)^2$ .

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

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

эбе

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

Доказательство. Имеем

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

Откуда

$$\int_{-\infty}^{\infty} \dots \int \frac{\partial L_n}{\partial u} dX_1 \dots dX_n = 0.$$
 (1.13)

Ho  $\mathbf{M} \, \widehat{\mathbf{u}} = \mathbf{u}$  или

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



Дифференцируя последнее равенство, получаем

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

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

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

Левую часть последнего равенства можно записать в виде

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

Итак, получено равенство

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

Обозначим  $Y=\widehat{u}-u_T,\ Z=\frac{\partial l_n}{\partial u}$ . 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  $\mathbf{M}^2YZ\leqslant \mathbf{M}\,Y^2\,\mathbf{M}\,Z^2$ . Поэтому

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

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

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

так как

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

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

$$I_n=-\sum_{i=1}^n \mathbf{M}\, rac{\partial^2 l_1}{\partial u^2}=nI_1,$$
 где  $I_1=-\mathbf{M}\, rac{\partial^2 l_1}{\partial u^2}.$ 

Таким образом,  $I_n$  карактеризует 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

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

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



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  $\hat{u}$  эффективна.

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

$$L_n = u^n e^{-u \sum_{i=1}^n X_i}, \quad X_i > 0, \quad i = 1, \dots, n,$$
  
 $l_n = n \ln u - u \sum_{i=1}^n X_i, \quad I_1 = \frac{1}{u^2}, \quad I_n = \frac{n}{u^2}.$ 

Итак, для неизвестного параметра экспоненциального закона информация Фишера зависит от значения самого искомого параметра. Для оценки  $\widehat{u}=n/\sum_{i=1}^{n}X_{i}$   $e_{n}<1$ .

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

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

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

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

Ho

$$\mathbf{M}\left(\frac{1}{L_n}\frac{\partial^2 L_n}{\partial u_i \partial u_j}\right) = \int_{-\infty}^{\infty} \dots \int_{-\infty} \frac{\partial^2 L_n}{\partial u_i \partial u_j} dX_1 \dots dX_n = 0,$$

$$\mathbf{M}\left(\frac{1}{L_n}\frac{\partial L_n}{\partial u_i}\right) = \int_{-\infty}^{\infty} \dots \left[\frac{\partial L_n}{\partial u_i}dX_1\dots dX_n = 0\right].$$

Поэтому

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

2 А.В. Коянев, Г.В. Лукин. Л.К. Удумян

Таким образом, информационная матрица Фишера является ковариационной матрицей совокупности центрированных случай. х величин  $\left(\frac{1}{L_n}\frac{\partial L_n}{\partial u_1},\dots,\frac{1}{L_n}\frac{\partial L_n}{\partial u_m}\right)$  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, ..., m.

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

Теорема 1.3. Ковариационная матрица  $K_u$  любой несмещенной оценки и 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 и. Оно означает, что симметричная матрица  $K_u - I_n^{-1}$  неотрицательна. Элементы главной диагонали матрицы  $K_u$  являются дисперсиями  $\sigma_u^2$  компонент  $\widehat{u}_i$  вектора оценки  $\widehat{u}$ . Поэтому из неравенства Фишера-Крамера-Рао (1.14), в частности, следуют неравенства  $\sigma_{ii}^2\geqslant I_{ii}^{-1},\ i=1,\dots,m,$  где  $I_{ii}^{-1}-i$ -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  $(K_u v, v) \geqslant (I_u^{-1} v, v)$  для любого  $v \in E^m$ . Обозначим  $b_i(u)$  смещение i-й компоненты вектора оцен-KH  $\widehat{u} = (\widehat{u}, \dots, \widehat{u}_m)^T$ , T. e.  $\mathbf{M} \widehat{u} = u_i + b_i(u)$ .

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

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

$$\mathbf{M}(\widehat{u} - u)^2 \geqslant \frac{(1 + b'(u))^2}{I_n},$$
 (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) [38].

### 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

$$\widehat{u}_{\mathsf{MM\Pi}} = \arg \max_{u} L_n(X_1, \dots, X_n; u). \tag{1.17}$$

Если функция правдоподобия дифференцируема по u, то  $MM\Pi$ -оценка (1.17) является решением системы уравнений правдоподобия

$$\operatorname{grad}_{u} L_{n}(X_{1}, \dots, X_{n}; \widehat{u}) = 0.$$
 (1.18)

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

$$\widehat{u}_{\mathsf{MM\Pi}} = \arg \max_{u} l_n(X_1, \dots, X_n; u), \tag{1.19}$$

$$\operatorname{grad}_{u} l_{n}(X_{1}, \dots, X_{n}; \widehat{u}) = 0.$$
 (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

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

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

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 Такое свойство оценок называют асимптотической эф- а ММП-оценка  $\widehat{\sigma}^2$  асимптотически несмещена.

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

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

откуда 
$$\widehat{u}_{\mathrm{MM}\Pi} = n / \sum_{i=1}^n X_i$$
.

Пример 1.6. Пусть

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

$$L_{n}(X_{1},\dots,X_{n},a,b) = \frac{1}{(b-a)^{n}}, \quad a \leqslant X_{\min}, \quad b \geqslant X_{\max},$$

гле

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

Cледовательно, максимум  $L_n$  достигается при  $\widehat{b}_{\mathsf{MM\Pi}} = X_{\max},$   $\widehat{a}_{\mathsf{MM\Pi}} = X_{\min}$ 

**Пример 1.7.** Пусть совместная плотность вероятностей членов выборки дается равенством

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

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

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

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

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

В частности, если  $K_x={
m diag}\,(\sigma_1^2,\ldots,\sigma_n^2)$ , то  $(K_x^{-1}\,1,1)=\sum_{i=1}^n \frac{1}{\sigma_i^2}$ ,

$$(K_x^{-1} 1, X) = \sum_{i=1}^n \frac{X_i}{\sigma_i^2}$$
 и  $\widehat{u}_{\mathsf{MM\Pi}} = \sum_{i=1}^n \frac{X_i}{\sigma_i^2} \Big/ \sum_{i=1}^n \frac{1}{\sigma_i^2} - \mathsf{средневзвешен}$ ная 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  $\widehat{u}_{jMM\Pi} = g_j(\widehat{u}_{MM\Pi})$ .

**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  $\mu$  и дисперсия  $d^2$  логнормального распределения связаны с  $m_0$  и  $\sigma^2$  соотношениями

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

Обозначим

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

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

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

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 достатистики  $T(X_1,\ldots,X_n)$  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 определе-

ния достаточной статистики.

**Теорема 1.6** (факторизации Фишера-Неймана). Для того чтобы статистика  $T(X_1,\ldots,X_n)$  была достаточной, необходимо и достаточно 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  $\varphi$  не зависит от искомого параметра и.

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

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

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

Поэтому набор двух скалярных статистик  $\left\{\overline{X}, \sum_{i=1}^n (X_i - \overline{X})^2\right\}$  — достаточная (векторная) статистика, если неизвестны  $m_0$  и  $\sigma^2$ . Ясно, что набор  $\left\{\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right\}$  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  $L_n(X_1,\dots,X_n;u)=1/u^n,\ X_i< u.$  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  $\widehat{u}_{\mathsf{MM\Pi}}$  зависит только от достаточной статистики  $T(X_1,\dots,X_n).$ 

Пример 1.11. Пусть

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

и выборка (1.1) повторная. Тогда  $\widehat{a}_{\mathsf{MM\Pi}} = X_{\min}$ ,  $\widehat{b}_{\mathsf{MM\Pi}} = X_{\max}$ , 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  $p_{\mathcal{E}}(x;u)$  называют экспоненциальной плотностью.

**Теорема 1.7.** Пусть  $1^{\circ}$  выборка (1.1) повторная и  $p_{\xi}(x;u)$  — экспоненциальная плотность;  $2^{\circ}$   $(m \times m)$ -симметричная матрица R с элементами  $R_{ij} = \frac{\partial^{2} u}{\partial u_{i} \partial u_{j}}$  положительно определена для всех  $u \in U \subset E^{m}$ . Тогда ММП-оценка  $\widehat{u}_{MM\Pi}$ 

единственна, является достаточной статистикой и корнем системы уравнений

$$-\frac{1}{n}\sum_{i=1}^{n}T_{j}(X_{i})=\frac{\partial\psi(\widehat{u}_{MM\Pi})}{\partial u_{j}}, \qquad j=1,\ldots,m.$$
 (1.21)

Доказательство. Имеем

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

Поэтому набор  $\left\{\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_m(X_i)\right\}$  является достаточ-

ной статистикой. Приравнивая к нулю частные производные  $\frac{\partial L_n}{\partial u_j}$ , 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 ММПоченки состоятельны, а именно: если (1.1) — повторная выборка, то  $\hat{u}_{\text{ММП}} \to u_T$  по вероятности при  $n \to \infty$ . Более того, для любого  $\varepsilon > 0$  существуют такие K,  $\alpha > 0$ , что  $\mathbf{P}\{|\hat{u}_{\text{ММП}} - u_T| \geqslant \varepsilon\} \leqslant Ke^{-\alpha n}$  для всех  $n \geqslant 1$ .

Ранее при рассмотрении примеров было отмечено, что ММПоценки являются либо эффективными, либо близки к нижней границе Фишера-Крамера-Рао при больших п. Оказывается, что и в общем случае при выполнении условий регулярности ММП-оценки являются асимптотически эффективными и асимптотически несмещенными [38].

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

Последовательность случайных величин  $a_n$  называют асимптомически нормальной, если найдутся такие детерминированные невырожденные матрицы  $A_n$  и векторы  $\mu_n$ , 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  $\sqrt{n}\,(\widehat{u}_{\mathsf{MM\Pi}}-u_T)$  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 асимп-

тотической 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,

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

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

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

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

$$\begin{split} \widehat{u}_{j\mathrm{MM\Pi}} - t(p_{\mathrm{dob}}) \sqrt{\frac{I_{jj}^{-1}(\widehat{u}_{\mathrm{MM\Pi}})}{n}} &\leqslant u_{jT} \leqslant \\ &\leqslant u_{j\mathrm{MM\Pi}} + t(p_{\mathrm{dob}}) \sqrt{\frac{I_{jj}^{-1}(\widehat{u}_{\mathrm{MM\Pi}})}{n}} \,, \end{split}$$

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

$$\frac{1}{\sqrt{2\pi}} \int_{0}^{t(p_{\text{MOB}})} e^{-s^2/2} \, ds = \frac{p_{\text{MOB}}}{2}.$$

Теоретическое обоснование ММП дано в монографии [38].

# 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.

#### 1. Метод Байеса

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

В большинстве прикладных задач искомый параметр и является детерминированной величиной, и поэтому искусственно навязанная байесовским подходом трактовка детерминированных величин и 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

$$p(x_1, ..., x_n, u) = p(x_1, ..., x_n \mid u) p_{anp}(u) = p(u \mid x_1, ..., x_n) p(x_1, ..., x_n),$$

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

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

Из вышесказанного следует формула

$$p(u \mid x_1, \dots, x_n) = rac{p(x_1, \dots, x_n \mid u) p_{\mathsf{anp}}(u)}{\int\limits_U p(x_1, \dots, x_n \mid u) p_{\mathsf{anp}}(u) \, du}.$$

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

$$p(u \mid X_1, \dots, X_n) = L_n(X_1, \dots, X_n; u) p_{\mathsf{anp}}(u) / C,$$

где нормировочный множитель

$$C = \int\limits_{U} L_n(X_1, \dots, X_n; u) p_{\mathsf{anp}}(u) \, du$$

HE SORUCUT OT 21

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

В качестве точечной байесовской оценки (Б-оценки)  $\widehat{u}_{\mathsf{B}}$  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

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

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

решение экстремальной задачи:

$$\widehat{u}_{\mathbb{B}} = \arg\min_{\widehat{u}} R(\widehat{u}; X_1, \dots, X_n), \tag{2.1}$$

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 апостернорной случайной величины и); 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  $p_{\rm anp}(u)$  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 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  $\tau=1/\sigma^2$  задана. Допустим, что  $p_{\rm anp}(u)\sim \mathcal{N}(\mu,1/\tau)$ , где  $\tau=1/\sigma^2$  — мера точности. Тогда апостериорное распреде-

ление является нормальным  $\mathcal{N}(\mu',1/(\tau+nr))$  со средним  $\mu'=\frac{\tau\mu+nr\overline{X}}{\tau+nr}$  и мерой точности  $\tau'=\tau+nr$ , где  $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$ .

Из теоремы 2.1 следует, что 6-оценки неизвестного параметра 6 для любой симметричной выпуклой функции потерь совпадают и даются формулой

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

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

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

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

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

$$u_{\rm B} = \left(\frac{n}{\sigma^2} K_x^{-1} + \frac{1}{\sigma_a^2} K_a^{-1}\right)^{-1} \left(\frac{n}{\sigma^2} K_x^{-1} \overline{X} + \frac{1}{\sigma_a^2} K_a^{-1} u_a\right), \tag{2.3}$$

где 
$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$
.

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

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

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

$$\widehat{u}_{\bar{b}} = \frac{\tau \mu + n \overline{X}}{\tau + n}, \quad \widehat{\sigma}_{\bar{b}}^2 = \frac{\widehat{\beta}'}{\alpha'}.$$
 (2.4)

Из (2.4) следует, что  $\widehat{u}_{\rm B}\sim\overline{X},\,\widehat{\sigma}_{\rm B}=\widehat{\sigma}_{\rm MM\Pi}^2$  при  $n\to\infty$  и  $\widehat{u}_{\rm B}\to\mu$  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  $\mu$  и матрицей точности  $rK_a^{-1}$ , а маргинальное распределение R-r гаммараспределение с параметрами  $\alpha,\beta>0$ . Тогда апостериорное совместное распределение R таково: условное распределение R при фиксированном R точное 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 c 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

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

Из (2.5) следует, что

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

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

$$\widehat{u}_{\mathsf{B}} - t(p_{\mathsf{QOB}}) \frac{1}{\sqrt{\tau'}} \leqslant u_T \leqslant \widehat{u}_{\mathsf{B}} + t(p_{\mathsf{QOB}}) \frac{1}{\sqrt{\tau'}},\tag{2.6}$$

TILE

$$au' = rac{1}{\sigma_a^2} + rac{n}{\sigma^2}, \quad rac{1}{\sqrt{2\pi}} \int\limits_0^{t(p_{
m mon})} e^{-s^2/2} \, ds = rac{p_{
m mon}}{2}.$$

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

$$\widehat{u}_{\rm B} = \arg \max_{x} \{ \ln p(X_1, \dots, X_n \mid u) + \ln p_{\rm anp}(u) \}.$$
 (2.7)

Пусть априорная информация о параметре u отсутствует. Тогда, согласно постулату Байеса, положим  $p_{\mathsf{anp}}(u) = \mathsf{const}$  и Б-оценка (2.7) примет вид  $\widehat{u}_\mathsf{B} = \mathsf{arg} \, \max \, l_n(X_1, \dots, X_n; u)$ , т. е.

Б-оценка будет совпадать с ММП-оценкой, основанной на выборке (1.1).

Теоретические основы метода Байеса представлены в монографиях [30, 38].

## 2. Минимаксный метод учета априорной информации

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

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

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

Таким образом, минимаксный метод выбирает такую оценку, которая предохраняет от возможно большего риска при изменении и 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

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

Нахождение ММ-оценки для произвольного априорного множества  $U_{\mathfrak{a}}$  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  $U_0-m$ -мерный эллипсоид (с центром в  $u_0$  и «радиуса»  $\rho$ , т.е.  $(u-u_0)^T B(u-u_0) \leqslant \rho^2$ , B— симметричная положительно определенная  $(m \times m)$ -матрица). Выберем квадратичную функцию потерь. Тогда MM-оценка (2.8) дается равенством

$$\hat{u}_{\text{MM}} = u_0 + \frac{1}{\sigma^2} \left( \frac{1}{\rho^2} B' + \frac{n}{\sigma^2} K_{\xi}^{-1} \right)^{-1} K_{\xi}^{-1} \left( \sum_{i=1}^{n} X_i - nu_0 \right), \quad (2.10)$$

где B'- модифицированная матрица B [92].

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

Если  $ho o \infty$ , то MM-оценка (2.10) также сближается с  $MM\Pi$ -оценкой  $\widehat{u}_{MM\Pi} = \overline{X}$  и тем самым «цена» априорной информации уменьшается. Если ho o 0, т. е. априорное множество

становится все уже, стягиваясь к центру  $u_0$ , то  $\widehat{u}_{\mathsf{MM}} \to u_0$ , а «цена» выборки (1.1) падает.

Теоретические основы ММ-оценивания представлены в монографии [92].

### 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

$$\widehat{u}_{\mathsf{OMM\Pi}} = \arg \max_{u \in U_a} L_n(X_1, \dots, X_n; u). \tag{2.11}$$

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 <sup>ВСПОЛЬЗ</sup>ОВАТЬ ЭКВИВАЛЕНТНОЕ представление

$$\widehat{u}_{\mathsf{OMM}\Pi} = \arg \min_{u \in U_a} -l_n(X_1, \dots, X_n; u). \tag{2.12}$$

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 [21].

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

$$\widehat{u}_{\mathrm{MM}\Pi} = \arg\min_{u \in E^m} -l_n(X_1, \dots, X_n; u) \in U_a.$$

Тогда  $\widehat{u}_{\text{ОММП}}=\widehat{u}_{\text{ММП}}$  и априорное множество (из-за своей «общирности») 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

$$-\operatorname{grad}_{u} l_{n}(X_{1}, \dots, X_{n}; u) + \alpha B(u - u_{0}) = 0, \tag{2.13}$$

где  $\alpha>0$  — множитель Лагранжа. Величина параметра  $\alpha$  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 известной ковариационной матрицей  $\sigma^2 K_{\xi}$ , выборка (1.1) повторная. Тогда

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

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

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

$$\widehat{u}_{\text{OMM}\Pi} = u_0 + \frac{1}{\sigma^2} \left( \alpha B + \frac{n}{\sigma^2} K_{\xi}^{-1} \right)^{-1} K_{\xi}^{-1} \left( \sum_{i=1}^{n} X_i - n u_0 \right). \tag{2.14}$$

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  $\Delta u$  оценок (2.10) и (2.14) соответственно равны

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

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)

$$D(\widehat{u}_{MM}) = \left(\frac{1}{\rho^2}B' + \frac{n}{\sigma^2}K_{\xi}^{-1}\right)^{-1}.$$
 (2.15)

Для ОММП-оценки (2.14)

$$D(\widehat{u}_{\mathsf{OMM\Pi}}) = (\alpha B + \frac{n}{\sigma^2} K_{\xi}^{-1})^{-1}.$$
 (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

$$\left(rac{1}{
ho^2}\,B' + rac{n}{\sigma^2}\,K_\xi^{-1}
ight)$$
 и  $\left(lpha B + rac{n}{\sigma^2}\,K_\xi^{-1}
ight)$ 

соответственно, предварительно обратив симметричную положительно определенную матрицу  $K_{\xi}$ . Одним из эффективных численных методов обращения симметричной положительно определенной матрицы A является метод квадратного корня (метод Холецкого), основанный на представлении  $A=S^TS$ , где S-Верхняя треугольная матрица.

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

$$\begin{split} \widehat{u}_{\text{MM}} &= \frac{d^2}{nd^2 + \sigma^2} \sum_{i=1}^n X_i + \left(1 - \frac{nd^2}{nd^2 + \sigma^2}\right) u_0, \\ \widehat{u}_{\text{OMMII}} &= \left\{ \begin{array}{l} \frac{1}{n} \sum_{i=1}^n X_i, & \left|\frac{1}{n} \sum_{i=1}^n X_i - u_0\right| \leqslant d, \\ u_0 + d \operatorname{sign}\left(\frac{1}{n} \sum_{i=1}^n X_i - u_0\right), & \left|\frac{1}{n} \sum_{i=1}^n X_i - u_0\right| > d. \end{array} \right. \end{split}$$

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  $-l_n(X_1,\dots,X_n;u)$  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  $\widehat{u}_{\mathsf{MM\Pi}}$  на априорное множество  $U_a$ .

# 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

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

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

$$L_o(u) = L(u) L_a(u),$$
 (2.18)

$$l_o(u) = l(u) + l_a(u).$$
 (2.19)

Следовательно, имеем равенство

$$I_o(u) = I(u) + I_a(u),$$
 (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 и определяется равенством

$$\widehat{u}_{\mathsf{OMM\Pi}} = \arg\max_{u} L_o(u). \tag{2.21}$$

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,  $l_a(u)=0$  — нулевая матрица и оценка (2.21) совпадает с  $MM\Pi$ 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 расматривать совместную ОММП-оценку всех параметров u,v:

$$(\widehat{u}_{\mathsf{OMM\Pi}}, \widehat{v}_{\mathsf{OMM\Pi}}) = \arg \max_{(u,v)} L_o(u,v).$$
 (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

$$\operatorname{grad}_{u} L_{o}(u) = 0, \tag{2.23}$$

$$\operatorname{grad}_{u} L_{o}(u, v) + \operatorname{grad}_{v} L_{o}(u, v) = 0$$
 (2.24)

соответственно.

Если на и, и 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 [21]. В частности, иногда некоторая часть параметров и (или и, и) по их содержательному смыслу неотрицательна. Тогда вместо задач на безусловный экстремум (2.21), (2.22) будем иметь задачи на условный экстремум:

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

$$(\widehat{u}_{\mathsf{OMM\Pi}}, \widehat{v}_{\mathsf{OMM\Pi}}) = \arg \max_{u,v \geqslant 0} L_o(u, v)$$
 (2.26)

соответственно.

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

В этом случае ОММП-оценка (2.21) дается равенством

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

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

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

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 и 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

$$K_{\widehat{u}_{OMMII}} = I_o^{-1}$$

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

$$\widehat{u}_{\mathsf{OMM\Pi}} = \arg\max_{u \in R_o} L_o(u). \tag{2.29}$$

Для решения задачи (2.29) можно использовать численные методы решения задач на условный экстремум [21]. 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 [48, 49].

#### 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  $p(x_1, ..., x_n; u)$  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 ж 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  $\{p(x_1,\ldots,x_n;u)\}$ . Плотность  $p_0(x_1,\ldots,x_n;u)$  называют устойчивой плотностью относительно класса  $\{p(x_1,\ldots,x_n;u)\}$ , а саму оценку  $\widehat{u}$  — робастной.

### 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  $\{p(x_1,\ldots,x_n;u\}$  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

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

rge

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

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

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

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

 $\dots dx_n$ 

Построим устойчивый метод оценивания параметра сдвига и 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

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

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

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  $\varphi(x) = (\ln p_0(x))^i$ . Робастная ММП-оценка (3.1)

удовлетворяет уравнению  $\sum_{i=1}^{n} \varphi(X_i - \widehat{u}) = 0.$ 

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

$$\widehat{\boldsymbol{u}} - \boldsymbol{u} \approx \frac{\sum_{i=1}^{n} \varphi(X_i - \boldsymbol{u})}{\sum_{i=1}^{n} \varphi'(X_i - \boldsymbol{u})} = \frac{\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i - \boldsymbol{u})}{\frac{1}{n} \sum_{i=1}^{n} \varphi'(X_i - \boldsymbol{u})} = \frac{\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i - \boldsymbol{u})}{\sum_{i=1}^{n} \varphi'(x - \boldsymbol{u}) p_{\boldsymbol{u}}(x - \boldsymbol{u}) dx}$$

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

$$D(p_0, p_u) = \int_{-\infty}^{\infty} \dots \int (\widehat{u} - u)^2 p_u(x_1 - u) \dots p_u(x_n - u) dx_1 \dots dx_n \approx \int_{-\infty}^{\infty} \dots \int \left(\sum_{i=1}^n \varphi(x_i - u)\right)^2 p_u(x_1 - u) \dots p_u(x_n - u) dx_1 \dots dx_n = \int_{-\infty}^{\infty} \dots \int \sum_{i=1}^n \varphi^2(x_i) p_u(x_1) \dots p_u(x_n) dx_1 \dots dx_n = \int_{-\infty}^{\infty} \dots \int \sum_{i=1}^n \varphi^2(x_i) p_u(x_1) \dots p_u(x_n) dx_1 \dots dx_n = \int_{-\infty}^{\infty} \dots \int \int_{-\infty}^{\infty} \varphi^2(x_i) p_u(x_i) dx = \int_{-\infty}^{\infty} \varphi^2(x_i) p_u(x_i) dx = \int_{-\infty}^{\infty} \varphi^2(x_i) p_u(x_i) dx = \int_{-\infty}^{\infty} \varphi^2(x_i) p_u(x_i) dx$$

$$\approx \frac{\int\limits_{-\infty}^{\infty} \varphi^{2}(x) p_{u}(x) dx}{n \Big(\int\limits_{-\infty}^{\infty} \varphi'(x) p_{u}(x) dx\Big)^{2}}$$

Итак.

$$D(p_0, p_u) \cong \frac{1}{n} \int_{-\infty}^{\infty} \left( \frac{p_0'(x)}{p_0(x)} \right)^2 p_u(x) \, dx / \left( \int_{-\infty}^{\infty} \left( \frac{p_0'(x)}{p_0(x)} \right)' p_u(x) \, dx \right)^2.$$

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

$$D(p_0, p_u) = \frac{1}{n} \int_{-\infty}^{\infty} \left( \frac{p'_0(x)}{p_0(x)}^2 \right) p_u(x) dx / \left( \int_{-\infty}^{\infty} \frac{p'_0(x)p'_u(x)}{p_0(x)} dx \right)^2$$

Используя неравенство Коши-Буняковского, получаем

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

$$\leqslant \int_{-\infty}^{\infty} \left(\frac{p_0'(x)}{p_0(x)}\right)^2 p_u(x) dx \cdot \int_{-\infty}^{\infty} \left(\frac{p_u'(x)}{p_u(x)}\right)^2 p_u(x) dx.$$

Поэтому

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

где

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

— информация Фишера о параметре сдвига и 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

$$\min_{p(x)\in\{p(x)\}} \int_{-\infty}^{\infty} \left(\frac{p'(x)}{p(x)}\right)^2 p(x) dx. \tag{3.3}$$

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  $H=\{p(x): p(x)=(1-\varepsilon)g(x)++\varepsilon h(x)\}$ , 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

истойчивая плотность

$$\mathbf{p}_{\mathbf{y}}(\mathbf{x}) = \begin{cases}
(1 - \varepsilon)g(x_0) \exp\{-\widetilde{K}|x - x_0|\}, & x < x_0, \\
(1 - \varepsilon)g(x), & x_0 \le x \le x_1, \\
(1 - \varepsilon)g(x_1) \exp\{-\widetilde{K}|x - x_1|\}, & x > x_1,
\end{cases} (3.4)$$

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

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

Доказательство теоремы Хьюбера основано на решении веклассической вариационной задачи

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

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  $x_0=-\widetilde{K}\sigma^2=-K\sigma$ ,  $x_1=K\sigma^2=K\sigma$ ,  $x_2=K\sigma$ ,

$$p_{y}(x) = \begin{cases} \frac{(1-\varepsilon)}{\sqrt{2\pi}\sigma} \exp\left\{\frac{K}{\sigma}(x+K\sigma) - \frac{K^{2}}{2}\right\}, & x < -K\sigma, \\ \frac{(1-\varepsilon)}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^{2}}{2\sigma^{2}}\right\}, & |x| \leq K\sigma, \\ \frac{(1-\varepsilon)}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{K}{\sigma}(x-K\sigma) - \frac{K^{2}}{2}\right\}, & x > K\sigma, \end{cases}$$
(3.5)

где K находится из равенства

$$\sqrt{\frac{\pi}{2}} \frac{1}{(1-\varepsilon)} = \int_{0}^{K} e^{-s^{2}/2} ds + \frac{1}{K} e^{-K^{2}/2}.$$

Параметр Хьюбера K однозначно определен по доле засорения  $\varepsilon$ , причем  $K(\varepsilon) \to \infty$  при  $\varepsilon \to +0$ ,  $K(\varepsilon) \to 0$  при  $\varepsilon \to 1-0$ ,  $K(\varepsilon) \to \infty$  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 [97]).

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

$$\begin{split} I_n(X_1,\dots,X_n;u) &= \sum_{i=1}^n \ln \, p_y(X_i-u) = n \ln \, \frac{(1-\varepsilon)}{\sqrt{2\pi}\,\sigma} \, + \\ &\quad + \frac{n_++n_-}{2}\, K^2 \, - \\ &\quad - \frac{K}{\sigma} \sum_{i\in I_+} (X_i-u) + \frac{K}{\sigma} \sum_{i\in I_-} (X_i-u) - \frac{1}{2\sigma^2} \sum_{i\in I_0} (X_i-u)^2, \end{split}$$
 The Magnesian

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  $n_+K\sigma-n_-K\sigma+\sum_{i\in I_0}(X_i-\widehat{u})=0$ . Последнее равенство перепишем в виде

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

где

$$\widehat{I}_{0} = \{i: |X_{i} - \widehat{u}| \leq K\sigma\}, \quad \widehat{I}_{+} = \{i: |X_{i} - \widehat{u}| > K\sigma\}, 
\widehat{I}_{-} = \{i: |X_{i} - \widehat{u}| < -K\sigma\},$$

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

Для решения нелинейного уравнения (3.6) удобно использовать итерационный процесс

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

ГДе

$$\begin{split} I_{0j} &= \{i: \ | X_i - \widehat{u}_j | \leq K\sigma \}, \quad I_{+j} &= \{i: \ X_i - \widehat{u}_j > K\sigma \}, \\ I_{-j} &= \{i: \ X_i - \widehat{u}_j < -K\sigma \}, \end{split}$$

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

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  $\frac{1}{n}\sum_{i=1}^{n}X_{i}$ .

При решении прикладных задач измерения являются неравноточными. В этих случаях члены  $X_i$  выборки (1.1) имеют различные дисперсии  $\sigma_i^2$ . 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) логариф-мическая функция правдоподобия дается равенством

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

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

$$l_n(X_1, \dots, X_n; u) = \sum_{i=1}^n \ln \frac{(1-\varepsilon)}{\sqrt{2\pi} \sigma_i} + \frac{n_+ + n_-}{2} K^2 - K \sum_{i \in I_+} \frac{X_i - u}{\sigma_i} + K \sum_{i \in I_-} \frac{X_i - u}{\sigma_i} - \frac{1}{2} \sum_{i \in I_0} \frac{(X_i - u)^2}{\sigma_i^2}.$$

9117

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

$$I_{0} = \{i: |X_{i} - u| \leq K\sigma_{i}\},$$

и вместо (3.7) получаем итерационный процесс

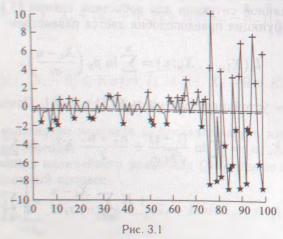
$$\widehat{u}_{j+1} = \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\right)^{-1} \left[\sum_{i \in I_{0j}} \frac{X_i}{\sigma_i^2} + \sum_{i \in I_{+j}} \frac{K\sigma_i + \widehat{u}_j}{\sigma_i^2} + \sum_{i \in I_{-j}} \frac{-K\sigma_i + \widehat{u}_j}{\sigma_i^2}\right],$$
(3.8)

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

Нахождение устойчивой минимаксной оценки Хьюбера относительно класса H предполагает знание доли засорения  $\varepsilon$ . 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  $\varepsilon$ , 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].



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 добав-

лением равномерно распределенных на интервале [-10, 10] величин. Горизонтальная пунктирная линия соответствует нулевому приближению  $\widehat{u}_0$ , сплошная линия — найденной в результате вычислений оценке. Символами  $(+, \bigstar)$  помечены значения с  $i \in I_+$  и  $i \in I_-$  соответственно.

#### 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 оценивание параметра сдвига с помощью усеченной средней

$$u = \overline{X}_{yc} = \frac{1}{n - 2m} \sum_{i=m+1}^{n-m} X_{(i)},$$
 (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  $\varepsilon$ , то можно взять  $m=[\varepsilon n/2]$ .

Модификацией усеченной средней является оценка Винзора

$$\widehat{u} = \overline{X}_{B} = \frac{1}{n} \left[ \sum_{i=m+1}^{n-m} X_{(i)} + m(X_{(m+1)} + X_{(n-m)}) \right].$$
 (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  $m=\lceil n/2 \rceil$  совпада-

ют с медианой выборки (1.1).

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

$$\widehat{u} = \sum_{i=m+1}^{n-m} \frac{X^{(i)}}{\sigma_{(i)}^2} \left( \sum_{i=m+1}^{n-m} \frac{1}{\sigma_{(i)}^2} \right)^{-1}, \tag{3.11}$$

$$\widehat{\mathbf{u}} = \left[ \sum_{i=m+1}^{n-m} \frac{X^{(i)}}{\sigma_{(i)}^2} + m \left( \frac{X^{(m+1)}}{\sigma_{(m+1)}^2} + \frac{X^{(n-m)}}{\sigma_{(n-m)}^2} \right) \right] \left( \sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^{-1}, \quad (3.12)$$

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:

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

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

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  $\widehat{u}_0 = \sum_{j=1}^n \frac{X_j}{\sigma_j^2} \Big(\sum_{j=1}^n \frac{1}{\sigma_j^2}\Big)^{-1}.$ 

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

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

где b, не зависят от u. Тогда ММП-оценка минимизирует сумму  $\sum \rho\left(\frac{X_i-u}{\sigma}\right)$ . В частности, для минимаксной робастной оценки Xьюбера из (3.8) имеем

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

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

Пусть задана четная выпуклая положительная функция ho(s).  $\varrho(0) = 0$ . М-оценкой называется значение u, которое минимизирует функционал  $\sum_{i=1}^{n} 
ho\left(rac{X_{i}-u}{\sigma_{i}}
ight),$ 

$$\pi \sum_{i=1}^{n} \rho \left( \frac{1}{\sigma_i} \right),$$

$$\widehat{u}_{\mathsf{M}} = \operatorname{Arg} \min_{u} \sum_{i=1}^{n} \rho\left(\frac{X_{i} - u}{\sigma_{i}}\right). \tag{3.13}$$

М-оценка однозначно определяется функцией ho(s).

Чтобы М-оценка была робастной по отношению к большим выбросам, если реперный закон распределения нормальный, необходим более медленный рост функции ho(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  $ho(s) = |s|^p$ , 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

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

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

Пример 3.3. Робастная оценка, порожденная функцией

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

 $\lambda \geq 0$ , называется оценкой Рамсея. Для малых |s|  $ho(s) \sim s^2$ ;  $a \, 
ho(s) o 1/\lambda^2$  при  $|s| o \infty$ . Параметр  $\lambda$  выберем так, чтобы точка

перегиба функции  $\rho(s)$  равнялась 3. Тогда  $\lambda=1/3$ . При увеличении  $\lambda$  «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

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

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

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

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

$$\sum_{i=1}^n (X_i - u) W_i = 0$$
, где  $W_i = \frac{1}{\sigma_i (X_i - u)} \, 
ho' \left( rac{X_i - u}{\sigma_i} 
ight)$ .

Определим итерационный процесс

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

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  $(o(s) = |s|^p)$  оба метода совпадают.

 $\hat{P}_{ACCMOTPИМ}$  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

$$p_{y}(x) = \begin{cases} \frac{1}{d} \frac{a}{(1+a)} \cos^{2} \frac{bx}{d}, & |x| \leq d, \\ \frac{1}{d} \frac{a}{(1+a)} \cos^{2} b \exp\left\{-2a\left(\frac{|x|}{d} - 1\right)\right\}, & |x| > d, \end{cases}$$
(3.14)

где а и в находятся из системы

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

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 [48, 82, 95, 97, 104].

#### 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).

## 1. Восстановление функции распределения

Достаточно просто по повторной выборке (1.1) дать оценку функции распределения F(x) случайной величины  $\xi$  (в дальнейшем, если не сделано специальных оговорок, предполагается, что  $\xi$  — 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  $\eta(x) = \left\{ egin{array}{ll} 1, & x > 0, \\ 0, & x < 0, \end{array} 
ight. - единичная функция. \end{array} 
ight.$ 

Теорема 4.1 (Гливенко-Кантелли).

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

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). Действи-

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

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

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

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

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

Имеем

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  $n o \infty$ . Далее,

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

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

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

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:

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

справедливое для любого закона распределения при достаточно большом n.

Обозначим  $K_n(u)$  функцию распределения случайной величины  $U_n = \sqrt{n}$  тах  $|F(x) - F_n(x)|$ . 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 и табулированы [16]. Используя таблицы  $K_n(u)$ , можно находить доверительные полосы для F(x):  $|F(x) - F_n(x)| \le u(p_{\text{дов}})/\sqrt{n}$  для  $x \in (-\infty, \infty)$ , где  $p_{\text{дов}} = K_n(u(p_{\text{дов}}))$ . Существует

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

(Колмогоров). Функция K(u) также табулирована [16]. Поэтому при большом n (практически, если n > 100) для нахождения  $u(p_{200})$  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.

# 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  $\Delta x_j$  (мы сохраняем за интервалом и его длиной одно обозначение). Обозначим  $n_j$  количество членов выборки (1.1), попавших в интервал  $\Delta x_j$ ,  $\sum_{j=1}^{N} n_j = n$ , Частоты  $\frac{n_j}{n}$  являются оценками вероятности  $p_j$  попадания значений  $\xi$  в интервал  $\Delta x_j$ , 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  $\Delta x_j$ , т. е.  $\frac{1}{\Delta x_j}\int\limits_{\Delta x_j} p_{\xi}(x)\,dx$ . Итак, за оценку плотности принимается

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

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

$$\int_{-\infty}^{\infty} \widehat{p}_{\Gamma}(x) dx = \sum_{j=1}^{N} \int_{\Delta x_{j}} \frac{n_{j}}{n \Delta x_{j}} dx = \sum_{j=1}^{N} \frac{n_{j}}{n} = 1.$$

Основной трудностью при построении гистограммы является выбор интервалов разбиения  $\Delta x_j$ . Рекомендуется интервалы  $\Delta r_j$  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  $\chi^2_{N-1}$  (теорема 1).

Статистику (4.3) можно представить в эквивалентном виде

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

Поэтому доверительная полоса для оценки (4.2) имеет вид

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

где

$$p_{ exttt{дов}} = \int\limits_{0}^{t_{N}(p_{ exttt{дов}})} p_{\chi^2_{n-1}}(t) \, dt.$$

Существуют стандартные программы построения гистограмм.

#### 3. Восстановление плотности распределения методом Розенблатта-Парзена

Дифференцируя эмпирическую функцию распределения, получаем оценку

 $\widehat{p}_{00}(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - X_i),$ 

где  $\delta(x)$  — функция Дирака. Эта оценка плотности неприемлема, поскольку является обобщенной функцией. Модифицируем оценку  $\widehat{p}_{06}$ , 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:

$$1^{\circ} \int_{-\infty}^{\infty} \psi_n(x-s) \, p_{\xi}(s) \, ds \to p_{\xi}(x) \, npu \, n \to \infty;$$
$$2^{\circ} \frac{1}{n^2} \max \psi_n(x) \to 0 \, npu \, n \to \infty.$$

Доказательство. Имеем

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

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

$$\begin{split} \mathbf{M} \hat{p}^{2}(x) &= \frac{1}{n^{2}} \sum_{i,j=1}^{n} \mathbf{M} \, \psi_{n}(x - X_{i}) \psi_{n}(x - X_{j}) = \\ &= \frac{1}{n^{2}} \int_{-\infty}^{\infty} \psi_{n}^{2}(x - s) p_{\xi}(s) \, ds + \frac{n-1}{n} \left( \int_{-\infty}^{\infty} \psi_{n}(x - s) p_{\xi}(s) \, ds \right)^{2}. \end{split}$$

Но

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

при  $n \to \infty$ , а

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

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

 $h_n o 0.\ h_n^2 \, n o \infty$  при  $n o \infty.$  Тогда последовательность функций

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

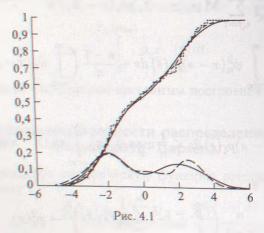
удовлетворяет условиям теоремы 4.2. Если

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

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.



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 [18, 29, 69,

125. 127].

#### 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  $\xi$  — 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

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

Имеем

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

Поэтому

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

 $lpha_j$  — начальный момент j-го порядка случайной величины  $\xi$ . Если начальные моменты  $lpha_j$  заменить их оценками  $\widehat{lpha}_j$  (на-

пример, оценками метода моментов 
$$\widehat{lpha}_j = rac{1}{n} \sum_{i=1}^n X_i^j$$
) и в (4.5)

ограничиться конечным 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  $\xi$  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  $\xi$  рассматривают нормированную случайную величину  $\xi_0 = (\xi - \alpha_1)/\sigma_\xi$  и проекционная оценка имеет вид

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

где 
$$\widehat{p}_0(y)=p_0(y)\sum_{k=0}^m \widehat{C}_{0k}q_k(y),$$
  $\widehat{C}_{0k}=\sum_{j=0}^k a_{kj}\widehat{\mu}_j,$   $\mu_j$  — нормирован-

ный 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

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

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  $\mu_k=\mu_k'/\sigma^k$ ,  $\mu_k'$  — центральный момент случайной величины  $\xi$ , и если  $\xi$  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  $\xi$  — полуось  $x\geqslant 0$  и в качестве реперного закона распределения взят экспоненциальный 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

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

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

Введем случайные величины  $Y_k=arphi_k(\xi)$ . 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 раз-

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

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 Фк г плотности распределения. Тогда плотность (4.11) представляет собой модель смеси с неизвестными весами  $C_k \in [0,1],$ 

 $\sum C_k =$  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 [18, 69, 99. 100].

#### 5. Восстановление плотности распределения регуляризованным методом гистограмм

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

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

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  $\overline{F}(x)$ , 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 спомощью 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

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

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

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

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

Поскольку F(x) неизвестна, матрицу  $K_w$  заменим ее оценкой  $\hat{\mathbb{K}}_w$ , 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 внтегральной теоремы Муавра-Лапласа, так как  $n_j$  — дискретные случайные величины, распределенные по биномиальному закону. Если  $n_j/n\to0$  при  $n\to\infty$ , то справедлива асимптотическая формула

$$\widehat{K}_{w} = \operatorname{diag}\left(\frac{n_{1}}{n^{2}}, \dots, \frac{n_{N}}{n^{2}}\right) (I + o(n)),$$

 $\mathbb{R}$  элементы матрицы  $o\left(n
ight)$  стремятся к нулю при  $n o\infty$ .

Пусть априорная информация об искомой плотности имеет втерминированный характер:  $p_{\xi}(x) \in G$ , где G — некоторое множество плотностей задаваемого класса (множество G может говокупностью равенств и неравенств с вхождением  $g_{\xi}(x)$ ).

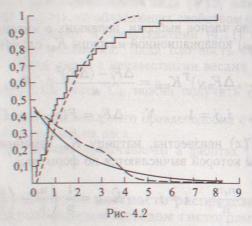
Определим ОММП-оценку плотности

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

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 [66, 84] или с помощью метода упорядоченной минимизации риска [6, 25].

На рис. 4.2 представлен пример восстановления плотности распределения с учетом априорной информации. В качестве системы 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 [79, 96].

### 6. Метод корневой оценки плотности распределения

В работе [15] 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 ncu-функцию  $\psi(x)$ , связанную с искомой плотностью p(x) равенством

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

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

$$\int_{-\infty}^{\infty} p(x) \, dx = 1$$
 следует равенство

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

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:

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

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

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

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

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

Для решения системы (4.20) в [15] предложен итерационный процесс

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

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

Поскольку из равенства (4.19) следует также равенство

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

2

$$\frac{1}{n}\sum_{k=1}^{n}\frac{\varphi_{i}(X_{k})\varphi_{j}(X_{k})}{p(X_{k})}\to\mathbf{M}\frac{\varphi_{i}(\xi)\varphi_{j}(\xi)}{p(\xi)}=\int_{-\infty}^{\infty}\frac{\varphi_{i}(x)\varphi_{j}(x)}{p(x)}p(x)\,dx=\delta_{ij}.$$

то полученные оценки являются состоятельными.

В работе [15] 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

$$arphi_i(x) = rac{1}{\sqrt{2^i i! \sqrt{\pi}}} H_i(x) e^{-x^2/2}, \quad i = 0, 1, \dots,$$

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.

## 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  $\omega^2$ .

Критерий согласия Колмогорова использует рассмотренную

ранее в разделе (1) случайную величину

$$V_n = \max_{x} |F(x) - F_n(x)|,$$
 (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 последователь-

ность 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  $V_n>V_{\rm крит}(n,p)$ , где  $V_{\rm крит}(n,p)$  — критическое значение величины (5.1), зависящее от объема выборки и заданного уровня значимости. Числовые значения  $V_{\rm крит}(n,p)$  можно найти в [16].

При применении критерия  $\omega^2$  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  $\omega_n^2$  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  $\omega_n^2 > \omega_{\rm крит}(n,p)$ , где  $\omega_{\rm крит}(n,p)$  — критическое значение величины (5.3), зависящее от объема выборки

и уровня значимости. Числовые значения  $\omega_{\text{крит}}(n,p)$  можно найти в [16].

Для проверки гипотезы H о соответствии тестируемой плотности распределения p(x) выборке (1.1) часто используют крите-

рий  $\chi^2$  (хи-квадрат) Пирсона.

Пусть область изменения значений случайной величины  $\xi$  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

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

**Теорема 5.1** (Пирсон). Пусть:  $1^{\circ}$  выборка (1.1) повторная;  $2^{\circ}\sum_{j=1}^{N}p_{j}=1$ . Тогда статистика  $\widehat{Z}_{n}$  из (5.4) при  $n\to\infty$  имеет  $\chi^{2}$ -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  $\chi^2$ -распределения найдем такое  $z(p_{\text{дов}})$ , что

 $P_{\mathrm{AN-I}}^{2}(z)\,dz=p_{\mathrm{дов}}.$  Если  $\widehat{Z}_{n}< z(p_{\mathrm{дов}})$ , то считают, что

гипотеза H подтверждается  $100\,p_{\text{дов}}$ -процентным уровнем доверия. Если же  $Z_n>z(p_{\text{дов}})$ , то считают, что гипотеза H о соответствии плотности распределения p(x) выборке (1.1) не подтверждается.

#### 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  $\chi^2$  Фишера. Пусть, как и выше, область изменения значений случайной величины  $\xi$  разбита на N интервалов  $\Delta x_j$ ,  $j=1,\ldots,N$ . Обозна-

чим  $n_j$  — число членов выборки (1.1), попавших в интервал  $\Delta x_j$ ,  $\widehat{p}_j = \int\limits_{x_j}^{x_{j+1}} p(x;\widehat{u})\,dx$  — вероятности попадания в интервал  $\Delta x_j$ ,

вычисленные по гипотетической оцененной плотности  $p(x; \overline{u})$ . Определим статистику

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

**Теорема 5.2** (Фишер). Пусть оценка  $\widehat{u}$  n-мерного параметра и является асимптотически эффективной и асимптотически нормальной (например,  $\widehat{u}=\widehat{u}_{MM\Pi}$ ). Тогда статистика  $\widehat{Z}_n$  при  $n\to\infty$  имеет  $\chi^2$ -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  $\chi^2$ -распределения найдем такое  $z(p_{\text{дов}})$ , что

$$\int\limits_{0}^{z(p_{\mathrm{dob}})}p_{\chi_{N-m-1}^{2}}(z)\,dz=p_{\mathrm{dob}}.$$

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

$$X_{k1}, \dots, X_{kn_k}, \qquad k = 1, \dots, l,$$
 (5.6)

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

рок (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  $\Delta x_j$ . Введем статистику

$$\widehat{Z}_{ln} = \sum_{k=1}^{l} n_k \sum_{j=1}^{N} \frac{1}{\widehat{p}_j} \left( \frac{n_{kj}}{n_k} - \widehat{p}_j \right).$$
 (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  $\chi^2$ -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

$$\int\limits_{0}^{Z(p_{ exttt{AOB}})}p_{\chi^2_{l(N-1)-m}}(z)\,dz=p_{ exttt{AOB}}.$$

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

стики (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 
$$p_j = \int_{x_j}^{x_{j+1}} p(x) dx$$
.

Закон распределения статистики (5.8) при  $n_k \to \infty$ ,  $k=1,\ldots,l$ , стремится к  $\chi^2$ -распределению с l(N-l) степенями свободы. Поэтому критическая граница  $z(p_{\rm дов})$ 

проверки гипотезы однородности выборок (5.6) определяется из равенства

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

Проверка гипотез о законе распределения представлена в работах [32, 38, 39, 43, 46].

#### 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 [83].

### 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  $\gamma$ . Лействительно, пусть задана произвольная случайная величина  $\xi$  с плотностью p(x) и функцией распределения

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

#### Теорема 6.1. Выполнение равенства

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

 $\mu_{\rm e}$ обходимо и достаточно для того, чтобы F(x) была функцией распределения случайной величины  $\xi$ .

Доказательство. Пусть F(x) — функция распределения случайной величины  $\xi$ . Тогда случайная величина  $\gamma$ , определенная равенством (6.1), распределена равномерно на промежутке [0,1]. Действительно, функция распределения  $F_{\gamma}(y)$  случайной величины  $\gamma$  определится равенством

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

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

Итак,

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

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

Пусть  $\gamma$  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  $F_{\xi}(x) = \mathbf{P}\left\{\xi < x\right\} = \mathbf{P}\left\{F^{-1}(\gamma) < x\right\} = \mathbf{P}\left\{\gamma < F(x)\right\} = F_{\gamma}(F(x)) = F(x)$ , поскольку  $F(x) \in [0,1]$ . Теорема доказана.

Из теоремы 6.1 следует, что повторная выборка

$$\gamma_1, \ldots, \gamma_n$$
 (6.2)

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

**Пример 6.1.** Пусть  $\xi$  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  $\xi$  распределена по экспоненциальному закону. Имеем:  $F_{\xi}(x)=1-e^{-\lambda x}, \ x\geqslant 0.$  Поэтому  $1-e^{-\lambda X_i}=\gamma_i,$ 

 $X_i = rac{1}{\lambda} \ln rac{1}{1-\gamma_i}$  . The first interest of the supplementary  $X_i = \frac{1}{\lambda} \ln \frac{1}{1-\gamma_i}$  .

Поскольку случанная величина  $1-\gamma_i$  также равномерно распределена на [0, 1], то повторную выборку (1.1) можно получить и по  $X_i = rac{1}{\lambda} \ln rac{1}{\gamma_i}$  , and  $X_i = rac{1}{\lambda} \ln rac{1}{\gamma_i}$ формуле

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 связано 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 эффек-

тивные 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  $\gamma_i$  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 про-

межутке [0, 1] со значениями на этом же промежутке.

Например. в методе вычетов Лемера R(x) = D(qx), 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

$$\gamma_{i+1} = D(g\gamma_i). \tag{6.4}$$

Есян в (6.4)  $\gamma_0 = m_0/M$ , где  $m_0$ , M, g — целые взаимно простые числа, то все  $\gamma_i$  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  $\gamma_{i+1},\ \gamma_i$ , 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 [83].

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  $\xi$  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

$$F_{\xi}(z) = \mathbf{P} \{ \xi < \mathbf{z} \} = \mathbf{P} \{ \xi^{(1)} < z \mid \xi^{(2)} < p(\xi^{(1)}) \} =$$

$$= \frac{\mathbf{P} \{ (\xi^{(1)} < z) \cap (\xi^{(2)} < p(\xi^{(1)})) \}}{\mathbf{P} \{ \xi^{(2)} < p(\xi^{(1)}) \}}.$$

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

$$P\{(\xi^{(1)} < z) \cap (\xi^{(2)} < p(\xi^{(1)}))\} =$$

$$= \int_{a}^{z} dx \int_{0}^{p(x)} \frac{1}{(b-a)C} dy = \frac{1}{C(b-a)} \int_{a}^{z} p(x) dx.$$

Знаменатель равен вероятности попадания случайной точки  $(\xi^{(1)}, \xi^{(2)})$  ниже кривой y = p(x) и поэтому

$$\mathbf{P}\left\{\xi^{(2)} < p(\xi^{(1)})\right\} = \int_{a}^{b} dx \int_{0}^{p(x)} \frac{1}{(b-a)C} dy = \frac{1}{C(b-a)}.$$

Итак,

$$F_{\xi}(z) = \left[\int_{a}^{z} \frac{p(x)}{C(b-a)} dx\right] / \frac{1}{C(b-a)} = \int_{a}^{z} p(x) dx,$$

т.е. p(x) — плотность вероятностей случайной величины  $\xi$ . Терема доказана.

Для получения одной реализации случайной величины  $\xi$  по методу Неймана требуются по крайней мере две независимые реализации случайной величины  $\gamma$ , причем, если случайная точка  $(\xi^{(1)},\xi^{(2)})$  попадает выше кривой y=p(x), то реализации  $\xi$  не будет получено. Поэтому для снижения числа пар  $(\xi^{(1)},\xi^{(2)})$ . для которых не происходит реализаций  $\xi$ , необходимо брать C равным наименьшему возможному значению  $C=\max p(x)$ .

Для получения выборки (1.1) для некоторых стандартных законов распределения используют специальные приемы, учитывающие специфику рассматриваемых распределений.

Пусть  $\xi$  — нормально распределенная случайная величина. Обозначим  $\xi_0 = (\xi - \mathbf{M}\,\xi)/\sigma_\xi$  нормированную случайную величину. Пусть

$$\xi_0^{(N)} = \sum_{j=1}^N \left( \gamma_j - \frac{1}{2} \right) / \sqrt{\frac{N}{12}} = \sqrt{\frac{3}{N}} \sum_{j=1}^N (2\gamma_j - 1)$$

— нормированная сумма N независимых случайных величин, равномерно распределенных на [0,1]. Согласно центральной предельной теореме Линдеберга—Леви закон распределения  $\mathbb{R}^{(N)}$  при  $N \to \infty$  стремится к  $\mathcal{N}(0,1)$ .

Поэтому при достаточно большом N закон распределения  $\xi^{(M)} = \sigma_{\xi} \xi^{(M)} + M \xi$  будет близок к  $\mathcal{N}(M \xi, \sigma_{\xi}^2)$ . Следовательно, при достаточном большом N одну реализацию случайной величины  $\xi$  можно подсчитывать по N независимым реализациям  $\gamma$ 

согласно формуле

$$X_i = \mathbf{M}\,\xi + \sigma_{\xi}\sqrt{\frac{3}{n}}\,\sum_{j=1}^{N}(2\gamma_j - 1).$$
 (6.5)

На практике ограничиваются N=12, и тогда формула (6.5)

$$X_i = \sigma_{\xi} \sum_{j=1}^{12} \gamma_j + \mathbf{M} \, \xi - 6\sigma_{\xi}.$$
 (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.  $\xi_N$  подчиняется  $\gamma$ -распределению. Действительно, для N=0утверждение доказано ранее. Предположим, что утверждение справедливо для N. Тогда

$$\xi_{N+1} = -\frac{1}{\lambda} \ln (\gamma_1 \dots \gamma_{N+2}) = \xi_N + \xi_0.$$

$$\xi_{N+1} = -\frac{1}{\lambda} \ln \left( \gamma_1 \dots \gamma_{N+2} \right) = \xi_N + \xi_0.$$
 Поскольку  $\xi_N$  и  $\xi_0$  независимы, 
$$p_{\xi_{N+1}}(x) = \int\limits_{-\infty}^{\infty} p_{\xi_N}(x-s) p_{\xi_0}(s) \, ds =$$
 
$$= \int\limits_{0}^{x} \frac{\lambda^{N+1}}{N!} \left( x-s \right)^N e^{-\lambda (x-s)} \lambda e^{-\lambda s} ds =$$
 
$$= \frac{\lambda^{N+2}}{N!} e^{-\lambda x} \int\limits_{0}^{x} (x-s)^N \, ds = \frac{\lambda^{N+2}}{(N+1)!} \, x^{N+1} e^{-\lambda x}.$$

Следовательно, если  $\xi$  подчиняется  $\gamma$ -распределению с плотностью вероятностей  $p_{\xi}(x)=rac{\lambda^{N+1}}{N!}\,x^Ne^{-\lambda x},\;x>0,$  то по N+1

независимым реализациям  $\gamma$  находим одну реализацию  $\xi$  согласно формуле  $X_i = -\frac{1}{\gamma} \ln{(\gamma_1 \dots \gamma_{N+1})}$ .

Рассмотрим теперь m-мерную нормально распределенную случайную величину с вектором математического ожидания  $a=(a_1,\dots,a_m)^T$  и ковариационной матрицей  $K_{\mathcal E}$ . Представим  $K=QQ^T$ , 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  $\eta$  распределена нормально. Имеем:  $\mathbf{M}\,\eta=Q^{-1}(\mathbf{M}\,\xi-a)=0$ , ковариационная матрица  $K_\eta=Q^{-1}K_{\mathcal E}(Q^{-1})^T=Q^{-1}Q\,Q^T(Q^T)^{-1}=I$  (единичная матрица). Следовательно,  $\eta=(\eta_1,\dots,\eta_m)^T$ , где  $\eta_j,j=1,\dots,m$  — независимые нормально распределенные величины с законом распределения  $\mathcal N$  (0,1) и одна реализация  $Y_i$  случайной величины  $\eta$  состоит из m независимых реализаций скалярной случайной величины с законом  $\mathcal N$  (0,1).

с законом  $\mathcal{N}(0,1)$ . Из вышесказанного следует, что реализация  $X_i$  случайной величины  $\xi$  может быть получена по формуле

$$X_i = a + QY_i. (6.7)$$

# 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 и решение задачи (и может быть скаляром, вектором, функцией, вектор-функцией и т. д. в зависимости от конкретной задачи). Подберем такую случайную величину  $\xi$ , что  $\mathbf{M}\,\xi=u$ . Тогда, осуществив достаточно большое число реализаций величины  $\xi$  и получив выборку ее значений  $X_1,\ldots,X_n$ , получают оценку  $\widehat{u}$ , например взяв в качестве  $\widehat{u}$  среднеарифметическую  $\widehat{u}=\frac{1}{n}\sum_{i=1}^n X_i$ .

Основная трудность состоит в подборе такой случайной величины  $\xi$ , что  $\mathbf{M}\,\xi=u$ . Конечно, подбор  $\xi$  неоднозначен и, если имеется возможность выбора, то следует отбирать случайную величину  $\xi$  с наименьшим разбросом относительно среднего значения u и учитывать (часто в первую очередь) вычислительные затраты на реализацию каждого значения  $X_i$  величины  $\xi$ . В конечном итоге выбор  $\xi$  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

$$J = \int_{G} f(x) dx, \tag{6.8}$$

где x-m-мерный вектор. Введем m-векторную случайную величину  $\eta$  с плотностью  $p(x)>0,\ x\in G$ , и скалярную случайную величину  $\xi=f(\eta)/p(\eta)$ . Имеем

$$\mathbf{M}\,\xi = \int_C \frac{f(x)}{p(x)} \, p(x) \, dx = J.$$

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  $\eta$ . 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:  $\mathbf{D}\hat{J} = -\mathbf{D}\xi$ . Таким образом, следует выбирать  $\xi$  с наименьшей дисперсией, или, что то же самое, с наименьшим  $\mathbf{M}\,\xi^2$ . Имеем

$$\mathbf{M} \, \xi^2 = \int_G \frac{f^2(x)}{p^2(x)} \, p(x) \, dx = \int_G \frac{f^2(x)}{p(x)} \, dx.$$

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

$$\left(\int_{G} |f(x)| dx\right)^{2} = \left(\int_{G} \frac{|f(x)|}{\sqrt{p(x)}} \sqrt{p(x)} dx\right)^{2} \leqslant$$

$$\leqslant \int_{G} \left(\frac{|f(x)|}{\sqrt{p(x)}}\right)^{2} dx \int_{G} \left(\sqrt{p(x)}\right)^{2} dx = \int_{G} \frac{f^{2}(x)}{p(x)} dx.$$

Положим

$$p(x) = \frac{|f(x)|}{\int_{G} |f(x)| \, dx}.$$
(6.10)

Имеем

$$\mathbf{M}\,\xi^2 = \int\limits_G \frac{f^2(x)}{p(x)}\,\mathrm{d}x = \int\limits_G \frac{f^2(x)}{|f(x)|}\,\mathrm{d}x\,\int\limits_G |f(x)|\,\mathrm{d}x = \left(\int\limits_G |f(x)|\,\mathrm{d}x\right)^2.$$

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  $\int f(x) \ dx$ , 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

 $p(x) = |f_a(x)| \cdot \left( \int_G |f_a(x)| \, dx \right)^{-1}$ 

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.

$$\int_{G} \frac{(f(x) - f_0(x))^2}{p(x)} dx \leqslant \varepsilon,$$

где arepsilon мало, и интеграл  $\int\limits_C f_0(x)\,dx=J_0$  известен. Введем скалярную случайную величину

$$\xi = J_0 + \frac{f(\eta) - f_0(\eta)}{p(\eta)}$$

и оценку

$$\widehat{J} = J_0 + \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i) - f_0(X_i)}{p(X_i)}.$$
(6.11)

Имеем

$$\mathbf{M}\,\xi=J,$$

$$D\xi = D\left(\frac{f(\eta) - f_0(\eta)}{p(\eta)}\right) = \int_G \frac{(f(x) - f_0(x))^2}{p(x)} dx - (J - J_0)^2 \leqslant \varepsilon^2.$$

Поэтому для дисперсии оценки справедливо неравенство  $\mathbf{D}\widehat{J}\leqslant \leqslant \varepsilon^2/n$ . 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

 $\varphi(x) = \int_{\Omega} K(x, x') \varphi(x') dx' + f(x)$ 

методами Монте-Карло.

Введем линейный интегральный оператор

$$K\varphi(x) = \int_G K(x, x')\varphi(x') dx'.$$

Обозначим  $K^{j}\varphi$  результат j-кратного применения оператора K $\kappa$  функции  $\varphi(x)$  и

$$(f,\varphi) = \int_G f(x)\varphi(x) dx$$

— скалярное произведение функций f(x) и arphi(x). Таким образом,  $(f,K^{\prime}arphi)$  — скалярное произведение функции f(x) и функции  $K^{j} \varphi$ , полученной в результате j-кратного применения оператора K к функции  $\varphi(x)$ . Например,

$$K^2 arphi(x) = \int\limits_G K(x,x') \int\limits_G K(x',x'') arphi(x'') \, dx'' \, dx' \equiv$$
 
$$\equiv \int\limits_G K_2(x,x') arphi(x') \, dx',$$
где

$$K_2(x,x') = \int_G K(x,x'')K(x'',x') dx''$$

— дважды итерированное ядро. Для произвольного j имеем

$$K^{j}\varphi(x) = \int_{G} K_{j}(x, x')\varphi(x') dx',$$

$$K_j(x, x') = \int_G K_{j-1}(x, x'') K(x'', x') dx''$$

— итерированные ядра.

Выберем в G такие плотность p(x)>0 и условную плотность  $p(x \mid x') > 0$ , 4TO  $\int p(x) dx = \int p(x \mid x') dx = 1$ .

Условную плотность  $p(x\mid x')$  называют плотностью вероятностей перехода из точки x' в точку x и часто обозначают  $p(x' \rightarrow x)$ .

Определим в G случайную траекторию

$$T_k = \{x_0 \to x_1 \to \dots \to x_k\},\,$$

где  $x_0$  — реализация случайной величины  $\xi_0$  с плотностью  $p(x),\ x_j$  — реализация случайной величины  $\xi_j$  с плотностью  $p(x_j\mid x_{j-1}),\ j=1,\ldots,k.$ 

Введем совокупность скалярных статистик

$$w_j = \frac{K(x_0, x_1)}{p(x_1 \mid x_0)} \frac{K(x_1, x_2)}{p(x_2 \mid x_1)} \dots \frac{K(x_{j-1}, x_j)}{p(x_j \mid x_{j-1})},$$

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  $\theta_j = w_j \frac{\varphi(x_0)}{p(x_0)} f(x_j)$ . Покажем, что  $\mathbf{M} \, \theta_j = (\varphi, K^j f)$ . Действительно,

$$\begin{split} \mathbf{M}\theta_{j} &= \mathbf{M} \frac{\varphi(x_{0})}{p(x_{0})} \frac{K(x_{0}, x_{1})}{p(x_{1} \mid x_{0})} \dots \frac{K(x_{j-1}, x_{j})}{p(x_{j} \mid x_{j-1})} f(x_{j}) = \\ &= \int_{G} \varphi(s_{0}) \, ds \int_{G} \dots \int_{G} K(s_{0}, s_{1}) \dots K(s_{j-1}, s_{j}) f(s_{j}) \, ds_{1} \dots ds_{j} = \\ &= \int_{G} \varphi(s_{0}) K^{j} f(s_{0}) \, ds = (\varphi, K^{j} f). \end{split}$$

 $\mathsf{Т}\mathsf{a}\mathsf{k}\mathsf{u}\mathsf{m}$  образом, реализовав n случайных траекторий

$$T_k = (T_{k1}, \dots, T_{kn}),$$

получаем оценки  $u_j$  скалярных произведений  $u_j=(arphi,K^jf)$  согласно формулам

 $\widehat{u}_j = \frac{1}{n} \sum_{i=1}^n \theta_{ji},\tag{6.13}$ 

где  $\theta_{ji},\,i=1,\ldots,n$  — реализации случайной величины  $\theta_j$  на i-й грасктории  $T_{ii}$ .

С помощью оценок (6.13) скалярных произведений  $(\varphi,K^jf)$  можно получить оценки значений функции  $K^jf(x_0)$ . Действительно, пусть  $\varphi(x)=\delta(x-x_0)$ . Тогда  $(\delta(x-x_0),K^jf(x))=K^jf(x_0)$ . Возьмем в качестве  $p(x)=\delta(x-x_0)$ , т. е. все траектории  $T_{ji}$  «стартуют» из фиксированной точки  $x_0$  и  $\theta_{j0}=w_jf(x_j)$ . Имеем: М  $\theta_{i0}=K^jf(x_0)$ .

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:

$$\varphi(x) = f(x) + Kf(x) + K^{2}f(x) + \dots$$
 (6.14)

Ограничимся в (6.14) конечным числом слагаемых. Тогда в фиксированной точке  $x=x_0$ 

$$\varphi(x_0) \approx f(x_0) + Kf(x_0) + \dots + K^k f(x_0)$$
 (6.15)

и приближенные значения слагаемых  $K^jf(x_0),\ j=1,\ldots,k$ , правой части (6.15) могут быть оценены по формулам (6.13), где  $\theta_{ji},\ i=1,\ldots,n$ , — реализации случайной величины  $\theta_j$  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  $K^jf(x_0)$  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 рассматри-

ваться 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

$$Au = f_T, (7.1)$$

где A — заданная  $(n \times m)$ -матрица. Заметим, что в § 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  $\varepsilon$  неизвестны, известен лишь закон распределения случайного вектора  $\varepsilon$  или даже не сам закон, а только некоторые его характеристики.

## 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  $\varphi_1(x),\dots,\varphi_m(x)$  — известные функции,  $u_1,\dots$ 

...,  $u_m$  — 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  $\varepsilon_i$ . Таким образом, имеем систему равенств

$$y_i - \bar{\varepsilon}_i = \sum_{j=1}^m u_j \varphi_j(x_i), \qquad i = 1, \dots, n.$$
 (7.3)

Введем векторы

$$f = (y_1, \dots, y_n)^T$$
,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ ,  $u = (u_1, \dots, u_m)^T$ 

и  $(n \times m)$ -матрицу A,  $A_{ij} = \varphi_j(x_i)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ . Тогда система равенств (7.3) запишется в эквивалентном виде (7.2).

**Пример 7.2.** Пусть между величинами y и  $x_1, \dots, x_m$  существует функциональная зависимость вида  $y = \sum\limits_{j=1}^m u_j x_j$ , 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  $y_i,\ v_i=1,\ldots,v_i$  причем измерения  $v_i,\ v_i=1,\ldots,v_i$  погрешности  $v_i$  содержат случайные погрешности  $v_i$ 

Таким образом, имеем систему равенств

$$y_i - \varepsilon_i = \sum_{j=1}^m u_j x_{ji}, \qquad i = 1, \dots, n.$$
 (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  $\alpha_j$  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

$$\sum_{j=1}^{m} K(x_i, s_j) \alpha_j u(s_j) = f(x_i) - \varepsilon(x_i), \qquad i = 1, \dots, n.$$
 (7.6)

Обозначим

$$f = (f(x_1), \dots, f(x_n))^T, \quad \varepsilon = (\varepsilon(x_1), \dots, \varepsilon(x_n))^T,$$
  
$$u = (u(s_1), \dots, u(s_m))^T,$$

A-(n imes m)-матрица с элементами  $A_{ij}=K(x_i,s_j)lpha_j$ . Тогда система (7.6) запишется в эквивалентном виде (7.2).

Пример 7.4. В приложениях часто приходится рассматривать системы интегральных равенств вида

$$\int_{a}^{b} K_{i}(s)u(s) ds = a_{iT}, \qquad i = 1, \dots, n,$$
 (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 томографин.

 ${
m H}_3$  эксперимента известны не точные значения  $a_{iT}$ , а их приближенные значения  $a_i$ , т. е.  $a_i = a_{iT} + \varepsilon_i$ , где  $\varepsilon_i$  — случайные погрешности.

Проведем дискретизацию системы интегральных венств (7.7). Вместо системы (7.7) получим ее дискретный

$$\sum_{i=1}^{m} K_i(s_j)\alpha_j u(s_j) = a_i - \varepsilon_i, \qquad i = 1, \dots, n,$$
 (7.8)

где известные коэффициенты  $\alpha_i$  зависят от выбранной инте-

гральной формулы.

Если обозначить  $f=(a_1,\ldots,a_n)^T,\ \varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)^T,\ A=0$  $(n \times m)$ -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  $n\geqslant m$ .

#### 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  $u \in R^m$ ;

Векторы ε и f случайные;

3. Отсутствуют систематические смещения в f, т.е.  $\mathbf{M} \varepsilon = \mathbf{0}$  — нулевой вектор;

4. Ковариационная матрица вектора  $\varepsilon$  дается равенством  $K_{\varepsilon}=\sigma^2 I_n$ , где  $I_n$  — единичная матрица порядка n, т.е. компоненты  $\varepsilon_i$  попарно не коррелируют и имеют одинаковую дисперсию  $\sigma^2$  ( $\sigma^2$  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  $\|arepsilon\|=\left(\sum_{i=1}^n arepsilon_i\right)^{1/2}$  — евклидова норма.

Согласно MHK за оценку искомого вектора u принимается вектор, который минимизирует сумму квадратов отклонений:

$$\hat{u}_{\mathsf{MHK}} = \arg\min Q(u). \tag{7.10}$$

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

 $\operatorname{grad}_{u} Q(u) = -2A^{T} f + 2A^{T} A u.$ 

Поэтому получаем равенство  $A^T A \widehat{u}_{\rm MHK} = A^T f$ . Так как A- матрица полного ранга, то симметричная  $(m\times m)$ -матрица  $A^T A-$  положительно определена. Следовательно,

$$\widehat{u}_{MHK} = (A^T A)^{-1} A^T f.$$
 (7.11)

Очевидно, что при  $u=\widehat{u}_{\rm MHK}$  функционал Q(u) имеет единственный (глобальный) минимум, поскольку  $\frac{\partial^2 Q(u)}{\partial u^2}=2A^TA$  положительно определенная матрица.

Как и любая оценка, МНК-оценка  $\widehat{u}_{\rm MHK}$  — (векторная) случайная величина. Из (7.11) следует, что  $\widehat{u}_{\rm MHK}$  — линейная оценка относительно вектора экспериментальных измерений f. Исследуем некоторые свойства МНК-оценки (7.11). Имеем:  $\mathbf{M} \, \widehat{u}_{\rm MHK} = \mathbf{M} \, (A^T A)^{-1} A^T f = (A^T A)^{-1} A^T \mathbf{M} \, f = (A^T A)^{-1} A^T A u = u$ , т. е. МНК-оценка не смещена.

Найдем ковариационную матрицу  $K_u$  МНК-оценки  $\widehat{u}_{\text{МНК}}$ . Имеем:  $K_u = BK_fB^T$ , где  $B = (A^TA)^{-1}A^T$ ,  $K_f$  — ковариацион-

ная 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,

$$K_{\rm u} = \sigma^2 (A^T A)^{-1}. \tag{7.12}$$

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  $\sigma^2$  и тем самым определяется однозначно и при неизвестной  $\sigma^2$ . В то же время ковариационная матрица МНКоценки (7.11), определяющая точность этой оценки, пропорцио-

нальна  $\sigma^2$ .

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  $\hat{u} = Rf$ . Имеем

$$\mathbf{M}\,\widehat{u}=u=R\mathbf{M}\,f=RAu.$$

Поэтому  $RA = I_m$ . Обозначим  $L = R - (A^T A)^{-1} A^T$ . Имеем

$$LA = RA - (A^TA)^{-1}A^TA = I_m - I_m = 0$$

— нулевая матрица. Далее, для ковариационной матрицы  $K_{\widehat{u}}$  оценки  $\widehat{u}$  получаем

$$K_{\widehat{u}} = RK_f R^T = R\sigma^2 I_n R^T = \sigma^2 RR^T = \sigma^2 [L + (A^T A)^{-1} A^T] \times$$

<sup>)</sup> Напомним, что оценка  $\widehat{u}^*$  называется эффективной в классе оценок  $\{\widehat{u}\}$ , если  $K_{\widehat{u}^*}\leqslant K_{\widehat{u}}$  для любой оценки  $\widehat{u}\in\{\widehat{u}\}.$ 

$$\times [L + (A^{T}A)^{-1}A^{T}]^{T} = \sigma^{2}[L + (A^{T}A)^{-1}A^{T}][L^{T} + A(A^{T}A)^{-1}] =$$

$$= \sigma^{2}[LL^{T} + LA(A^{T}A)^{-1} + (LA(A^{T}A)^{-1})^{T} + (A^{T}A)^{-1}] =$$

$$= \sigma^{2}[(A^{T}A)^{-1} + LL^{T}].$$

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 локазана.

Из теоремы Гаусса-Маркова, в частности, следует, что дисперсия любой компоненты МНК-оценки (7.11) не превосходит лисперсию 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  $u_i$ ,  $j=1,\ldots,m$ , 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  $\widehat{v}_{\mathrm{MHK}} = C \widehat{u}_{\mathrm{MHK}}$  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  $\sigma^2$  погрешностей измерений  $f_i$  неизвестна. Оказывается, что в этом случае можно получить оценки не только искомого вектора u, но и дополнительного неизвестного параметра  $\sigma^2$ . Обозначим  $f_{\rm MHK}=A\widehat{u}_{\rm MHK},\ e=f-f_{\rm MHK}$ . Компоненты  $e_1, \dots, e_n$  вектора e называют невязками, а сам вектор e вектором невязок.

Введем положительную статистику

$$s^2 = \frac{\|e\|^2}{n-m}. (7.13)$$

**Теорема 7.2.** Статистика  $s^2$  — несмещенная оценка дисперсии  $\sigma^2$ . Доказательство. Имеем

$$arepsilon = Au + arepsilon - A\widehat{u}_{\mathrm{MHK}} = Au + arepsilon - A(A^TA)^{-1}A^Tf =$$

$$= Au + arepsilon - A(A^TA)^{-1}A^T(Au + arepsilon) =$$

$$= Au + arepsilon - Au - A(A^TA)^{-1}A^T arepsilon = (I_n - A(A^TA)^{-1}A^T) arepsilon = Carepsilon,$$
где  $C = I_n - A(A^TA)^{-1}A^T$ . Поскольку

$$C^{2} = [I_{n} - A(A^{T}A)^{-1}A^{T}]^{2} = I_{n} - 2A(A^{T}A)^{-1}A^{T} + A(A^{T}A)^{-1}A^{T}A(A^{T}A)^{-1}A^{T} = I_{n} - A(A^{T}A)^{-1}A^{T} = C,$$

то C — проекционная матрица. Кроме того, C — симметричная матрица. Далее,

$$\mathbf{M} \|e\|^2 = \mathbf{M} e^T e = \mathbf{M} \varepsilon^T C^T C \varepsilon = \mathbf{M} \varepsilon^T C \varepsilon = \mathbf{M} \sum_{i=1}^n \sum_{j=1}^n C_{ij} \varepsilon_i \varepsilon_j =$$

$$= \sigma^2 \sum_{i=1}^n C_{ii} = \sigma^2 \operatorname{tr} C = \sigma^2 (n - \operatorname{tr}(A(A^T A)^{-1} A^T)) = \sigma^2 (n - m),$$

так как 
$$\operatorname{tr}(A(A^TA)^{-1}A^T)=\operatorname{tr}((A^TA)^{-1}A^TA)=\operatorname{tr}I_m=m.$$
 Итак.  $\mathbf{M}\,s^2=\sigma^2.$  Теорема доказана.

Если  $\sigma^2$  неизвестна, то в качестве несмещенной оценки ковариационной матрицы МНК-оценки (7.11) можно взять матрицу  $K_{\mathfrak{A}} = s^2 (A^T A)^{-1}$ .

Заметим, что корреляционная матрица  $cor(\widehat{u})$  МНКоценки (7.11) не зависит от  $\sigma^2$  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  $y = u_1 + u_2 x$  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

$$A_n u = f_n - \varepsilon_n, \tag{7.14}$$

где  $A_n$  — последовательность матриц с неограниченно увеличивающимся числом строк n.

Для того чтобы последовательность МНК-оценок

$$\widehat{u}_n = (A_n^T A)^{-1} A_n^T f_n \tag{7.15}$$

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  $(A_n^TA_n)^{-1}\to 0$  — 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  $\lambda_{\min}(A_n^TA_n) \to +\infty$  при  $n \to \infty$ . Здесь  $\lambda_{\min}(A_n^TA_n) -$  минимальное собственное значение симметричной положительно определенной матрицы  $A_n^TA_n$ .

Доказательство. Пусть  $\widehat{u}_n$  сходится в среднем квадратичном к u, т.е.  $(A_n^TA_n)^{-1} \to 0$  при  $n \to \infty$ . Из последнего соотношения следует, что все собственные значения последовательности матриц  $(A_n^TA_n)^{-1}$  стремятся к нулю, в частности  $\lambda_{\max}(A_n^TA_n)^{-1} \to 0$ . Но тогда  $\lambda_{\min}(A_n^TA_n) = 1/\lambda_{\max}(A_n^TA_n)^{-1} \to \infty$ . Пусть теперь  $\lambda_{\min}(A_n^TA_n) \to \infty$ . Тогда  $\lambda_{\max}(A_n^TA_n)^{-1} = 1/\lambda_{\min}(A_n^TA_n) \to 0$  при  $n \to \infty$ . Следовательно, все собственные значения последовательности матриц  $(A_n^TA_n)$  стремятся к нулю, что эквивалентно стремлению к нулевой матрице последовательности матриц  $(A_n^TA_n)^{-1}$ . Теорема доказана.

Условие Эйкера,

$$\lambda_{\min}(A_n^T A_n) \to +\infty$$
 при  $n \to \infty$ , (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  $(m \times m)$ -матрича 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=ux, где u — неизвестный параметр, x — независимая переменная. Пусть измерения производятся в точках  $x=i\Delta x$ ,  $i=1,\ldots,n$ , где  $\Delta x$  — расстояние между двумя соседними точками фиксации x. Имеем:  $A_n=\Delta x(1,2,\ldots,n)^T$  — матрица-столбец,  $A_n^T=\Delta x(1,2,\ldots,n)$  — матрица-строка,  $A_n^TA_n=(\Delta x)^2\sum_{i=1}^n i^2=(\Delta x)^2n(n+1)(2n+1)/6$  — положительное число. Имеем:  $A_n^TA_n/n^3\to (\Delta x)^2/3$  при  $n\to\infty$ . Поэтому условие теоремы 7.4 выполнено, и, следовательно, послеловательность МНК-оценок  $u_n=6(n(n+1)(2n+1)\Delta x)^{-1}\sum_{i=1}^n iy_i$  сходится в среднем квадратичном к точному значению параметра u. Конечно, в данном примере состоятельность в среднем квадратичном легко проверяется непосредственно, так как

$$K_n=\sigma^2(\widehat{u}_n)=rac{6\sigma^2}{n(n+1)(2n+1)(\Delta x)^2} o 0$$
 при  $n o \infty.$ 

Приведем еще один достаточный критерий состоятельности в среднем квадратичном последовательности МНК-оценок (7.15). Определим матрицы сопряженности  $R_n$  равенством

$$(R_n)_{ij} = rac{(A_n^T A_n)_{ij}}{\sqrt{(A_n^T A_n)_{ii}(A_n^T A_n)_{jj}}}, \quad i,j = 1, \dots, m.$$

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

$$1^{\circ} \ (A_n^T A_n)_{jj} = \sum_{i=1}^n a_{ij}^2 o \infty \ npu \ n o \infty \ для любого \ j=1, \ldots$$

 $2^{\circ}$   $\lim_{n\to\infty}R_n=R$ , где R — симметричная положительно

определенная (m × m)-матрица.

Тогда последовательность МНК-оценок (7.15) состоятельна в среднем квадратичном.

Доказательство. Введем диагональные  $(m \times m)$ -матрицы

$$D_n = \operatorname{diag}\left(\sqrt{(A_n^T A_n)_{11}}, \dots, \sqrt{(A_n^T A_n)_{mm}}\right).$$

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  $y=u_1+u_2x$ , где  $u_1,\ u_2$  — неизвестные искомые параметры, x — независимая переменная. Пусть измерения производятся в точках  $x_i=i\Delta x,\ i=1,\dots,n$ , где  $\Delta x$  — расстояние между двумя соседними точками фиксации  $x_i$ . Имеем

$$A_{n} = \begin{pmatrix} 1 & \Delta x \\ 1 & 2\Delta x \\ \vdots & n\Delta x \end{pmatrix}$$

$$A_{n}^{T} A_{n} = \begin{pmatrix} n & \frac{n(n+1)}{2} \Delta x \\ \frac{n(n+1)}{2} \Delta x & \frac{n(n+1)(2n+1)}{6} (\Delta x)^{2} \end{pmatrix}.$$

Здесь  $A_n - (n \times 2)$ -матрица,  $A_n^T A_n$  — симметричная положительно определенная  $(2 \times 2)$ -матрица. Матрицы сопряженности даются равенствами

$$R_n = \begin{pmatrix} 1 & \frac{\sqrt{6}}{2} \sqrt{\frac{n+1}{2n+1}} \\ \frac{\sqrt{6}}{2} \sqrt{\frac{n+1}{2n+1}} & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & 1 \end{pmatrix} \text{ при } n \rightarrow \infty.$$

Следовательно, условия теоремы 7.5 выполнены.

Выше была доказана несмещенность оценки (7.13) неизвестной дисперсии  $\sigma^2$ . 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  $C=I_n-A_n(A_n^TA_n)^{-1}A_n^T$  — 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  $\frac{1}{n-m}\sum_{i=1}^n \varepsilon_i^2 o \sigma^2$  по вероятно-

сти при  $n\to\infty$ . Покажем, что  $\frac{1}{n-m} \varepsilon^T A_n (A_n^T A_n)^{-1} A_n^T \varepsilon \to 0$  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

$$\|\mathbf{r}_n\|^2 = \mathbf{r}_n^T \mathbf{r}_n = \frac{1}{n-m} \varepsilon^T A_n P_n P_n^T A_n^T \varepsilon = \frac{1}{n-m} \varepsilon^T A_n (A_n^T A_n)^{-1} A_n^T \varepsilon.$$

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.

Имеем:  $\mathbf{M}\,v_n^0=0$  — нулевой вектор. Для ковариационной матрицы  $V_n^0$  вектора  $v_n^0$  получаем равенства  $V_n^0=V_n^{-1/2}V_nV_n^{-1/2}=V_n^{-1/2}V_n^{1/2}V_n^{1/2}V_n^{-1/2}=I_m$ . 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).

$$\widehat{u}_{n}^{0} = \frac{1}{\sigma} (A_{n}^{T} A_{n})^{1/2} (\widehat{u}_{n} - u) = \frac{1}{\sigma} (A_{n}^{T} A_{n})^{1/2} ((A_{n}^{T} A_{n})^{-1} A_{n}^{T} f - u) = 
= \frac{1}{\sigma} (A_{n}^{T} A_{n})^{1/2} ((A_{n}^{T} A_{n})^{-1} A_{n}^{T} (A_{n} u + \varepsilon) - u) = 
= \frac{1}{\sigma} (A_{n}^{T} A_{n})^{-1/2} A_{n}^{T} \varepsilon = B_{n} \varepsilon,$$

где  $B_n = \frac{1}{\sigma} (A_n^T A_n)^{-1/2} A_n^T - (m \times n)$ -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  $\varepsilon_i$ ,  $i=1,\ldots,n$  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, ..., m, i = 1, ..., 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  $\frac{1}{\sqrt{\alpha_n}}\left(\widehat{u}_n-u\right)\sim \mathcal{N}(0,\sigma^2R^{-1})$  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  $\frac{1}{\sqrt{\alpha_n}} \left(\widehat{u}_n - u\right)$  дается равенством

$$K_{\overline{u}} = \frac{1}{\alpha_n} \sigma^2 (A_n^T A_n)^{-1} = \sigma^2 (\alpha_n A_n^T A_n)^{-1}.$$

Поэтому  $K_{\overline{u}} o \sigma^2 R^{-1}$  при  $n o \infty$ . 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  $\mathcal{N}(u,\sigma^2(A^TA)^{-1})$ . 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  $p_{\mbox{\scriptsize дов}}$  истинное значение  $u_j$  принадлежит интервалу

$$\widehat{u}_{j\mathrm{MHK}} - \sigma t(p_{\mathtt{ДOB}}) \sqrt{(A^T A)_{jj}^{-1}} \leqslant u_j \leqslant \widehat{u}_{j\mathrm{MHK}} + \sigma t(p_{\mathtt{ДOB}}) \sqrt{(A^T A)_{jj}^{-1}} \,,$$

где

$$rac{2}{\sqrt{2\pi}}\int\limits_{0}^{t(p_{_{\mathrm{AOB}}})}e^{-s^{2}/2}\,ds=p_{_{\mathrm{JOB}}},\quad j=1,\ldots,m.$$

До сих пор предполагалось, что закон случайных величин  $\varepsilon_i,\ i=1,\dots,n$ , произволен. Предположим теперь, что случайные величины  $\varepsilon_i$  распределены по нормальному закону. Тогда с учетом предположений классической схемы МНК совместная плотность вероятностей  $p_\varepsilon(x)$  совокупности случайных величин  $\varepsilon$  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  $\widehat{u}_{\text{ММП}}=\arg\max L(u)$ , которое в нашем случае эквивалентно равенству  $\widehat{u}_{\text{ММП}}=\arg\min\|Au-f\|^2$ .

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  $\mathcal{N}(u, \sigma^2 (A^T A)^{-1})$ .

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 МНКоценка (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

$$\operatorname{grad}_{u} \ln L(u) = \frac{1}{\sigma^{2}} A^{T} (f - Au) = \frac{1}{\sigma^{2}} A^{T} \varepsilon.$$

Отсюва

$$(\operatorname{grad}_u \ln L(u))^T = \frac{1}{\sigma^2} \varepsilon^T A,$$

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

$$I = \mathbf{M} \left( \frac{1}{\sigma^4} A^T \varepsilon \varepsilon^T A \right) = \frac{1}{\sigma^4} A^T \mathbf{M} \left( \varepsilon \varepsilon^T \right) A = \frac{1}{\sigma^2} A^T A,$$
$$I^{-1} = \sigma^2 (A^T A)^{-1}.$$

откуда

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  $y=\sum_{i=1}^m a_j u_j=a^T u$ , 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  $x_1, \dots, x_n$ , где производятся экспериментальные измерения функции, но и при других значениях независимой переменной. Прогноз по формуле

$$\widehat{y}_{\mathsf{MHK}} = a^T \widehat{u}_{\mathsf{MHK}} \tag{7.18}$$

будем называть МНК-прогнозом.

**Теорема 7.10.** МНК-прогноз (7.18) несмещен и эффективен в классе всех линейных несмещенных прогнозов.

Доказательство. Имеем

$$\mathbf{M} \, \widehat{y}_{\mathsf{MHK}} = \mathbf{M} \, a^T \widehat{u}_{\mathsf{MHK}} = a^T \mathbf{M} \, \widehat{u}_{\mathsf{MHK}} = a^T u = y,$$

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  $L=R-(A^TA)^{-1}A^T$ . При доказательстве теоремы 7.1 было показано, что LA=0. Имеем

$$\sigma^2(\hat{y}_{\mathsf{MHK}}) = \sigma^2 a^T (A^T A)^{-1} a,$$

$$\begin{split} \sigma^{2}(\widehat{y}) &= a^{T}RK_{f}R^{T}a = \sigma^{2}a^{T}RR^{T}a = \sigma^{2}a^{T}(L + A^{T}A)^{-1}A^{T}) \times \\ &\times (L^{T} + A(A^{T}A)^{-1})a = \sigma^{2}a^{T}((A^{T}A)^{-1} + LL^{T})a = \\ &= \sigma^{2}a^{T}(A^{T}A)^{-1}a + \sigma^{2}a^{T}LL^{T}a \geqslant \sigma^{2}(\widehat{y}_{\mathsf{MHK}}). \end{split}$$

Последнее неравенство следует из неотрицательности симметричной матрицы  $LL^T$ . Теорема доказана.  $\square$ 

Если в предположениях классической схемы вектор погрешностей  $\varepsilon$  распределен нормально, то МНК-прогноз (7.18) распределен по одномерному нормальному закону  $\mathcal{N}(y,\sigma^2a^T(A^TA)^{-1}a)$  и является эффективным в классе всех несмещенных прогнозов (линейных и нелинейных).

### 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  $\varepsilon$  дается равенством  $K_{\varepsilon} = \sigma^2 \Omega$ , где  $\Omega$  — положительно определенная  $(n \times n)$ -матрица  $(\sigma^2$  может быть неизвестной). Таким образом, теперь возможно наличие корреляционных связей между компонентами случайного вектора  $\varepsilon$ .

Поскольку матрица  $\Omega$  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

$$\mathbf{M}\,\boldsymbol{\xi} = \mathbf{M}\,S^{-1}\boldsymbol{\varepsilon} = \boldsymbol{0},$$

$$K_{\xi} = S^{-1}K_{\xi}(S^{-1})^T = \sigma^2 S^{-1}\Omega(S^{-1})^T = \sigma^2 S^{-1}SS^T(S^T)^{-1} = \sigma^2 I_n.$$

Таким образом, регрессионная модель (7.19) удовлетворяет всем предположениям классической схемы МНК. Следовательно, МНК-оценка искомого вектора и определяется равенством

$$\widehat{u}_{\text{MHK}} = \arg\min Q(u), \tag{7.20}$$

rae  $Q(u) = \|Bu - \varphi\|^2$ .

Преобразуем функционал Q(u). Имеем

$$Q(u) = (Bu - \varphi)^{T}(Bu - \varphi) = (S^{-1}Au - S^{-1}f)^{T}(S^{-1}Au - S^{-1}f) = (Au - f)^{T}(S^{-1})^{T}S^{-1}(Au - f) = (Au - f)^{T}(SS^{T})^{-1}(Au - f) = (Au - f)^{T}\Omega^{-1}(Au - f).$$

Итак, МНК-оценка (7.20) определяется равенством

$$\widehat{u}_{\text{MHK}} = \arg\min\left(Au - f\right)^T \Omega^{-1} (Au - f). \tag{7.21}$$

Имеем

$$\widehat{u}_{MHK} = (B^T B)^{-1} B^T \varphi = ((S^{-1} A)^T S^{-1} A)^{-1} (S^{-1} A)^T S^{-1} f =$$

$$= (A^T (S^{-1})^T S^{-1} A)^{-1} A^T (S^{-1})^T S^{-1} f = (A^T \Omega^{-1} A)^{-1} A^T \Omega^{-1} f.$$

Следовательно

$$\widehat{\mathbf{u}}_{MHK} = (A^T \Omega^{-1} A)^{-1} A^T \Omega^{-1} f. \tag{7.22}$$

Оценку (7.22) принято называть взвешенной МНК-оценкой Гаусса-Маркова.

Взвешенная МНК-оценка (7.22) несмещена. Действительно,

$$\mathbf{M} \, \widehat{u}_{\mathsf{MHK}} = (A^T \Omega^{-1} A^T)^{-1} A^T \Omega^{-1} f = (A^T \Omega^{-1} A)^{-1} A^T \Omega^{-1} A u = u.$$

Найдем ковариационную матрицу оценки (7.22). Имеем

$$K_{v} = (A^{T}\Omega^{-1}A)^{-1}A^{T}\Omega^{-1}K_{f}((A^{T}\Omega^{-1}A)^{-1}A^{T}\Omega^{-1})^{T} =$$

$$= \sigma^{2}(A^{T}\Omega^{-1}A)^{-1}A^{T}\Omega^{-1}\Omega\Omega^{-1}A(A^{T}\Omega^{-1}A)^{-1} = \sigma^{2}(A^{T}\Omega^{-1}A)^{-1}.$$

Злесь использована симметричность матриц  $(A^T\Omega^{-1}A)^{-1}$  и  $\Omega^{-1}$ .  $K_u = \sigma^2 (A^T \Omega^{-1} A)^{-1}. \tag{7.23}$ 

Теорема 7.11. При сделанных выше предположениях обобшенная МНК-оценка эффективна в классе линейных несмешенных сценок.

Схема доказательства теоремы 3 аналогична схеме доказательства теоремы 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  $A_n u = f_n - arepsilon_n$ ,  $K_arepsilon = \sigma^2 \Omega_n$ .

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  $\hat{u}_n$  состоятельна в среднем квадратичном.

Доказательство. Достаточно доказать, что  $\lambda_{\min}(A_n^T\Omega_n^{-1}A_n) o$  $\rightarrow \infty$  при  $n \rightarrow \infty$ . Имеем

$$\begin{split} \lambda_{\min}(A_n^T\Omega^{-1}A_n) \geqslant \lambda_{\min}(A_n^TA_n)\lambda_{\min}(\Omega_n^{-1}) &= \frac{\lambda_{\min}(A_n^TA_n)}{\lambda_{\max}(\Omega_n)} \geqslant \\ \geqslant \frac{1}{r}\lambda_{\min}(A_n^TA_n) \to \infty \end{split}$$

при  $n \to \infty$ . Теорема доказана.

Пусть теперь дополнительно к вышеуказанным предположениям вектор погрешностей є 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,

 $\widehat{u}_{\mathsf{MM\Pi}} = \mathrm{arg} \, \max \, L(u) = \mathrm{arg} \, \min \, (f - Au)^T \Omega^{-1} (f - Au) = \widehat{u}_{\mathsf{MHK}}.$ 

Более того, сама взвешенная МНК-оценка будет распределена по нормальному закону  $\mathcal{N}(u,\sigma^2(A^T\Omega^{-1}A)^{-1})$ .

Часто  $\sigma^2$  неизвестна. Из формулы (7.22) следует, что взвешенная МНК-оценка не зависит от  $\sigma^2$ , но ковариационная матрица взвешенной МНК-оценки, определяющая точность этой оценки, пропорциональна  $\sigma^2$  (см. формулу (7.23)). Если  $\sigma^2$  неизвестна, то ее можно оценить по формуле

$$\widehat{\sigma}^2 = \frac{(A\widehat{u}_{\text{MHK}} - f)^T \Omega^{-1} (A\widehat{u}_{\text{MHK}} - f)}{n - m}.$$
 (7.24)

Оценка (7.24) несмещена и состоятельна по вероятности при  $n \to \infty$ .

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:  $\mathbf{M}\,\widehat{u}_{\mathrm{MHK}} = \mathbf{M}\,(A^TA)^{-1}A^Tf = (A^TA)^{-1}A^T\mathbf{M}\,f = u,$  т.е. оценка остается несмещенной без всяких дополнительных

условий.

Предположим, что выполнены условия теоремы 7.12. Имеем

$$\begin{split} K_{\overline{v}} &= (A_n^T A_n)^{-1} A_n^T K_f [(A_n^T A_n)^{-1} A_n^T]^T = \\ &= \sigma^2 (A_n^T A_n)^{-1} A_n^T \Omega_n A_n (A_n^T A_n)^{-1}. \end{split}$$

Поэтому

$$\begin{split} \lambda_{\max}(K_{\widehat{u}}) &\leqslant \sigma^2 \lambda_{\max}(\Omega_n) \lambda_{\max}((A_n^T A_n)^{-1} A_n^T A_n (A_n^T A_n)^{-1}) = \\ &= \sigma^2 \lambda_{\max}((A_n^T A_n)^{-1}) \lambda_{\max}(\Omega_n) \leqslant \frac{\sigma^2 r}{\lambda_{\min}(A_n^T A_n)} \to 0 \end{split}$$

при  $n \to \infty$ . Следовательно, при вышеуказанных условиях МНКоценка (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 и 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

$$Bu = b, (7.25)$$

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  $\hat{u}$ , удовлетворяющую системе уравнений (7.25).

Пусть выполняются предположения классической схемы МНК.

Определим обобщенную МНК-оценку  $\widehat{u}_{\mathrm{OF}}$  соотношением

$$\widehat{u}_{\mathsf{OB}} = \arg\min \|Au - f\|^2 \tag{7.26}$$

при условии (7.25).

Найдем оценку  $\widehat{u}_{\mathrm{OB}}$  и выясним ее свойства.

**Теорема 7.13.** Оценка  $\widehat{u}_{OB}$  (7.26) единственна, несмещена и дается равенством

$$\widehat{u}_{OB} = \widehat{u}_{MHK} + (A^T A)^{-1} B^T C (b - B \widehat{u}_{MHK}),$$
 (7.27)

где  $\widehat{u}_{MHK}$  — МНК-оценка (7.11),  $C = (B(A^TA)^{-1}B^T)^{-1}$ . Для ковариационной матрицы оценки  $\widehat{u}_{OB}$  справедливо равенство

$$K_{\widehat{u}} = \sigma^2 (A^T A)^{-1} (I_m - B^T C B (A^T A)^{-1}).$$
 (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  $\alpha$  — вектор множителей Лагранжа размерности k.

Имеем:  $Q(u) = (Au - f, Au - f) + \alpha^T (Bu - b),$ 

$$\operatorname{grad}_{u} Q(u) = 2A^{T}Au - 2A^{T}f + B^{T}\alpha = 0,$$
 (7.29)

откуда  $u=(A^TA)^{-1}(A^Tf-\frac{1}{2}\,B^T\alpha)$ . Подставляя последнее выражение в (7.25), найдем вектор  $\alpha=2(B(A^TA)^{-1}B^T)^{-1}(B\widehat{u}_{\rm MHK}-b)$ . Если теперь в уравнении (7.29) заменить  $\alpha$  последним выражением, то получаем оценку (7.27).

Далее,

$$\begin{split} \mathbf{M}\,\widehat{u}_{\mathrm{OB}} &= \mathbf{M}\,\widehat{u}_{\mathrm{MHK}} + \mathbf{M}\,(A^TA)^{-1}B^TC(b-B\widehat{u}_{\mathrm{MHK}}) = u + \\ &\quad + (A^TA)^{-1}B^TC(b-Bu) = u, \\ K_{\widehat{u}} &= LK_{\mathbf{u}_{\mathrm{MHK}}}L^T, \end{split}$$

гле  $L = I_m - (A^T A)^{-1} B^T C B$ . Следовательно,

$$K_{ii} = \sigma^{2}(I_{m} - (A^{T}A)^{-1}B^{T}CB)(A^{T}A)^{-1}(I_{m} - B^{T}CB(A^{T}A)^{-1}) =$$

$$= \sigma^{2}((A^{T}A)^{-1} - (A^{T}A)^{-1}B^{T}CB(A^{T}A)^{-1})(I_{m} -$$

$$- B^{T}CB(A^{T}A)^{-1}) = \sigma^{2}(A^{T}A)^{-1}(I_{m} - B^{T}CB(A^{T}A)^{-1})^{2} =$$

$$= \sigma^{2}(A^{T}A)^{-1}(I_{m} - B^{T}CB(A^{T}A)^{-1}).$$

Теорема доказана.

Из доказательства теоремы 7.13 следует, что

$$P = I_m - B^T C B (A^T A)^{-1}$$

- 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

$$K_{\widehat{u}} = \sigma^2 (A^T A)^{-1} - \sigma^2 (A^T A)^{-1} B^T C B (A^T A)^{-1} \leqslant$$
  
  $\leqslant \sigma^2 (A^T A)^{-1} = K_{u_{MHK}},$ 

поскольку матрица  $(A^TA)^{-1}B^TCB(A^TA)^{-1}$  симметрична и неотрицательна.

Итак, обобщенная МНК-оценка (7.27) эффективней МНКоценки (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  $K_{\widehat{\mathbf{u}}}=0$  — нулевая матрица. Действительно, имеем

$$C = (B(A^T A)^{-1}B^T)^{-1} = (B^T)^{-1}(A^T A)B^{-1},$$

$$\begin{split} \widehat{u}_{\text{OB}} &= \widehat{u}_{\text{MHK}} + (A^T A)^{-1} B^T (B^T)^{-1} (A^T A) B^{-1} (b - B \widehat{u}_{\text{MHK}}) = \\ &= \widehat{u}_{\text{MHK}} + B^{-1} b - \widehat{u}_{\text{MHK}} = B^{-1} b, \end{split}$$

$$K_{\hat{u}} = \sigma^2 (A^T A)^{-1} (I_m - B^T (B^T)^{-1} (A^T A) B^{-1} B (A^T A)^{-1}) =$$
  
=  $\sigma^2 (A^T A)^{-1} (I_m - I_m) = 0.$ 

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 мат-

— для МНК-оценки (7.11) —  $A^T A$ ;

— для взвешенной МНК-оценки (7.22) —  $\Omega, A^T \Omega^{-1} A^T;$ 

— для обобщенной МНК-оценки (7.27) —  $A^TA$ ,  $B(A^TA)^{-1}B^T$ .

Матрицы  $A^TA$ ,  $A^T\Omega^{-1}A^T$  — размерности  $(m \times m)$ ,  $\Omega$  —

размерности  $(n \times n)$ ,  $B(A^TA)^{-1}B^T$  — размерности  $(k \times k)$ . Поскольку все матрицы  $A^TA$ ,  $\Omega$ ,  $A^T\Omega^{-1}A$ ,  $B(A^TA)^{-1}B^T$ 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 и. В этих случаях необходимо использовать дополнительную априорную информацию о векторе и 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 представле-

ны в работах [2, 6, 17, 31, 77, 78, 81].

### 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);  $\varepsilon$  — помеха, носящая хаотический характер, причем  $\mathbf{M} \, \varepsilon = 0$ ;  $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_m$  — параметры, подлежащие оценке; Y — искомая величина (отклик системы).

Модель (7.30) может быть использована для прогнозирования искомой величины Y для будущих моментов времени, если зна-

чения 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  $u_0, u_1, \dots, u_m$  на этапах «обучения» и «переобучения» модели (7.30) применим МНК.

Запишем модель (7.30) в виде двух равенств:

$$\mathbf{M}Y = u_0 + \sum_{j=1}^{m} u_j \, \mathbf{M} \, X_j,$$
 (7.31)

$$Y^0 = \sum_{j=1}^m u_j X_j^0 + \varepsilon, \tag{7.32}$$

гле  $Y^0 = Y - \mathbf{M} Y$ ,  $X_j^0 = X_j - \mathbf{M} X_j$  — центрированные величины.

Формулу для определения  $u_1, \ldots, u_m$  получаем из (7.32), минимизируя дисперсию  $\sigma^2(\varepsilon) = \mathbf{M}(Y^0 - \sum_{j=1}^m u_j X_j^0)^2$  по  $u_1, \ldots, u_m$ :  $u = K_x^{-1} \operatorname{cov}(Y, X), \tag{7.33}$ 

где  $u=(u_1,\ldots,u_m)^T,\;K_x$  — положительно определенная ковариационная  $(m\times m)$ -матрица факторов  $X_1,\ldots,X_m;$  соv (Y,X)=  $=(\operatorname{cov}(Y,X_1),\ldots,\operatorname{cov}(Y,X_m))^T$  — вектор, компоненты которого  $\operatorname{cov}(Y,X_j)=\mathbf{M}(Y^0X_j^0)$  — ковариации величины Y и факторов  $X_i,\;j=1,\ldots,m.$ 

На практике на этапах «обучения» или «переобучения» модели (7.30) следует использовать формулу (7.33) с заменой  $K_x$  и  ${\rm cov}\,(Y,X)$  их оценками, полученными на основе «историче-

ских» данных.

Если база «исторических» данных, используемых для «обучения», состоит из реализованных значений

$$Y_i, X_{ij}, j = 1, \dots, m, i = 1, \dots, n,$$
 (7.34)

то оценки элементов матрицы  $K_x$  и вектора  $\mathrm{cov}\,(Y,X)$  определяются равенствами

$$(\widehat{K}_x)_{rl} = \frac{1}{n-1} \sum_{i=1}^n (X_{ir} - \overline{X}_r) (X_{il} - \overline{X}_l), \quad \overline{X}_r = \frac{1}{n} \sum_{i=1}^n X_{ir},$$

$$\widehat{\text{cov}}(Y, X_r) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y}) (X_{ir} - \overline{X}_r), \quad \overline{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Таким образом, окончательные оценки параметров  $u_0, u_1, \ldots, u_m$ , полученные на этапе «обучения» модели (7.30), определяются равенствами

$$\widehat{u}_0 = \overline{Y} - \sum_{j=1}^m \widehat{u}_j \, \overline{X}_j, \quad \widehat{u} = (\widehat{u}_1, \dots, \widehat{u}_m)^T = \widehat{K}_x^{-1} \widehat{\text{cov}}(Y, X), \quad (7.35)$$

где  $\widehat{\operatorname{cov}}(Y,X) = (\widehat{\operatorname{cov}}(Y,X_1),\ldots,\widehat{\operatorname{cov}}(Y,X_m))^T$ .

После этапа «обучения» модели (7.30), определяемого формулами (7.35), модель (7.30) можно использовать для прогнозирования искомой величины Y согласно равенству

$$Y_{ ext{npor}} = \widehat{u}_0 + \sum_{j=1}^m \widehat{u}_j \, X_j,$$

гле  $X_1, \dots, X_m$  — 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  $\sigma_i^2$ , причем ни реальная величина таких погрешностей, ни их номера i не известны.

При наличии больших выбросов необходимо применять робастные методы оценивания [97], позволяющие значительно снизить вредное влияние больших выбросов на оценку и получить приемлемую итоговую оценку искомых параметров.

### 1. Робастные М-оценки параметров линейных моделей

Существует несколько робастных методов оценивания [82, 95, 97, 104]. 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 экстремаль-

ную задачу

$$\widehat{u}_{po6} = \arg\min_{u} \sum_{i=1}^{n} |(Au)_i - f_i|^p.$$
 (8.1)

где  $(Au)_i - i$ -я компонента вектора Au,  $1 \le p < 2$ .

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  $K_f = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ .

Тогда взвешенная МНК-оценка будет определяться равен-

$$\widehat{u}_{\text{MHK}} = \arg\min_{u} \sum_{i=1}^{n} ((Au)_i - f_i)^2 \frac{1}{\sigma_i^2}.$$
 (8.2)

При наличии больших выбросов некоторые измерения  $f_i$  будут иметь обльшие значения дисперсий, чем соответствующие им дисперсии  $\sigma_i^2$ , причем ни номера i больших выбросов, ни истинные значения дисперсий  $\sigma^2(f_i)$  не известны.

Введем М-класс робастных оценок, определяемых равенством

$$\widehat{u}_{po6} = \arg\min_{u} \sum_{i=1}^{n} \rho\left(\frac{(Au)_i - f_i}{\sigma_i}\right), \tag{8.3}$$

где ho(x) — выпуклая четная положительная функция, ho(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 [82, 95, 97, 104]. 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

$$\rho(x) = \begin{cases} x^2/2, & |x| \le K, \\ -K^2/2 + K|x|, & |x| > K, \end{cases}$$

где K — параметр Хьюбера [97].

Часто робастная оценка Хьюбера при наличии больших выбросов предпочтительней других робастных оценок, так как она является минимаксной оценкой в классе плотностей, образуемом моделью смеси [97].

Множество индексов  $I = \{i: i = 1, \dots, n\}$  разобьем на три подмножества:

$$I_{0} = \{i : |f_{i} - (\widehat{Au}_{po6})_{i}| \leq K\sigma_{i}\},\$$

$$I_{-} = \{i : f_{i} - (\widehat{Au}_{po6})_{i} < -K\sigma_{i}\},\$$

$$I_{+} = \{i : f_{i} - (\widehat{Au}_{po6})_{i} > K\sigma_{i}\}.$$
(8.4)

Очевидно, что  $I = I_0 \cup I_- \cup I_+$ .

Уравнение  $\operatorname{grad}_u\sum_{i=1}^n \rho((A\widehat{u}_{\mathsf{poo}})_i-f_i)/\sigma_i)=0$  для определения робастной оценки Хьюбера имеет вид

$$\sum_{i \in I_0} \frac{a_{ij}}{\sigma_i^2} \left( (A \widehat{u}_{\mathsf{po6}})_i - f_i \right) + \sum_{i \in I_-} \frac{K a_{ij}}{\sigma_i} - \sum_{i \in I_+} \frac{K a_{ij}}{\sigma_i} = 0, \quad j = 1, \dots, m.$$

Последние уравнения перепишем в эквивалентном виде

$$\begin{split} \sum_{i \in I_0} \frac{a_{ij}}{\sigma_i^2} \left( (A \widehat{u}_{\text{po6}})_i - f_i \right) + \sum_{i \in I_-} \frac{K \sigma_i a_{ij} + a_{ij} (A \widehat{u}_{\text{po6}})_i - a_{ij} (A \widehat{u}_{\text{po6}})_i}{\sigma_i^2} - \\ - \sum_{i \in I_+} \frac{K \sigma_i a_{ij} + a_{ij} (A \widehat{u}_{\text{po6}})_i - a_{ij} (A \widehat{u}_{\text{po6}})_i}{\sigma_i^2} = 0. \end{split}$$

Отсюда получаем

$$\begin{split} & \sum_{i=1}^{n} \frac{a_{ij}}{\sigma_{i}^{2}} \left( A \widehat{u}_{\text{po6}} \right)_{i} = \\ & = \sum_{i \in I_{0}} \frac{a_{ij}}{\sigma_{i}^{2}} f_{i} + \sum_{i \in I_{-}} \frac{a_{ij}}{\sigma_{i}^{2}} \left[ (A \widehat{u}_{\text{po6}})_{i} - K \sigma_{i} \right] + \sum_{i \in I_{+}} \frac{a_{ij}}{\sigma_{i}^{2}} \left[ (A \widehat{u}_{\text{po6}})_{i} + K \sigma_{i} \right], \end{split}$$

 $j=1,\dots,m$ . Из последней системы следует, что  $\widehat{u}_{\mathsf{po6}}$  определяется равенством

$$\widehat{u}_{\mathsf{po6}} = (A^T W A)^{-1} A^T W \widehat{f},\tag{8.5}$$

где  $W = \mathrm{diag}\,(1/\sigma_1^2,\dots,1/\sigma_n^2),\; \widetilde{f} = (\widetilde{f}_1,\dots,\widetilde{f}_n)^T;\; \widetilde{f}_i = f_i,\; \mathrm{если}\; i \in I_0;\; \widetilde{f}_i = (A\widehat{u}_{\mathrm{po6}})_i - K\sigma_i,\; \mathrm{если}\; i \in I_-;\; \widetilde{f}_i = (A\widehat{u}_{\mathrm{po6}})_i + K\sigma_i,\; \mathrm{если}\; i \in I_+.$ 

## 2. Численные методы нахождения робастных оценок параметров линейных моделей

Нахождение робастной оценки (8.3) связано с минимизацией функционала  $\sum_{i=1}^{n} \rho((Au)_i - f_i)/\sigma_i)$ , 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

$$\sum_{i=1}^{n} \rho' \left( \frac{(Au)_i - f_i}{\sigma_i} \right) \frac{a_{ij}}{\sigma_i} = 0, \qquad j = 1, \dots, m.$$
 (8.6)

Систему (8.6) с учетом четности  $\rho(x)$  запишем в эквивалентном виде

$$\sum_{i=1}^{n} W_i((Au)_i - f_i)a_{ij} = 0, \qquad j = 1, \dots, m,$$
 (8.7)

$$W_{i} = \frac{|\rho'((Au)_{i} - f_{i})/\sigma_{i})|}{|(Au)_{i} - f_{i}|\sigma_{i}}.$$
(8.8)

Если бы «веса»  $W_i$  были известны, то решение системы (8.7) давалось бы равенством

$$u = (A^T W A)^{-1} A^T W f, (8.9)$$

где  $W = \operatorname{diag}(W_1, \ldots, W_n)$ .

На основе (8.8), (8.9) предлагается итерационный метод на-хождения робастной оценки (8.3):

$$W^{(l)} = \operatorname{diag}(W_1^{(l)}, \dots, W_n^{(l)}), \quad W_i^{(l)} = \frac{|\rho'(((Au^{(l)})_i - f_i)/\sigma_i)|}{|(Au^{(l)})_i - f_i|\sigma_i}$$
$$u^{(l+1)} = (A^T W^{(l)} A)^{-1} A^T W^{(l)} f.$$
(8.10)

В качестве первого приближения рекомендуется брать средневзвешенную МНК-оценку  $u^{(1)} = (A^T W^{(0)} A)^{-1} A^T W^{(0)} f$ , где  $W^{(0)} = \operatorname{diag}(1/\sigma_1^2, \dots, 1/\sigma_n^2)$ .

Итерационный метод (8.10) нахождения робастной оцен-

ки (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

$$\widehat{u}_{po6} = \arg\min_{u} \sum_{i=1}^{n} \widetilde{W}_{i} ((Au)_{i} - f_{i})^{2},$$
 (8.11)

ГЛЕ

$$\widetilde{W}_{i} = \frac{\rho((Au)_{i} - f_{i})/\sigma_{i}}{((Au)_{i} - f_{i})^{2}}.$$
(8.12)

Если бы «веса»  $\widetilde{W}_i$  были известны, то решение экстремальной задачи (8.11) давалось бы равенством

$$u = (A^T \widetilde{W} A)^{-1} A^T \widetilde{W} f,$$
 (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):

$$\widetilde{W}^{(l)} = \operatorname{diag}(\widetilde{W}_{1}^{(l)}, \dots, \widetilde{W}_{n}^{(l)}), \quad \widetilde{W}_{i}^{(l)} = \frac{\rho(((Au^{(l)})_{i} - f_{i})/\sigma_{i})}{((Au^{(l)})_{i} - f_{i})^{2}},$$

$$u^{(l+1)} = (A^{T}\widetilde{W}^{(l)}A)^{-1}A^{T}\widetilde{W}^{(l)}f. \tag{8.14}$$

В качестве первого приближения рекомендуется брать средневзвешенную МНК-оценку  $u^{(1)}=(A^TW^{(0)}A)^{-1}A^TW^{(0)}f$ , где  $W^{(0)}=\mathrm{diag}\,(1/\sigma_1^2,\ldots,1/\sigma_n^2).$ 

Итерационный метод (8.14) нахождения робастной оценки (8.3) называется методом вариационно-взвешенных квадра-

тических приближений.

Если  $\rho(x) = |x|^p$ ,  $1 \le p < 2$ , т. е. рассматривается робастная  $L_p$ -оценка, то итеративный МНК (8.10) и метод вариационновзвешенных квадратических приближений (8.14) эквивалентны.

Сходимость итераций (8.10), (8.14) для случая  $\rho(x) = |x|^p$ ,  $1 \le p < 2$ , 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

(ПП):

$$u^{(l+1)} = (A^T W A)^{-1} A^T W f^{(l)}, (8.15)$$

где

$$\begin{split} f^{(l)} &= (f_1^{(l)}, \dots, f_n^{(l)})^T; \quad f_i^{(l)} = f_i, \quad \text{если} \quad i \in I_0^{(l)}; \\ f_i^{(l)} &= (Au^{(l)})_i + K\sigma_i, \quad \text{если} \quad i \in I_+^{(l)}; \\ f_i^{(l)} &= (Au^{(l)})_i - K\sigma_i, \quad \text{если} \quad i \in I_-^{(l)}; \\ I_0^{(l)} &= \{i: \mid (Au^{(l)})_i - f_i \mid \leqslant K\sigma_i\}, \\ I_-^{(l)} &= \{i: \mid f_i - (Au^{(l)})_i < -K\sigma_i\}, \\ I_+^{(l)} &= \{i: \mid f_i - (Au^{(l)})_i > K\sigma_i\}. \end{split}$$

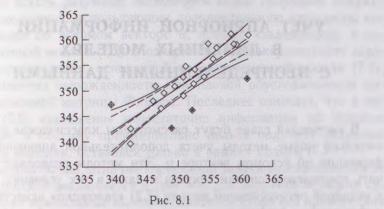
В качестве первого приближения рекомендуется брать

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  $(A^TW^{(l)}A)^{-1}A^TW^{(l)}$ , в том числе обращая матрицу  $A^TW^{(l)}A$ . В ИП (8.15) формирование матрицы  $(A^TWA)^{-1}A^TW$  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  $(A^TW^{(l)}A)^{-1}A^TW^{(l)}$ .

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 § 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  $I_-^{(l)}$ .



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 (+,  $\star$ ) помечены значения с  $i \in I_+$ 

Робастные методы для линейных моделей представлены также в работах [8–10, 82, 98, 104].

 $u i \in I_{-}$ 

#### 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 и.

Выше были рассмотрены только несмещенные (или, по крайней мере, асимптотически несмещенные) оценки искомого вектора и 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  $\delta = b(\widehat{u}, u_T) + \eta(\widehat{u})$ , гле  $b(\widehat{u}, u_T) = M\widehat{u} - u_T$  — вектор смещения,  $\eta(\widehat{u}) = \widehat{u} - M\widehat{u}$  — 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 плохой 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 *и* 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.

### 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

$$p(u \mid f) = \frac{p(u)p(f \mid u)}{\int p(u)p(f \mid u) du},$$
 (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 апостериорное 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  $\varepsilon$ ) 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.

$$p(f \mid u) = \frac{1}{(2\pi)^{n/2} \sigma^n |\Omega|^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} (f - Au)^T \Omega^{-1} (f - Au) \right\},$$

$$p(u) = \frac{1}{(2\pi)^{m/2} \sigma_a^m |K_a|^{1/2}} \exp \left\{ -\frac{1}{2\sigma_a^2} (u - u_0)^T K_a^{-1} (u - u_0) \right\},$$

где  $u_0$ .  $\sigma_a^2 K_a$  — центр и ковариационная матрица априорного распределения.

 $\dot{\mathbf{A}}$ ля рассматриваемого случая, когда законы распределения вектора f и априорный нормальные, все три вышеуказанных

 $_{
m Bap}$ ианта Б-оценок искомого вектора u совпадают и определяются формулой

$$\widehat{u}_{\mathsf{B}} = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} K_a^{-1}\right)^{-1} \left(\frac{1}{\sigma^2} A^T \Omega^{-1} f + \frac{1}{\sigma_a^2} K_a^{-1} u_0\right). \tag{9.2}$$

В частности, если  $u_0 = 0$ , то

$$\widehat{u}_{\rm B} = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} K_a^{-1}\right)^{-1} A^T \Omega^{-1} f. \tag{9.3}$$

Если  $u_0 \neq u_T$ , то Б-оценка смещена. Действительно,

$$\begin{split} \mathbf{M} \, \widehat{u}_{\mathrm{B}} &= \\ &= \left( \frac{1}{\sigma^2} \, A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} \, K_a^{-1} \right)^{-1} \left( \frac{1}{\sigma^2} \, A^T \Omega^{-1} A u_T + \frac{1}{\sigma_a^2} \, K_a^{-1} u_0 \right) = \\ &= u_T + \left( \frac{1}{\sigma^2} \, A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} \, K_a^{-1} \right)^{-1} \, \frac{1}{\sigma_a^2} \, K_a^{-1} (u_0 - u_T) \neq u_T, \end{split}$$

если  $u_0 \neq u_T$ .

Таким образом, смещение Б-оценки дается равенством

$$b(\widehat{u}_{\mathsf{B}}, u_T) = \frac{1}{\sigma_a^2} \left( \frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} K_a^{-1} \right)^{-1} K_a^{-1} (u_0 - u_T).$$

Чем лучше «угадан» вектор  $u_0$ , тем меньше величина смещения. Подсчитаем ковариационную матрицу Б-оценки (9.2). Имеем

$$\begin{split} K_{\widehat{u}_{\mathrm{B}}} &= (\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma_{a}^{2}}K_{a}^{-1})^{-1}\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}\Omega\Omega^{-1}A\times \\ &\times (\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma_{a}^{2}}K_{a}^{-1})^{-1} = \frac{1}{\sigma^{2}}(\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma^{2}}K_{a}^{-1})^{-1}\times \\ &\times A^{T}\Omega^{-1}A(\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma_{a}^{2}}K_{a}^{-1})^{-1} = \\ &= (\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma_{a}^{2}}K_{a}^{-1})^{-1}(\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma_{a}^{2}}K_{a}^{-1} - \frac{1}{\sigma^{2}}K_{a}^{-1})\times \\ &\times (\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma_{a}^{2}}K_{a}^{-1})^{-1} = (\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma_{a}^{2}}K_{a}^{-1})^{-1} - \\ &- \frac{1}{\sigma^{2}}(\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{1}{\sigma^{2}_{a}}K_{a}^{-1})^{-1}K_{a}^{-1}(\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}A^{T} + \frac{1}{\sigma^{2}_{a}}K_{a}^{-1})^{-1}. \end{split}$$

Но симметричная матрица

$$\frac{1}{\sigma_a^2} \Big( \frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} K_a^{-1} \Big)^{-1} K_a^{-1} \Big( \frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} K_a^{-1} \Big)^{-1}$$

положительно определена. Поэтому

$$K_{\widehat{u}_{\mathbf{g}}} < \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\sigma_a^2} K_a^{-1}\right)^{-1} < \sigma^2 (A^T \Omega^{-1} A)^{-1} = K_{\widehat{u}_{\mathbf{MHK}}}.$$

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 и, 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  $\frac{1}{\sigma_a}A^T\Omega^{-1}A+\frac{1}{\sigma_a^2}K_a^{-1}$  положительно определена вне зависимости от обусловленности матрицы Фишера  $\frac{1}{\sigma_a^2}A^T\Omega^{-1}A$  за счет слагаемого  $\frac{1}{\sigma_a^2}K_a^{-1}$ .

Теоретические основы метода Байеса представлены в работах [30, 38].

В настоящее время метод Байеса широко применяется при решении некорректных задач, в том числе некорректных задач обработки экспериментальных данных и задач восстановления [8, 25, 35, 36, 57, 67, 87].

### 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

$$\widehat{u}_{\mathsf{MM}} = \arg\min_{\widehat{u} \in \{\widehat{u}\}} J(\widehat{u}). \tag{9.5}$$

Если априорная информация отсутствует, т. е.  $R_a = R^m$ , и в (9.4) взята евклидова норма, то MM-оценка (9.5) совпадает с MHK-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»  $\rho$ :

 $R_a = \{ u : (u - u_0)^T B(u - u_0) \le \rho^2 \}, \tag{9.6}$ 

где B — симметричная положительно определенная матрица. Тогда MM-оценка дается равенством

$$\widehat{u}_{\text{MM}} = u_0 + \frac{1}{\sigma^2} \left( \frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\rho^2} B' \right)^{-1} A^T \Omega^{-1} (f - A u_0), \quad (9.7)$$

где модифицированная матрица B' такова, что  $\widehat{u}_{\mathsf{MM}}$  принадлежит границе множества  $R_a$ .

В частности, если  $u_0 = 0$ , то

$$\widehat{u}_{MM} = \left(A^T \Omega^{-1} A + \frac{\sigma^2}{\rho^2} B'\right)^{-1} A^T \Omega^{-1} f.$$
 (9.8)

Пусть теперь  $R_a - m$ -мерный параллелепипед с центром в  $u_0$ :

$$R_a = \{u : (u - u_0, \varphi_k)^2 \le \lambda_k^2, k = 1, ..., m\},$$
 (9.9)

где  $\{\varphi_k\}$  — ортонормированный базис пространства  $R^m$ ,  $\{\lambda_k^2\}$  — заданные положительные числа,

Обозначим

$$B = U \operatorname{diag}(1/\lambda_1^2, \dots, 1/\lambda_m^2)U^T,$$

где  $U=(\varphi_1,\ldots,\varphi_m)$  — ортогональная матрица, k-й столбец которой образован m компонентами вектора  $\varphi_k$ . 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

$$\widehat{u}_{MM} = u_0 + \frac{1}{\sigma^2} \left( \frac{1}{\sigma^2} A^T \Omega^{-1} A + B' \right)^{-1} A^T \Omega^{-1} (f - Au_0), \quad (9.10)$$

где B' — модифицированная матрица B [92].

В частности, если  $u_0 = 0$ , то

$$\widehat{u}_{MM} = (A^T \Omega^{-1} A + \sigma^2 B')^{-1} A^T \Omega^{-1} f.$$
 (9.11)

В приложениях априорное множество  $R_a$  часто задается в виде совокупности неравенств:

$$a_i \leqslant u_i \leqslant b_i, \qquad i = 1, \dots, m, \tag{9.12}$$

где  $a_i$ ,  $b_i$  — заданные числа.

Совокупность неравенств (9.12) определяет параллелепи-

$$u_0 = \frac{1}{2} (a_1 + b_1, \dots, a_m + b_m)^T,$$

$$B = 4 \operatorname{diag} \left( \frac{1}{(b_1 - a_1)^2}, \dots, \frac{1}{(b_m - a_m)^2} \right),$$

$$U = I_m, \quad \varphi_k = l_k = (0, \dots, 1, \dots, 0)^T, \quad \lambda_k^2 = \frac{1}{4} (b_k - a_k)^2.$$

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 «уже» 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 ho o 0 (для оценки (9.7)) или  $\max \lambda_k^2 o 0$  (для оценки (9.10)). Наоборот, если  $ho o \infty$  (для оценки (9.7)) или  $\min \lambda_k^2 o \infty$  (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 (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 MM-оценки смещены, точность MM-оценки можно определить с помощью симметричной неотрицательной  $(m \times m)$ -матрицы  $\mathbf{D}(\widehat{u}_{\mathsf{MM}}) = \mathbf{M}(\widehat{u}_{\mathsf{MM}} - u_T)(\widehat{u}_{\mathsf{MM}} - u_T)^T$ , в частности, i-й элемент главной диагонали матрицы  $\mathbf{D}(\widehat{u}_{\mathsf{MM}})$  равен  $\mathbf{M}(\widehat{u}_{i\mathsf{MM}} - u_i)^2$ , где  $\widehat{u}_{i\mathsf{MM}}$ ,  $u_i - i$ -е компоненты векторов  $\widehat{u}_{\mathsf{MM}}$  и  $u_T$  соответственно. Следовательно,  $\operatorname{Sp} \mathbf{D}(\widehat{u}_{\mathsf{MM}}) = \sum_{i=1}^{m} \mathbf{M}(\widehat{u}_{i\mathsf{MM}} - u_i)^2$  — среднеквадратическая ошибка MM-оценки.

Для ММ-оценок (9.7), (9.10)

$$\mathbf{D}(\widehat{u}_{\mathsf{MM}}) = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{1}{\rho^2} B'\right)^{-1},$$

$$\mathbf{D}(\widehat{u}_{\mathsf{MM}}) = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + B'\right)^{-1}$$

соответственно.

Поскольку матрицы B' в обеих MM-оценках (9.7), (9.10) положительно определены, то справедливо неравенство  $D\left(\widehat{u}_{MM}\right) < K_{\widehat{u}_{MHK}}$  и для оценки (9.7), и для оценки (9.10), где  $K_{\widehat{u}_{MHK}} = \sigma^2(A^T\Omega^{-1}A)^{-1}$  — ковариационная матрица МНК-оценки (7.22).

Итак, ММ-оценки (9.7), (9.10) эффективней МНКоценки (7.22). Это и неудивительно, так как ММ-оценки используют дополнительную информацию об искомом век-

Tope  $u_T$ .

Для фактического нахождения ММ-оценок (9.7), (9.10) необходимо обратить симметричные положительно определенные  $(m \times m)$ -матрицы  $\left(\frac{1}{\sigma^2}A^T\Omega^{-1}A+\frac{1}{\sigma^2}B'\right)$ ,  $\left(\frac{1}{\sigma^2}A^T\Omega^{-1}A+B'\right)$  соответственно. Для обращения этих матриц рекомендуется пользоваться методом квадратного корня.

ММ-оценки (9.8), (9.11) эквивалентны смещенным риджоценкам, 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 В' 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 [31].

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) [84].

Минимаксный метод решения некорректно поставленных задач представлен в [92].

### 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 и некорректной задачи восстановления без учета дополнительной априорной информации о векторе и, 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  $\varepsilon$ . 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  $u_{\text{ммП}}$  максимизирующий функцию правдоподобия:

$$\widehat{u}_{MM\Pi} = \arg\max_{u} L(u). \tag{9.13}$$

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,

$$M_{ij} = -\mathbf{M} \frac{\partial^2 \ln L(u)}{\partial u_i \partial u_j}, \quad i, j = 1, \dots, m.$$

Матрица Фишера представима также в виде

$$M = \mathbf{M} (\operatorname{grad}_u \ln L(u)) (\operatorname{grad}_u \ln L(u))^T$$
.

Поскольку  $\mathbf{M}$  grad<sub>u</sub> ln L(u)=0, то M — ковариационная матрица случайного вектора grad<sub>u</sub> 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 [38], то ММП-оценка единственна и является единственным решением системы уравнений правдоподобия

$$\operatorname{grad}_{u} \ln L(u) = 0. \tag{9.14}$$

Напомним, что  $MM\Pi$ -оценки обладают оптимальными свойствами среди всех асимптотически несмещенных оценок. Оптимальность  $MM\Pi$ -оценок означает их асимптотическую эффективность, т.е. приближение при увеличении объема выборки n к нижней границе точности несмещенных оценок, определяемой неравенством Фишера—Крамера—Рао

$$K_{\overline{u}} \geqslant M^{-1}, \tag{9.15}$$

где  $K_{\overline{u}}$  — 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  $\varepsilon$ , когда

$$p_{arepsilon}(x) = rac{1}{(2\pi)^{n/2} \sigma^n |\Omega|^{1/2}} \, \exp \left[ -rac{x^T \Omega^{-1} x}{2\sigma^2} 
ight],$$

МНК-оценка (7.22) совпадает с ММП-оценкой,  $M=\frac{1}{\sigma^2}A^T\Omega^{-1}A$ , 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  $\xi$ , независимая от случайной величины  $\varepsilon$  из (7.2), плотность вероятностей которой параметрически зависит от искомого вектора u. Пусть далее известна выборка

 $Y_1, \dots, Y_N$  (9.16)

значений случайной величины  $\xi$ . 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

$$p(y_1,\ldots,y_N;u)=\prod_{i=1}^N p_{\xi}(y_i;u),$$

гле p(y;u) — плотность вероятностей случайной величины  $\xi$ . 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  $\varepsilon, Y_1, \dots, Y_N$  равна  $p_\varepsilon(x)p(y_1, \dots, y_N; u)$ . Функцию

$$L_0(u) = p_{\varepsilon}(f - Au)p(Y_1, \dots, Y_N; u)$$

назовем обобщенной функцией правдоподобия.

$$L_0(u) = L(u) L_a(u),$$
 (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 правдо-

подобия 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

$$M_0 = M + M_a, (9.19)$$

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  $u_T$  принимается вектор  $\widehat{u}_{\text{ОММП}}$ , максимизирующий обобщенную функцию правдоподобия:

$$\widehat{u}_{\mathsf{OMM\Pi}} = \arg\max L_0(u). \tag{9.20}$$

Очевидно, что решения экстремальных задач (9.20) и  $\widehat{u} = \arg\max \ln L_0(u)$  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  $\varepsilon$  и  $\xi$ . Тогда согласно ОММП вместо (9.20) необходимо решать экстремальную задачу

$$(\widehat{u}_{\mathsf{OMM\Pi}}, \widehat{\alpha}_{\mathsf{OMM\Pi}}) = \arg \max_{(u,\alpha)} L_0(u,\alpha),$$
 (9.22)

где  $\alpha=(\alpha_1,\dots,\alpha_l)^T$  — 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  $M_0, M, M_a$ , определяемые равенствами

$$(\widetilde{M}_0)_{ij} = -\mathbf{M} \frac{\partial^2 \ln L_0(u, \alpha)}{\partial v_i \partial v_j}, \quad (\widetilde{M})_{ij} = -\mathbf{M} \frac{\partial^2 \ln L(u, \alpha)}{\partial v_i \partial v_j},$$

$$(\widetilde{M}_a)_{ij} = -\mathbf{M} \frac{\partial^2 \ln L_a(u, \alpha)}{\partial v_i \partial v_j}, \quad i, j = 1, \dots, m + l,$$

$$v = (u_1, \dots, u_m, \alpha_1, \dots, \alpha_l)^T \equiv (v_1, \dots, v_{m+l})^T.$$

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

$$L(u) = \frac{1}{(2\pi)^{n/2} \sigma^n |\Omega|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2} (f - Au)^T \Omega^{-1} (f - Au)\right\},$$

$$L_{\alpha}(u) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{m/2} \sigma_i^m |K_i|^{1/2}} \exp\left\{-\frac{1}{2\sigma_i^2} (Y_i - u)^T K_i^{-1} (Y_i - u)\right\},$$

$$M = \frac{1}{\sigma^2} A^T \Omega^{-1} A, \quad M_a = \sum_{i=1}^{N} \frac{1}{\sigma^2} K_i^{-1}.$$

Имеем

$$M_0 = \frac{1}{\sigma^2} A^T \Omega^{-1} A + \sum_{i=1}^N \frac{1}{\sigma_i^2} K_i^{-1}.$$
 (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 плообусловлена 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  $\sigma_i^2$ , либо при увеличении объема априорной выборки N.

Для рассматриваемого случая решение системы обобщенных уравнений правдоподобия (9.21) существует, единственно и дается формулой

$$\widehat{u}_{\text{OMM}\Pi} = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1}\right)^{-1} \times \left(\frac{1}{\sigma^2} A^T \Omega^{-1} f + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1} Y_i\right). \tag{9.24}$$

В частности, если  $Y_i \sim \mathcal{N}(u, \sigma_a^2 K_a)$ , то

$$\widehat{\boldsymbol{u}}_{\text{OMM}\Pi} = \left(\frac{1}{\sigma^2} A^T \Omega^{-1} A + \frac{N}{\sigma_a^2} K_a^{-1}\right)^{-1} \left(\frac{1}{\sigma^2} A^T \Omega^{-1} f + \frac{1}{\sigma_a^2} K_a^{-1} \sum_{i=1}^N Y_i\right). \tag{9.25}$$

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  $\sigma_a^2$ ,  $K_a$ ,  $u_0$  различен. Таким образом, для частного случая, когда имеется лишь одна реализация априорной случайной величины, а законы распределения случайных величин  $\varepsilon$ ,  $\xi$  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..., f. Пусть, например,  $f_j \sim \mathcal{N}(f_T, \sigma_{fj}^2 K_{fj})$ . Тогда

$$L(u) = \frac{1}{(2\pi)^{nr/2}} \prod_{j=1}^{r} \frac{1}{\sigma_{fj}^{n} |K_{fj}|^{1/2}} \exp\left\{-\frac{1}{2\sigma_{fj}^{2}} (f_{j} - Au)^{T} K_{fj}^{-1} (f - Au)\right\}$$

и ОММП-оценка дается равенством

$$\widehat{u}_{\text{OMMII}} = \left(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^{2}} A^{T} K_{fj}^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_{i}^{2}} K_{i}^{-1}\right)^{-1} \times \left(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^{2}} A^{T} K_{fj}^{-1} f_{j} + \sum_{i=1}^{N} \frac{1}{\sigma_{i}^{2}} K_{i}^{-1} Y_{i}\right). \quad (9.27)$$

ОММП-оценку (9.27) можно записать в эквивалентном виде

$$\widehat{u}_{\text{OMMII}} = \left( A^T K_{\widehat{f}}^{-1} A + K_a^{-1} \right)^{-1} \left( A^T K_{\widehat{f}}^{-1} \widehat{f} + K_a^{-1} \widehat{u}_a \right), \quad (9.28)$$

гле  $K_{\widehat{f}} = \left(\sum_{j=1}^{\tau} \frac{1}{\sigma_{fj}^2} K_{fj}^{-1}\right)^{-1}$  — ковариационная матрица ММПоценки  $\widehat{f}$  вектора f по реализациям  $f_1, \dots, f_r$ ,

$$\widehat{f} = \left(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^{2}} K_{fj}^{-1}\right)^{-1} \sum_{j=1}^{r} \frac{1}{\sigma_{fj}^{2}} K_{fj}^{-1} f_{j}, \quad K_{a} = \left(\sum_{i=1}^{N} \frac{1}{\sigma_{i}^{2}} K_{i}^{-1}\right)^{-1}$$

— ковариационная матрица ММП-оценки  $\widehat{u}_a$  искомого вектора  $u_T$  по реализации априорной выборки  $Y_1, \dots, Y_N$ ,

$$\widehat{u}_a = \left(\sum_{i=1}^N \frac{1}{\sigma_i^2} K_i^{-1}\right)^{-1} \sum_{i=1}^N \frac{1}{\sigma_i^2} K_i^{-1} Y_i.$$

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, находится ММП-оценка  $\hat{f}$  и ее ковариационная матрица  $K_{\hat{i}}$ , а по реализациям априорной выборки (9.16) находится ММП-оценка  $\hat{u}_a$  и ее ковариационная матрица  $K_a$ ; на втором этапе по формуле (9.28) экончательно находится ОММП-оценка искомого вектора u. В частности, если  $f_j \sim \mathcal{N}(f_T, \sigma^2\Omega), \ j=1,\dots,r,$  и  $Y_i \sim \mathcal{N}(u_T, \sigma_a^2 K_a)$ , то оценка (9.27) принимает вид

$$\widehat{u}_{\text{ОММП}} = \left(\frac{r}{\sigma^2} A^T \Omega^{-1} A + \frac{N}{\sigma_a^2} K_a^{-1}\right)^{-1} \times \left(\frac{1}{\sigma^2} A^T \Omega^{-1} \sum_{j=1}^r f_j + \frac{1}{\sigma_a^2} K_a^{-1} \sum_{i=1}^N Y_i\right). \quad (9.29)$$

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  $\xi$ . Если, например, усреднение производится только по реализациям f при фиксированном числе реализаций  $\xi$ , то ОММП-оценки, вообще говоря, смещены.

Ковариационные матрицы ОММП-оценок (9.24)-(9.29) дают-

ся, соответственно, формулами 
$$K_{\widehat{u}} = \left(\frac{1}{\sigma^2}A^T\Omega^{-1}A + \sum_{i=1}^N \frac{1}{\sigma_i^2}K_i^{-1}\right)^{-1},$$
 
$$K_{\widehat{u}} = \left(\frac{1}{\sigma^2}A^T\Omega^{-1}A + \frac{N}{\sigma_a^2}K_a^{-1}\right)^{-1},$$
 
$$K_{\widehat{u}} = \left(\frac{1}{\sigma^2}A^T\Omega^{-1}A + \frac{1}{\sigma_a^2}K_a^{-1}\right)^{-1},$$
 
$$K_{\widehat{u}} = \left(\sum_{j=1}^r \frac{1}{\sigma_{fj}^2}A^TK_{fj}^{-1}A + \sum_{i=1}^N \frac{1}{\sigma_i^2}K_i^{-1}\right)^{-1},$$
 
$$K_{\widehat{u}} = (A^TK_{\widehat{f}}^{-1}A + K_a^{-1})^{-1}, \quad K_{\widehat{u}} = \left(\frac{r}{\sigma^2}A^T\Omega^{-1}A + \frac{N}{\sigma_a^2}K_a^{-1}\right)^{-1}.$$

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

$$\widehat{u}_{\mathsf{OMM\Pi}} = u_T + R_1 \Delta f + R_2 \Delta u,$$

где

$$R_{1} = \left(\frac{r}{\sigma^{2}} A^{T} \Omega^{-1} A + \frac{N}{\sigma_{a}^{2}} K_{a}^{-1}\right)^{-1} \frac{r}{\sigma^{2}} A^{T} \Omega^{-1},$$

$$R_{2} = \left(\frac{r}{\sigma^{2}} A^{T} \Omega^{-1} A + \frac{N}{\sigma_{a}^{2}} K_{a}^{-1}\right)^{-1} \frac{N}{\sigma_{a}^{2}} K_{a}^{-1},$$

$$\Delta f = \hat{f} - f_{T}, \quad \Delta u = \hat{u} - u_{T}, \quad \hat{f} = \frac{1}{r} \sum_{j=1}^{r} f_{j}, \quad \hat{u} = \frac{1}{N} \sum_{i=1}^{N} Y_{i}.$$

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  $\Delta f$ ,  $\Delta u$  следует

$$K_{\widehat{u}_{\text{CIM-MLIT}}} = R_1 K_{\Delta f} R_1^T + R_2 K_{\Delta u} R_2^T = R_1 K_{\widehat{f}} R_1^T + R_2 K_{\widehat{u}} R_2^T.$$

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

$$U^{T} \frac{1}{\sigma^{2}} A^{T} \Omega^{-1} A U = \operatorname{diag}(\lambda_{1}, \dots, \lambda_{l}, 0, \dots, 0), \quad U^{T} \frac{1}{\sigma_{\alpha}^{2}} K_{\alpha}^{-1} U = I_{m},$$

где  $\lambda_1,\dots,\lambda_l>0$  — сингулярные числа матрицы  $\frac{1}{\sigma^2}\,A^T\Omega^{-1}A,\,l$  — ранг матрицы  $A,\,l\leqslant m.$ 

Имеем

$$\begin{split} U^{-1}R_{1}K_{\overline{f}}R_{1}^{T}(U^{T})^{-1} &= rU^{-1}(\frac{r}{\sigma^{2}}A^{T}\Omega^{-1}A + \frac{N}{\sigma_{a}^{2}}K_{a}^{-1})^{-1} \times \\ &\times (U^{T})^{-1}U^{T}\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}AUU^{-1}(\frac{r}{\sigma^{2}}A^{T}\Omega^{-1}A + \\ &\quad + \frac{N}{\sigma_{a}^{2}}K_{a}^{-1})^{-1}(U^{T})^{-1} = \\ &= r(rU^{T}\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}AU + NU^{T}\frac{1}{\sigma_{a}^{2}}K_{a}^{-1}U)^{-1} \times \\ &\times U^{T}\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}AU(rU^{T}\frac{1}{\sigma^{2}}A^{T}\Omega^{-1}AU + NU^{T}\frac{1}{\sigma_{a}^{2}}K_{a}^{-1}U)^{-1} = \end{split}$$

$$= r(\operatorname{diag}(N + r\lambda_1, \dots, N + r\lambda_l, N, \dots, N))^{-1} \times \operatorname{diag}(\lambda_1, \dots, \lambda_l, 0, \dots, 0) \times \times (\operatorname{diag}(N + r\lambda_1, \dots, N + r\lambda_l, N, \dots, N))^{-1} = \operatorname{diag}(\frac{r\lambda_1}{(N + r\lambda_1)^2}, \dots, \frac{r\lambda_l}{(N + r\lambda_l)^2}, 0, \dots, 0).$$

Следовательно,  $R_1K_{\widehat{f}}R_1^T$  стремится к нулевой матрице при  $N \to \infty$  для любого r.

Аналогично получаем равенство

$$U^{-1}R_2K_{\widehat{u}}R_2^T(U^T)^{-1} = \operatorname{diag}\left(\frac{N}{(N+r\lambda_1)^2}, \dots, \frac{N}{(N+r\lambda_l)^2}, \frac{1}{N}, \dots, \frac{1}{N}\right).$$

Следовательно,  $R_2K_{\widehat{u}}R_2^T$  стремится к нулевой матрице при  $N\to\infty$  для любого r. Итак,  $K_{\widehat{u}_{\rm OMM\Pi}}$  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  $K_{\widehat{u}_{\text{ОММП}}} = R_1 K_{\widehat{f}} R_1^T + R_2 K_{\widehat{u}} R_2^T$  из предыдущей теоремы. Из доказательства теоремы 9.1 следует, что  $R_1 K_{\widehat{f}} R_1^T$  стремится к нулевой матрице не только при  $N \to \infty$ , но и при  $r \to \infty$ , а  $R_2 K_{\widehat{u}} R_2^T$  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

$$U^{-1}R_2\Delta u = U^{-1} \left(\frac{r}{\sigma^2} A^T \Omega^{-1} A + \frac{N}{\sigma_a^2} K_a^{-1}\right)^{-1} \frac{N}{\sigma_a^2} K_a^{-1} U \Delta v =$$

$$= \operatorname{diag} \left(\frac{N}{N + r\lambda_1}, \dots, \frac{N}{N + r\lambda_l}, 1, \dots, 1\right) \Delta v.$$

Следовательно,  $R_2\Delta u$  сходится в среднем квадратичном при  $r \to \infty$  к случайному вектору

$$u^* = U \operatorname{diag}(0, ..., 0, 1, ..., 1)U^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}Y_i - u_T\right).$$

В частности, если A — матрица полного ранга (l=m), то  $u^*=0$ . Имеем

$$\begin{split} U^T \frac{1}{\sigma^2} A^T \Omega^{-1} A u^* &= U^T \frac{1}{\sigma^2} A^T \Omega^{-1} A U \operatorname{diag} \left(0, \dots, 0, 1, \dots, 1\right) U^{-1} \times \\ &\times \left(\frac{1}{N} \sum_{i=1}^N Y_i - u_T\right) = \operatorname{diag} \left(\lambda_1, \dots, \lambda_l, 0, \dots, 0\right) \times \\ &\times \operatorname{diag} \left(0, \dots, 0, 1, \dots, 1\right) U^{-1} \left(\frac{1}{N} \sum_{i=1}^N Y_i - u_T\right) = 0. \end{split}$$

Поэтому  $A^T\Omega^{-1}Au^*=0$ . Но из последнего равенства следует  $Au^*=0$ . Действительно, представим  $\Omega$  в виде  $\Omega^{1/2}\Omega^{1/2}$ , где  $\Omega^{1/2}$  симметрична и положительно определена. Имеем

$$0 = (A^{T} \Omega^{-1} A u^{*}, u^{*}) = (A^{T} \Omega^{-1/2} \Omega^{-1/2} A u^{*}, u^{*}) =$$

$$= (\Omega^{-1/2} A u^{*}, \Omega^{-1/2} A u^{*}) = ||\Omega^{-1/2} A u^{*}||^{2}.$$

Следовательно.  $Au^* = 0$ . Теорема доказана.

Теоремы 9.1, 9.2 гарантируют состоятельность ОММПошенки 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  $\sigma^2 \to 0$ .

**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 вышепривеленных 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:

 $1^{\circ}\ u\in R$ , где R либо все пространство  $R^{m}$ , либо множество, для которого определена p(z;u);

 $2^{\circ} \frac{\partial p(z,u)}{\partial u_k}, \ k=1,\ldots,m,$  существуют для любого  $u\in R;$ 

 $3^\circ \int \left| \frac{\partial p(z,u)}{\partial u_k} \right| dz, \ k=1,\ldots,m$ , существуют и конечны для любого  $u\in R$ ;

4° все элементы  $(M_0)_{ij},\ i,j=1,\dots,m$ , 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  $f_1, \dots, f_r$  правой части регрессионной модели (7.2) и априорной выборки (9.16).

**Теорема 9.5.** Пусть выполнены обобщенные условия Фишера-Крамера-Рао и интегралы  $\left|\left|\widehat{u}_i(z)\frac{\partial p(z;u)}{\partial u_j}\right|dz\right|$  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

$$K_{\widehat{u}} \geqslant M_0^{-1}.$$
 (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  $f_r$  теряет смысл (не существует). Чтобы в этом случае была возможность получать приемлемые оценки искомого вектора u, необходимо к реализациям  $f_1$  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 и), 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).

$$M = \sum_{j=1}^{r} \frac{1}{\sigma_{jj}^{2}} A^{T} K_{jj}^{-1} A, \quad M_{a} = \sum_{i=1}^{N} \frac{1}{\sigma_{i}^{2}} K_{i}^{-1}, \tag{9.3i}$$

7.5.

$$M_0 = \sum_{j=1}^r \frac{1}{\sigma_{fj}^2} \, A^T K_{fj}^{-1} A + \sum_{i=1}^N \frac{1}{\sigma_i^2} \, K_i^{-1}.$$

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  $K_{\widehat{a}} = M_0^{-1}$ .

Напомним, что ОММП-оценки (9.24)—(9.29) были получены, когда законы распределения  $f_1,\ldots,f_\tau$  и членов априорной выборни (9.16) нормальные. Следовательно, в условиях, когда законы распределения реализаций  $f_1,\ldots,f_\tau$  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  $\tau$  и N.

Если законы распределений  $f_1, \dots, f_r$  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 асимптотиче-

ски 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 неизвест-

ны уровни шума  $\sigma^2$  и  $\sigma^2$ .

Пусть для определенности выполнены условия, при которых получена ОММП-оценка (9.29), но  $\sigma^2$  и  $\sigma_a^2$  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

$$\begin{split} \widehat{\boldsymbol{u}}_{\text{OMM}\Pi} &= \left(\frac{r}{\widehat{\sigma}^2} A^T \Omega^{-1} A + \frac{N}{\widehat{\sigma}_a^2} K_a^{-1}\right)^{-1} \left(\frac{r}{\widehat{\sigma}^2} A^T \Omega^{-1} \widehat{\boldsymbol{f}} + \frac{N}{\widehat{\sigma}_a^2} K_a^{-1} \widehat{\boldsymbol{u}}_a\right), \\ \widehat{\sigma}^2 &= \frac{1}{rn} \sum_{j=1}^r \|A \widehat{\boldsymbol{u}}_{\text{OMM}\Pi} - f_i\|_{\Omega^{-1}}^2, \\ \widehat{\sigma}_a^2 &= \frac{1}{Nm} \sum_{i=1}^N \|\widehat{\boldsymbol{u}}_{\text{OMM}\Pi} - Y_i\|_{K_a^{-1}}^2, \\ \widehat{\boldsymbol{f}} &= \frac{1}{r} \sum_{j=1}^r f_j, \quad \widehat{\boldsymbol{u}}_a &= \frac{1}{N} \sum_{i=1}^N Y_i. \end{split}$$

 $\mathsf{OMM}\mathsf{\Pi}$ -оценку для u можно записать в эквивалентном виде

$$\widehat{\boldsymbol{u}}_{\mathsf{OMM}\Pi} = (\boldsymbol{A}^T \boldsymbol{\Omega}^{-1} \boldsymbol{A} + \widehat{\boldsymbol{\alpha}} \boldsymbol{K}_a^{-1})^{-1} (\boldsymbol{A}^T \boldsymbol{\Omega}^{-1} \widehat{\boldsymbol{f}} + \widehat{\boldsymbol{\alpha}} \boldsymbol{K}_a^{-1} \widehat{\boldsymbol{u}}_a),$$

PITO

$$\widehat{\alpha} = \frac{N^2 m \sum_{j=1}^{T} \|A\widehat{u}_{\text{OMM}\Pi} - f_j\|_{\Omega^{-1}}^2}{r^2 n \sum_{i=1}^{N} \|\widehat{u}_{\text{OMM}\Pi} - Y_i\|_{K_n^{-1}}^2}.$$
(9.32)

Параметр  $\hat{\alpha}$  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 зада-

выборку (9.16), чтобы обобщенная информационная матрица  $M_0$  стала положительно определенной.

Теорема 9.5 остается в силе, даже если обе информационные матрицы M,  $M_{\rm B}$  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),

$$M = \sum_{j=1}^{r} \frac{1}{\sigma_{fj}^{2}} A^{T} K_{fj}^{-1} A, \quad M_{a} = \sum_{i=1}^{N} \frac{1}{\sigma_{i}^{2}} K_{i}^{-1}, \quad (9.31)$$

T.e

$$M_0 = \sum_{j=1}^{\tau} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1}.$$

Но для ОММП-оценки (9.27)

$$K_{\text{fi}} = \left(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} K_i^{-1}\right)^{-1},$$

и поэтому  $K_{\hat{u}} = M_0^{-1}$ .

Напомним. что ОММП-оценки (9.24)–(9.29) были получены, когда законы распределения  $f_1, \ldots, f_r$  и членов априорной выборни (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  $f_1, \dots, f_r$  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 неизвест-

ны уровни шума  $\sigma^2$  и  $\sigma_a^2$ .

Пусть для определенности выполнены условия, при которых получена ОММП-оценка (9.29), но  $\sigma^2$  и  $\sigma_a^2$  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

$$\begin{split} \widehat{u}_{\mathrm{OMM\Pi}} &= \Big(\frac{r}{\widehat{\sigma}^2} A^T \Omega^{-1} A + \frac{N}{\widehat{\sigma}_a^2} K_a^{-1}\Big)^{-1} \Big(\frac{r}{\widehat{\sigma}^2} A^T \Omega^{-1} \widehat{f} + \frac{N}{\widehat{\sigma}_a^2} K_a^{-1} \widehat{u}_a\Big), \\ \widehat{\sigma}^2 &= \frac{1}{rn} \sum_{j=1}^r \|A \widehat{u}_{\mathrm{OMM\Pi}} - f_i\|_{\Omega^{-1}}^2, \\ \widehat{\sigma}_a^2 &= \frac{1}{Nm} \sum_{i=1}^N \|\widehat{u}_{\mathrm{OMM\Pi}} - Y_i\|_{K_a^{-1}}^2, \\ \widehat{f} &= \frac{1}{r} \sum_{i=1}^r f_i, \quad \widehat{u}_a = \frac{1}{N} \sum_{i=1}^N Y_i. \end{split}$$

 $\mathsf{OMM}\Pi$ -оценку для u можно записать в эквивалентном виде

$$\widehat{u}_{\mathsf{OMMII}} = (A^T \Omega^{-1} A + \widehat{\alpha} K_a^{-1})^{-1} (A^T \Omega^{-1} \widehat{f} + \widehat{\alpha} K_a^{-1} \widehat{u}_a),$$

где

$$\widehat{\alpha} = \frac{N^2 m \sum_{j=1}^{r} \|\widehat{Au}_{\text{OMM}\Pi} - f_j\|_{\Omega^{-1}}^2}{r^2 n \sum_{i=1}^{N} \|\widehat{u}_{\text{OMM}\Pi} - Y_i\|_{K_n^{-1}}^2}.$$
(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 зада-

вать 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 и как случайной величины. Конечно, любая оценка искомого вектора и, 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 баиесовским 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 и, 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 задается выборкой эначений 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  $\Omega$  и  $K_a$  часто неизвестны. Тогда вместо оптимальной оценки (9.25) естественно рассмотреть оценку

$$\widehat{u} = \left(\frac{r}{\sigma^2} A^T A + \frac{N}{\sigma_a^2} B\right)^{-1} \left(\frac{r}{\sigma^2} A^T \widehat{f} + \frac{N}{\sigma_a^2} B \widehat{u}_a\right), \tag{9.33}$$

где B — симметричная положительно определенная матрица, вообще говоря отличная от  $K_{\sigma}^{-1}$ .

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:  $r \to \infty; N \to \infty; \sigma^2 \to 0; \sigma^2 \to 0.$ 

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.

### 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  $\mathbb{R}^m$ .

Согласно ОММП за оценку искомого вектора u берется вектор  $\widehat{u}_{\mathsf{ОММП}}$ , максимизирующий функцию правдоподобия L(u) при

ограничении  $u \in R_a$ . т. е.

$$\widehat{u}_{\mathsf{OMM\Pi}} = \arg \max_{u \in R_a} L(u). \tag{9.34}$$

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  $f_1, \ldots, f_r$ части регрессионной модели (7.2) нормальный  $f_j \sim N(f_T, \sigma_{f_j}^2 K_{f_j})$ , а сами реализации независимы. Тогда ОММП-оценка (9.34) является решением экстремальной задачи

$$\widehat{\mathbf{u}}_{\text{OMMII}} = \arg\min_{\mathbf{u} \in R_u} \sum_{j=1}^{r} \frac{1}{\sigma_{fj}^2} (f_j - Au)^T K_{fj}^{-1} (f_j - Au). \tag{9.35}$$

Пусть  $R_a$  — эллипсоид (9.6). Тогда ОММП-оценка (9.35) существует, единственна и дается равенством

HOMMI =

$$= \left(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^{2}} A^{T} K_{fj}^{-1} A + \alpha^{*} B\right)^{-1} \left(\sum_{j=1}^{r} \frac{1}{\sigma_{fj}^{2}} A^{T} K_{fj}^{-1} f_{j} + \alpha^{*} B u_{0}\right),$$
(9.36)

где множитель Лагранжа  $\alpha^* > 0$  находится из равенства

$$\|\widehat{u}_{\text{OMM}\Pi} - u_0\|_B^2 = \rho^2. \tag{9.37}$$

Если все реализации  $f_1,\dots,f_r$  независимы и одинаково распределены по закону  $\mathcal{N}(f_T,\sigma^2\Omega)$ , то ОММП-оценка (9.36) принимает вид

$$\widehat{u}_{\text{OMM}\Pi} = \left(\frac{r}{\sigma^2} A^T \Omega^{-1} A + \alpha^* B\right)^{-1} \left(\frac{r}{\sigma^2} A^T \Omega^{-1} \widehat{f} + \alpha^* B u_0\right), \quad (9.38)$$

где 
$$\widehat{f} = \frac{1}{r} \sum_{j=1}^{r} f_j$$
.

В частности, если имеется только одна реализация f, то

$$\widehat{u}_{\text{OMMII}} = (A^T \Omega^{-1} A + \alpha^* B)^{-1} (A^T \Omega^{-1} f + \alpha^* B u_0), \qquad (9.39)$$

где  $\alpha^* > 0$  находится из равенства (9.37).

ОММП-оценки (9.36)–(9.39), вообще говоря, смещены, т. е. М  $\widehat{u}_{\rm OMM\Pi}=u_T-\Delta u$ , где смещение  $\Delta u$  для оценок (9.36)–(9.39) имеет, соответственно, величину

$$\Delta u = \alpha^* \Big( \sum_{j=1}^r \frac{1}{\sigma_{fj}^2} A^T K_{fj}^{-1} A + \alpha^* B \Big)^{-1} B(u_0 - u_T),$$

$$\Delta u = \alpha^* \Big( \frac{r}{\sigma^2} A^T \Omega^{-1} A + \alpha^* B \Big)^{-1} B(u_0 - u_T),$$

$$\Delta u = \alpha^* \Big( \frac{1}{\sigma^2} A^T \Omega^{-1} A + \alpha^* B \Big)^{-1} B(u_0 - u_T).$$

Рассмотрим, например, семейство оценок (9.39) с  $\alpha^*=\alpha>0$ . При  $\alpha=0$ 

 $\widehat{u} = \widehat{u}_{\mathsf{MM\Pi}} = (A^T \Omega^{-1} A)^{-1} A^T \Omega^{-1} f.$ 

При 
$$lpha o \infty$$

$$\widehat{u} o u_0$$
.

Если  $\widehat{u}_{\mathsf{MM\Pi}} \notin R_a$ , то ОММП-оценка (9.39) вычисляется по формуле (9.39) при условии  $\|\widehat{u}_{\mathsf{ОММ\Pi}} - u_0\|_B^2 = \rho^2$ . Если  $\widehat{u}_{\mathsf{ММ\Pi}} \in R_a$  то ОММП-оценка (9.39) совпадает с  $\widehat{u}_{\mathsf{MM\Pi}}$  (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  $lpha^*$ , при котором

$$(A^T\Omega^{-1}A + \alpha^*B)^{-1}(A^T\Omega^{-1}f + \alpha^*Bu_0) \in \overline{R}_a \setminus R_a,$$

и полагаем

$$\widehat{u}_{OMM\Pi} = (A^T \Omega^{-1} A + \alpha^* B)^{-1} (A^T \Omega^{-1} f + \alpha^* B u_0).$$

Если при lpha=0  $\widehat{u}_{\mathsf{MM\Pi}}$  существует и принадлежит  $R_a$ , то полагаем  $\widehat{u}_{\mathsf{OMM\Pi}}=\widehat{u}_{\mathsf{MM\Pi}}.$ 

Заметим, что для случая  $\widehat{u}_{\mathsf{MM\Pi}} \in R_a$  имеем  $\alpha^* = 0$  и поэтому  $\Delta u = 0$ , т.е. смещение отсутствует. Итак, если  $\widehat{u}_{\mathsf{MM\Pi}} \in R_a$ , то дополнительная априорная информация  $u \in R_a$  «не срабатывает», если же  $\widehat{u}_{\mathsf{MM\Pi}}$  не существует или не удовлетворяет по точности (матрица  $A^T\Omega^{-1}A$  вырождена или плохо обусловлена) или  $\widehat{u}_{\mathsf{MM\Pi}} \notin R_a$  то  $\widehat{u}_{\mathsf{OMM\Pi}}$  «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 квад-

ратичном 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  $\rho_0$ , что  $u_{\text{ОММП}}=u_{\text{ММП}}$ , 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 [40, 56, 59, 63–66, 71, 72, 84–87, 129].

Обобщенный метод максимального правдоподобия представлен в публикациях [10, 72, 107].

## 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 задачтакая 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  $x_1$  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  $\Pi_m$  размерности m.

Систему (7.1) запишем в эквивалентном виде

$$(b_i, x) = 0, i = 1, ..., n,$$
 (9.40)

гле  $b_i^T - i$ -я строка матрицы B.

При фиксированном  $x\in\Pi_m$  множество векторов  $S_m=\{v:v\in R^{m+1},\ (v,x)=0\}$  образует подпространство размерности m. Для произвольного  $v\in R^{m+1}$  имеет место ортогональное разложение  $v=v_S+v_N$ , где  $v_S\in S_m,\ v_N\bot S_m$ .

Имеем

$$v_N = rac{(v,x)x}{\|x\|^2}, \quad x 
eq 0, \quad 
ho^2(v,S_m) = \|v_N\|^2 = rac{(v,x)^2}{\|x\|^2},$$

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.

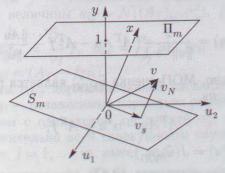


Рис. 9.1

Рассмотрим функционал

$$J(u) = \frac{\|Au - f\|^2}{\|u\|^2 + 1}. (9.41)$$

Напомним, что согласно МНК в рамках классической схемы требуется найти вектор и, минимизирующий функционал невяз-

 $KH \|Au - f\|^2.$ 

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) [47], 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  $S_m$  ( $x\bot S_m$ ), 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 и, минимизирующий функционал ортогональных проекций (9.41). Подсчет градиента функционала (9.41) дает формулу

$$\operatorname{grad}_{u} J(u) = \frac{2}{\|u\|^{2} + 1} \Big( A^{T} A u - A^{T} f - \frac{\|Au - f\|^{2}}{\|u\|^{2} + 1} u \Big).$$

Следовательно, МОП-оценка  $\widehat{u}_{\text{МОП}}$  является решением нелинейного уравнения

$$A^T A u - \alpha^* u = A^T f, \tag{9.42}$$

Fig.  $\alpha^* = \min J(u) = J(u_{\text{MOH}}).$ 

Нелинейность уравнения (9.42) обусловлена зависимостью параметра  $\alpha^* > 0$  от искомой МОП-оценки.

Можно оценить снизу значение  $lpha^*$ . 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

$$B^T B v = \lambda v, \tag{9.43}$$

где  $\lambda$  — множитель Лагранжа.

Уравнение (9.43) — задача на собственные значения симметричной неотрицательной (m+1) imes (m+1)-матрицы  $B^T B$ . Пусть  $\lambda^* \geqslant 0$  — наименьшее собственное значение, а  $v^*$  ( $||v^*|| = 1$ ) соответствующий ему собственный вектор матрицы  $B^TB$ . По-

$$\sum_{i=1}^n (b_i,v)^2 = (B^TBv,v),$$

$$\min_{\|v\|=1} \sum_{i=1}^{n} (b_i, v)^2 = \min_{\|v\|=1} (B^T B v, v) = \lambda^*.$$

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,  $lpha^*=\lambda^*=0$ , 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  $f_{\mathrm{H}}=rac{\sigma_{a}}{\sigma_{f}}f,\ u_{\mathrm{H}}=rac{\sigma_{a}}{\sigma_{f}}u,\ f_{T\mathrm{H}}=rac{\sigma_{a}}{\sigma_{f}}f_{T},\ B_{\mathrm{H}}=$ 

 $B_{CH} = (A, -f_{H}), \ B_{TH} = (A_{T}, -f_{TH}).$ Все элементы  $b_{Hij}$  нормализованной расширенной  $n \times (m+1)$ +1)матрицы  $B_{
m H}$  — независимые нормально распределенные случайные величины с одинаковой дисперсией  $\sigma_a^2$ . Функция правдоподобия относительно всей совокупности случайных величин  $b_{{
m H}ij},\,i=1,\ldots,n,\,j=1,\ldots,m$ , имеет вид

$$L(B_{TH}) = (2\pi\sigma_a^2)^{-n(m+1)/2} \exp\Big\{-\frac{1}{2\sigma_a^2} \sum_{i=1}^n \sum_{j=1}^{m+1} (b_{Hij} - b_{THij})^2\Big\},\,$$

 $^{\mathsf{ГДе}}\ b_{T\mathrm{H}ij}$  — неизвестные элементы нормализованной расширенной матрицы  $B_{TH}$ .

Согласно ММП для нахождения ММП-оценок элементов  $b_{THij}$  необходимо максимизировать  $L(B_{TH})$ , что эквивалентно минимизации функционала

$$J(B_{TH}) = \sum_{i=1}^{n} \sum_{j=1}^{m+1} (b_{Hij} - b_{THij})^2 \equiv \sum_{i=1}^{n} ||b_{Hi} - b_{THi}||^2,$$

где  $b_{\text{Hi}}^T$ ,  $b_{T\text{Hi}}^T$  — i-е строки матриц  $B_{\text{H}}$  и  $B_{T\text{H}}$  соответственно.

Согласно вышесказанному, векторы  $b_{TH}$ ; принадлежат подпространству  $S_{Hm}$ , ортогональному искомому нормализованному вектору  $x_H = (u_H, 1)$ . Следовательно, функционал  $J(B_{TH})$  является функционалом ортогональных проекций для нормализованной системы  $Au_H = f_H$ . Итак, для нормализованной системы ММП эквивалентен МОП при выполнении вышеуказанных условий относительно случайных погрешностей. Уравнение (9.42) для нормализованной системы относительно ММПоценки  $u_H$  искомого нормализованного вектора  $u_H$  имеет вид

$$A^TA\widehat{u}_{\mathrm{H}}-lpha^*\widehat{u}_{\mathrm{H}}=A^Tf_{\mathrm{H}},$$
 где  $lpha^*=\min_{\widehat{u}_{\mathrm{H}}\in E^m}rac{\|Au_{\mathrm{H}}-f_{\mathrm{H}}\|^2}{\|\widehat{u}_{\mathrm{H}}\|^2+1}$ 

Возвращаясь к исходным векторам, получаем

$$A^T A \widehat{u}_{\mathsf{MO\Pi}} - \alpha^* \widehat{u}_{\mathsf{MO\Pi}} = A^T f$$
, где  $\alpha^* = \sigma_a^2 \min_{u \in R^m} \frac{\|Au - f\|^2}{\sigma_a^2 \|u\|^2 + \sigma_f^2}$ . (9.44)

Из последнего равенства, в частности, следует, что  $\alpha^* \to 0$  при  $\sigma^2 \to 0$ . т.е. МОП-оценка стремится к МНК-оценке при неограниченном уменьшении уровня погрешностей в элементах

матрицы А.

МОП позволяет найти не только оценку искомого вектора и, но и переоценить неизвестные значения матрицы  $A_T$  и вектора  $f_T$ . Действительно, МОП-оценки  $\widehat{b}_{THi}$  неизвестных векторов  $b_{THi}$  принадлежат одному подпространству  $\widehat{S}_{Hm}$ , ортогональному вектору  $\widehat{x}_H^T = (\widehat{u}_H^T, 1)$ , а сами  $\widehat{b}_{THi}$  равны проекциям векторов  $b_{Hi}$  на подпространство  $\widehat{S}_{Hm}$ . Следовательно,  $\widehat{b}_{THi} = b_{Hi} - (b_{Hi}.\widehat{x}_H)\widehat{x}_H/\|\widehat{x}_H\|^2$ ,  $i=1,\ldots,n$ . Последние равенства запишем в эквивалентном виде

$$\widehat{a}_{Ti} = a_{i} - \frac{(a_{i}, \widehat{u}_{H}) - f_{Hi}}{\|\widehat{u}_{H}\|^{2} + 1} \widehat{u}_{H}, \quad \widehat{f}_{THi} = f_{Hi} + \frac{(a_{i}, \widehat{u}_{H}) - f_{Hi}}{\|\widehat{u}_{H}\|^{2} + 1} (9.45)$$

$$i = 1, \dots, n,$$

где  $a_i^T-i$ -я строка матрицы A,  $f_{\mathrm{H}i}-i$ -й элемент нормализованного вектора  $f_{\mathrm{H}}$ .

Векторно-матричная форма системы равенств (9.45) относи-

тельно исходных векторов u, f имеет вид

$$\widehat{A}_{\mathsf{MO\Pi}} = A + \sigma_a^2 \frac{\rho(\widehat{u}_{\mathsf{MO\Pi}}) \widehat{u}_{\mathsf{MO\Pi}}^T}{\sigma_a^2 \|\widehat{u}_{\mathsf{MO\Pi}}\|^2 + \sigma_f^2}, \ \widehat{f}_{\mathsf{MO\Pi}} = f - \sigma_f^2 \frac{\rho(\widehat{u}_{\mathsf{MO\Pi}})}{\sigma_a^2 \|\widehat{u}_{\mathsf{MO\Pi}}\|^2 + \sigma_f^2}, \ (9.46)$$

где  $\rho(\widehat{u}_{\mathsf{MOH}}) = f - A\widehat{u}_{\mathsf{MOH}}$ .

Заметим, что МОП-оценки  $A_T$  и  $f_T$  сходятся в среднем квадратичном к  $A_T$  и  $\widehat{f}_{\rm MHK}=A_{\rm T}\widehat{u}_{\rm MHK}$  соответственно при  $\sigma_a^2 \to 0$ .

Если  $\sigma_a^2$  неизвестна, то, используя ММП, можно получить оценку  $\widehat{\sigma}_a^2$ . Действительно,

$$\frac{\partial \ln L(B_{TH})}{\partial \sigma_a^2} = -\frac{n(m+1)}{2\widehat{\sigma}_a^2} + \frac{1}{2\widehat{\sigma}_a^4} \sum_{i=1}^n \sum_{j=1}^{m+1} (b_{Hij} - \widehat{b}_{THij})^2 = 0.$$

Отсюда  $\|\widehat{u}_{\text{МОП}}\|^2\widehat{\sigma}_a^2 = (n(m+1))^{-1}\|A\widehat{u}_{\text{МОП}} - f\|^2 - \sigma_f^2$ . Последнее равенство можно записать в виде

$$||A\widehat{u}_{MO\Pi} - f||^2 - \mu^2 ||\widehat{u}_{MO\Pi}||^2 = \delta^2,$$
 (9.47)

где  $\mu^2 = \hat{\sigma}_n^2 n(m+1), \ \delta^2 = \sigma_n^2 n(m+1).$ 

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

$$\widehat{u}_{\gamma} = \arg \min_{\|u\| \le K} \frac{\|Au - f\|^2}{\sigma_c^2 \|u\|^2 + \sigma_f^2}.$$
 (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

$$\gamma = \frac{\alpha(\sigma_f^2 + \sigma_a^2 ||\widehat{u}_\gamma||^2)}{\sigma_f^2} - \frac{\sigma_a^2 ||\widehat{Au}_\gamma - f||^2}{\sigma_a^2 ||\widehat{u}_\gamma||^2 + \sigma_f^2},$$

 $\alpha > 0$  — 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  $\sigma_a^2$  и  $\sigma_f^2$ . Значение

параметра  $\gamma$  определяется согласно равенству (9.47), записанному относительно  $\widehat{u}_{\gamma}$ :

$$||A\widehat{u}_{\gamma} - f||^2 - \mu^2 ||\widehat{u}_{\gamma}||^2 = \delta^2.$$
 (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  $\sigma_a^2=0$  РМОПоценка искомого вектора и определяется системой уравнений  $A^TA\hat{u}+\alpha\hat{u}=A^Tf$ , где параметр регуляризации  $\alpha>0$  определяется из равенства  $\|A\hat{u}-f\|^2=\delta^2$ . Таким образом, при отсутствии погрешностей в элементах матрицы A РМОП-оценки искомого вектора u совпадают с приближенным решением, даваемым вариационным методом регуляризации A. 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 [31, 43].

# 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  $\varepsilon_i$ . 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  $\mathbb{R}^m$  в про-

странство  $R^n$ .

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  $\varphi$  нелинейно зависит от компонент искомого вектора u; дискретизированные аналоги нелинейных интегральных уравнений, включая нелинейные интегральные уравнения первого рода.

### 1. МНК-оценки параметров нелинейных моделей и их свойства

Пусть компоненты  $\varepsilon_i$  вектора погрешностей  $\varepsilon$  имеют одина-ковую дисперсию  $\sigma^2$  и попарно не коррелируют.

Определим МНК-оценку искомого вектора *и* нелинейной регрессионной модели (10.1) равенством

$$\widehat{u}_{\text{MHK}} = \arg\min_{u \in R^m} ||A(u) - f||^2, \tag{10.2}$$

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  $\varepsilon$  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  $\varepsilon$  имеются корреляционные связи и известна корреляционная матрица  $K_{\varepsilon}=\sigma^2\Omega$  ( $\sigma^2$  может быть неизвестной). Тогда МНК-оценку определим равенством

$$\widehat{u}_{MHK} = \arg \min (A(u) - f)^T \Omega^{-1} (A(u) - f).$$
 (10.3)

Если  $\varepsilon$  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  $\operatorname{grad}_{\mathbf{x}} Q(u) = \mathbf{0}$ , определяющее МНК-оценку для нелинейных задач восстановления, является нелинейным.

Более того, для нелинейных задач восстановления функционал  $Q(\mathbf{u})$  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 є независимы и одинаково распределены;

 $2^{\circ} \|A(u_1) - A(u_0)\|^2/n$  равномерно относительно  $u_1, u_2 \in R$  сходится  $\kappa \varphi(u_1, u_0)$ , где R — ограниченное множество;

 $3^{\circ} \varphi(u_1, u_0) = 0$  тогда и только тогда, когда  $u_1 = u_0$ ;

 $4^{\circ}$  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

$$\frac{1}{n}\|A(u)-f\|^2=\frac{2}{n}\left(A(u_T)-A(u),\varepsilon\right)+\frac{1}{n}\|\varepsilon\|^2+\frac{1}{n}\|A(u)-A(u_T)\|^2.$$
 Далее,

 $\mathbf{M}\,\frac{1}{n}\left(A(u_T) - A(u), \varepsilon\right) = 0,$ 

$$\sigma^2\left(\frac{1}{n}\left(A(u_T) - A(u), \varepsilon\right)\right) = \frac{\sigma^2}{n^2} \|A(u_T) - A(u)\|^2 \to 0$$

при  $n \to \infty$ . Поэтому  $(A(u_T) - A(u), \varepsilon)/n$  сходится в среднем квадратичном к нулю. Так как  $\varepsilon_i^2$ ,  $i=1,\dots,n$ , распределены одинаково, то  $\|\varepsilon\|^2/n$  сходится по вероятности к  $\sigma^2$  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  $u^* = \lim_{n \to \infty} u_n$ . Из последнего неравенства имеем  $u^* = u_T$ . Теорема доказана.

Условие  $2^{\circ}$  теоремы 10.1 является аналогом сильной регулярности матриц  $A_n$  последовательности линейных регрессионных задач (7.14). Действительно, для линейной задачи

$$\frac{1}{n} ||A(u_1) - A(u_0)||^2 = 
= \frac{1}{n} ||A_n(u_1 - u_0)||^2 = (u_1 - u_0)^T \frac{1}{n} A_n^T A_n (u_1 - u_0).$$

Таким образом, выполнение условия  $2^\circ$  эквивалентно существованию предела  $A_n^TA_n/n\to L$  при  $n\to\infty$ , где L — положительно определенная матрица. Отметим, что неограниченность последо-

вательности 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  $\varepsilon$ .

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 МНКоченки  $\widehat{u}_{MHK}$  строить приближенные доверительные интервалы для компонент  $\widehat{u}_{Tj},\ j=1,\dots,m$ , искомого вектора  $u_T$ :

$$\begin{split} \widehat{u}_{\text{MHK}_j} - t(p_{\text{los}}) \sigma((P^T(\widehat{u}_{\text{MHK}})P(\widehat{u}_{\text{MHK}}))_{jj}^{-1})^{1/2} \leqslant \\ \leqslant u_{Tj} \leqslant \widehat{u}_{\text{MHK}_j} + t(p_{\text{los}}) \sigma((P^T(\widehat{u}_{\text{MHK}})P(\widehat{u}_{\text{MHK}}))_{jj}^{-1})^{1/2}, \end{split}$$

где 
$$p_{\text{дов}}$$
 — доверительная вероятность,  $\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{t(p_{\text{дов}})}e^{-s^2/2}\,ds=\frac{p_{\text{дов}}}{2}.$ 

Если  $\sigma^2$  неизвестна, то в качестве состоятельной оценки можно взять статистику  $s^2 = \|A(\widehat{u}_{\text{MHK}}) - f\|^2/n$ .

Заметим, что МНК-оценки нелинейной регрессионной модели (10.1), как правило, смещены. Ясно, что для состоятельных МНК-оценок смещение стремится к нулю при  $n \to \infty$ ,

### 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. Итерационный метод Ньютона для решения задачи (10.2) имеет вид

$$u^{(k+1)} = u^k - \left(\frac{\partial^2 Q}{\partial u^2}\right)^{-1} \cdot \nabla Q(u^{(k)}),$$
 (10.4)

где  $\frac{\partial^2 Q}{\partial u^2}$  — матрица вторых производных (гессиан) целевой функции Q(u), элемент которой  $\frac{\partial^2 Q}{\partial u_i \partial u_j}$  рассчитывается по формуле

емент которой  $\frac{\partial}{\partial u_i \partial u_j}$  рассчитывается по формул $\frac{\partial^2 Q}{\partial u_j} = \frac{\pi}{2} \frac{\partial \varphi(x_i; u)}{\partial \varphi(x_i; u)} \frac{\partial \varphi(x_i; u)}{\partial \varphi(x_i; u)}$ 

$$\frac{\partial^2 Q}{\partial u_i \partial u_j} = 2 \cdot \sum_{l=1}^n \frac{\partial \varphi(x_l; u)}{\partial u_j} \cdot \frac{\partial \varphi(x_l; u)}{\partial u_i} +$$

$$+2 \cdot \sum_{l=1}^{n} (\varphi(x_l; u) - f_l) \cdot \frac{\partial^2 \varphi(x_l; u)}{\partial u_i \partial u_j}.$$
 (10.5)

Если матрица  $\frac{\partial^2 Q}{\partial u^2}$  хорошо обусловлена в окрестности искомого решения и начальный вектор  $u^{(0)}$  взят из малой окрестности вектора  $u^{(0)}$ , то итерационный процесс (10.4) сходится к искомому вектору  $u_{\text{мин}}$ , причем скорость сходимости квадратичная.

К сожалению, итерационный процесс не всегда применим. Во-первых, расчет вторых производных (10.5) может быть затруднен или невозможен, например, если функция  $\varphi(x;u)$  не является дважды непрерывно дифференцируемой по u или она задается неаналитически. Во-вторых, матрица вторых производ-

ных  $\frac{\partial^2 Q}{\partial u^2}$  может быть плохо обусловлена в окрестности искомого решения  $u^*$ .

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

$$A(u) = A(u^{(k)}) + P_k(u - u^{(k)}), \tag{10.6}$$

176

где  $P_k = (n \times m)$ -матрица первых производных компонент вектор-функции A(u), вычисленных для значения  $u = u^{(k)}$ , т.е.

$$(P_k)_{ij} = \frac{\partial A_i(u^{(k)})}{\partial u_j}, \qquad i = 1, \dots, n, \quad j = 1, \dots, m.$$

Формула (10.6) определяет линеаризацию вектор-функции A(u) в окрестности точки  $u=u^{(k)}$ . Заменяя в (10.1) A(u) согласно (10.6), получаем вместо нелинейной задачи (10.1) линейную задачу восстановления:

$$P_k \delta u = \delta f - \varepsilon, \tag{10.7}$$

где  $\delta u = u - u^{(k)}$  — искомый вектор,  $\delta f = f - A(u^{(k)})$ .

МНК-оценка  $\delta u_{
m MHK}$  линейной регрессионной модели (10.7) дается равенством

$$\delta u_{\text{MHK}} = (P_k^T P_k)^{-1} P_k^T \delta f. \tag{10.8}$$

Положим  $u^{(k+1)} = u^{(k)} + \delta u_{\text{MHK}}$ . Из (10.8) получаем

$$u^{(k+1)} = u^{(k)} + (P_k^T P_k)^{-1} P_k^T (f - A(u^{(k)})).$$
 (10.9)

Формула (10.9) определяет итерационный процесс (ИП) Ньютона—Гаусса. Таким образом, последующее приближение ИП Ньютона—Гаусса является МНК-оценкой искомого вектора и для линеаризованной в окрестности предыдущего приближения исходной нелинейной регрессионной модели.

Задача 10.2. Пусть ковариационная матрица компонент вектора погрешностей  $\varepsilon$  определяется равенством  $K_{\varepsilon}=\sigma^2\Omega$ , где  $\Omega$  — 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 (10.9) к стационарной точке, обращающей в нуль  $\mathbf{grad}_{k} Q(u)$ , 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 (10.9) необходимо обращение симметричной положительно определенной  $(m \times m)$ -матрицы  $P_k^T P_k$ . При практической реализации ИП (10.9) 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 (10.9) является ИП гра-

диентного метода:

$$u^{(k+1)} = u^{(k)} + P_k^T(f - A(u^{(k)})). (10.10)$$

не требующий обращения матрицы  $P_k^T P_k$ .

Часто вместо ИП (10.9) и (10.10) используют, соответственно. ИП

$$u^{(k+1)} = u^{(k)} + \alpha_k (P_k^T P_k)^{-1} P_k^T (f - A(u^{(k)})), \tag{10.11}$$

$$u^{(k+1)} = u^{(k)} + \alpha_k P_k^T (f - A(u^{(k)})), \tag{10.12}$$

где оптимальное значение ускоряющего множителя  $\alpha_k$  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 (10.11) и  $v^{(k)} = P_k^T(f - A(u^{(k)}))$  для (10.12).

Для ИП (10.11) и (10.12) в качестве подходящих значений  $\alpha_k$  можно брать

 $\alpha_k = \frac{(P_k v^{(k)}, f - A(u^{(k)}))}{\|P_k v^{(k)}\|^2}.$ 

Часто основной вычислительной трудностью для реализации ИП Ньютона—Гаусса (10.9) является плохая обусловленность матрицы  $P_k^T P_k$  при некоторых значениях  $u^{(k)}$ . Левенбергом была предложена модификация ИП (10.9), свободная от этого недостатка. ИП Левенберга определяется равенством

$$u^{(k+1)} = u^{(k)} + (P_k^T P_k + \alpha_k I_m)^{-1} P_k^T (f - A(u^{(k)})),$$
 (10.13)

где параметр регуляризации  $lpha_k>0$  выбирается так, чтобы (m imes m)-матрица  $P_k^T P_k + lpha_k I_m$  была хорошо обусловлена.

Используется и другая регуляризованная модификация

ИП (10.9):

$$u^{(k+1)} = u^{(k)} + (P_k^T P_k + \alpha_k B_k)^{-1} P_k^T (f - A(u^{(k)})),$$
 (10.14)

где  $B_k$  — симметричные положительно определенные  $(m \times m)$ -матрицы. В частности, Марквардт предложил в качестве  $B_k$  выбирать диагональные матрицы  $B_k = \mathrm{diag}\,((P_k^T P_k)_{11}, \ldots, (P_k^T P_k)_{mm})$ , 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  $B\colon R^n\to R^l$ , что B(A(U))=Cu, где  $C-(I\times m)$ -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  $Cu=\varphi,\ u^{(1)}=0$ 

 $=(C^TC)^{-1}\hat{C}^T\varphi$ , rae  $\varphi=B(f)$ .

**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 [13, 31, 68, 118].

# 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 совокупно-

стью точек  $(t_k, y_k), k = 1, ..., n$ .

Одним из эффективных методов выделения детерминированной компоненты из экспериментальной функциональной зависимости является представление функциональной зависимости по базисной системе линейно независимых функций  $\{\varphi_j(t): j=1,\ldots,m\}$ 

$$y = y(t) = \sum_{j=1}^{m} u_j \varphi_j(t) + \varepsilon(t), \tag{11.1}$$

где  $\varepsilon(t)$  — хаотическая компонента,  $u_1, \dots, u_m$  — коэффициенты разложения, подлежащие оценке на основе известной совокупно-

сти точек  $(t_k, y_k)$ .

В настоящей и следующей главах мы будем трактовать хаотическую компоненту как случайный шум [42, 76]. Однако существует другой подход, в рамках которого хаотическая компонента трактуется как детерминированный динамический хаос [73, 105].

## 1. Метод сглаживающих ортогональных полиномов

Часто в качестве базисной системы функций в (11.1) берут нормальную систему полиномов. Напомним, что система полиномов  $\{\varphi_j(t),\ j=1,\dots,m\}$  называется нормальной, если в ней

присутствуют все полиномы степени 0, 1, ..., m-1. Не умаляя общности, можно считать, что  $\varphi_{i}(t)$  — полином степени i-1.

простой нормальной системой полиномов Наиболее 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  $(t_k, y_k)$ ,  $k = 1, \dots, n$ 

Действительно, МНК-оценка вектора  $u=(u_1,\ldots,u_m)^T$  ли-

нейной молели

$$y_k = \sum_{j=1}^m u_j \varphi_j(t_k) + \varepsilon(t_k), \qquad k = 1, \dots, n,$$
 (11.2)

нмеет вид (см. § 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  $A,\ \Omega$  — ковариационная матрица случайных величин  $\varepsilon(t_i),\ i=1,\ldots,n$ . Поэтому элементы матрицы  $A^TA$  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  $\Omega=\sigma^2I_n$ . Тогда

$$(A^T A)_{ij} = n \sum_{k=1}^n \frac{t_k^{i+j-2}}{n} \approx n \int_0^1 t^{i+j-2} dt = \frac{n}{i+j-1}, \quad i, j = 1, \dots, m.$$

Следовательно, матрица  $A^TA$  при достаточно большом nблизка к известной матрице Гильберта  $H_m$  (с точностью до множителя п). Но при увеличении порядка т 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  $\{t_k\}, \, k=1,\dots,n.$ 

Если в модели (11.2)  $\Omega = \mathrm{diag}\,(\sigma_1^2,\ldots,\sigma_n^2)$  и система базисных функций  $\{arphi_j(t)\},\ j=1,\ldots,n,$  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

$$\widehat{u}_j = \sum_{k=1}^n \frac{1}{\sigma_k^2} \varphi_j(t_k) y_k, \qquad j = 1, \dots, m.$$
 (11.4)

Устойчивость оценок (11.4) очевидна.

Для любого множества точек  $\{t_k\}$  и весов  $\{1/\sigma_k^2\},\ k=1,\dots$ , n, можно построить нормальную ортонормированную систе-

му полиномов любого порядка m.

Для построения ортонормированной системы полиномов можно использовать устойчивую схему Форсайта [81], согласно которой полиномы системы рассчитываются по рекуррентным формулам

$$a_{1} = \sum_{k=1}^{n} \frac{t_{k}}{\sigma_{k}^{2}} \left(\sum_{k=1}^{n} \frac{1}{\sigma_{k}^{2}}\right)^{-1}, \quad c_{1} = \left(\sum_{k=1}^{n} \frac{(t_{k} - a_{1})^{2}}{\sigma_{k}^{2}}\right)^{-1/2},$$

$$\varphi_{0}(t) = \left(\sum_{k=1}^{n} \frac{1}{\sigma_{k}^{2}}\right)^{-1/2}, \quad \varphi_{1}(t) = (t - a_{1})c_{1},$$

$$a_{i+1} = \sum_{k=1}^{n} \frac{t_{k}\varphi_{i}^{2}(t_{k})}{\sigma_{k}^{2}}, \quad b_{i+1} = \sum_{k=1}^{n} \frac{t_{k}\varphi_{i}(t_{k})\varphi_{i-1}(t_{k})}{\sigma_{k}^{2}}$$

$$\tilde{\varphi}_{i+1}(t) = (t - a_{i+1})\varphi_{i}(t) - b_{i+1}\varphi_{i-1}(t),$$

$$c_{i+1} = \left(\sum_{k=1}^{n} \frac{\tilde{\varphi}_{i+1}^{2}(t_{k})}{\sigma_{k}^{2}}\right)^{-1/2}, \quad \varphi_{i+1}(t) = c_{i+1}\tilde{\varphi}_{i+1}(t), \quad i = 1, 2, \dots, m.$$

После формирования ортонормированной нормальной системы полиномов по формулам (11.5) рассчитываем МНК-оценки параметров  $\widehat{u}=(\widehat{u}_1,\dots,\widehat{u}_m)^T$  по формулам (11.4).

Сглаженная детерминированная компонента  $y_{\text{дет}}(t)$  исходного временного ряда  $\{t_k,y_k\},\ k=1,\dots,n,$  определяется равенством

$$y_{\text{ger}}(t) = \sum_{j=0}^{m} \widehat{u}_j \varphi_j(t). \tag{11.6}$$

В частности, значения детерминированной компоненты в точ-ках  $t_k$  даются равенствами  $y_{\text{дет}}(t_k)=\sum_{j=0}^m \widehat{u}_j \varphi_j(t_k),\ k=1,\dots,n.$ 

Степень сглаживания зависит от числа членов разложения m. чем меньше m, тем больше степень сглаживания.

Подходящее значение т можно находить, используя различные критерии, основанные на анализе остатков (хаотическая компонента)

$$y_{\text{xaor}}(t_k) = y_k - y_{\text{ger}}(t_k), \qquad k = 1, \dots, n.$$
 (11.7)

Для выбора подходящего значения m можно использовать, например, показатель Херста [89, 93]. 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

$$y_k = y_{\text{MeT}}(t_k) + y_{\text{XMOT}}(t_k), \tag{11.8}$$

неправомерно и ведет к значительному искажению  $y_{\text{дет}}(t_k)$ .

При наличии больших выбросов вместо (11.8) необходимо использовать представление

$$y_k = y_{\text{net}}(t_k) + y_{\text{BMG}}(t_k) + y_{\text{xaor}}(t_k),$$
 (11.9)

где  $y_{\text{выо}}(t_k)$  — 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

$$\widehat{u}_{\text{po6}j} = \sum_{k=1}^{n} \frac{1}{\sigma_k^2} \varphi_j(t_k) \widetilde{y}_k, \qquad j = 1, \dots, m,$$
 (11.10)

где 
$$\widetilde{y}_k = \begin{cases} y_k, & \text{если } k \in I_0, \\ \sum\limits_{j=1}^m \widehat{u}_{\text{роб}j}\varphi_j(t_k) + K\sigma_k, & \text{если } k \in I_+, \\ \sum\limits_{j=1}^m \widehat{u}_{\text{роб}j}\varphi_j(t_k) - K\sigma_k, & \text{если } k \in I_-. \end{cases}$$

$$I_{0} = \left\{k: |y_{k} - \sum_{j=1}^{m} \widehat{u}_{\text{po6}j} \varphi_{j}(t_{k})| \leq K \sigma_{k}\right\},$$

$$I_{+} = \left\{k: |y_{k} - \sum_{j=1}^{m} \widehat{u}_{\text{po6}j} \varphi_{j}(t_{k}) > K \sigma_{k}\right\},$$

$$I_{-} = \left\{k: |y_{k} - \sum_{j=1}^{m} \widehat{u}_{\text{po6}j} \varphi_{j}(t_{k}) < -K \sigma_{k}\right\},$$

К - параметр Хьюбера.

Следовательно, итерационный процесс (8.15) для численного нахождения оценок (11.10), (11.11) принимает вид

ождения оценов (11.10), (11.17) примине 
$$I_0^{(l)}$$
,  $y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0^{(l)}, \\ \sum\limits_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k) + K \sigma_k, & \text{если } k \in I_+^{(l)}, \\ \sum\limits_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k) - K \sigma_k, & \text{если } k \in I_-^{(l)}, \end{cases}$   $u_j^{(l+1)} = \sum\limits_{k=1}^n \frac{1}{\sigma_k^2} \varphi_j(t_k) y_k^{(l)}, \quad j = 1, \dots, m,$   $I_0^{(l)} = \left\{ k: \ |y_k - \sum\limits_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k)| \leqslant K \sigma_k \right\}.$ 

$$I_{+}^{(l)} = \left\{ k : \ y_k - \sum_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k) > K \sigma_k \right\},$$

$$I_{-}^{(l)} = \left\{ k : \ y_k - \sum_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k) < -K \sigma_k \right\}.$$

Подходящее значение *т* можно находить, используя различные критерии, основанные на анализе остатков (хаотическая компонента)

$$y_{\text{xaor}}(t_k) = y_k - y_{\text{aer}}(t_k), \qquad k = 1, \dots, n.$$
 (11.7)

Для выбора подходящего значения m можно использовать, например, показатель Херста [89, 93]. 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

$$y_k = y_{\text{mer}}(t_k) + y_{\text{xaor}}(t_k), \tag{11.8}$$

неправомерно и ведет к значительному искажению  $y_{\text{дет}}(t_k)$ .

При наличии больших выбросов вместо (11.8) необходимо использовать представление

$$y_k = y_{\text{MeT}}(t_k) + y_{\text{BMG}}(t_k) + y_{\text{XAOT}}(t_k),$$
 (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

$$\widehat{u}_{pooj} = \sum_{k=1}^{n} \frac{1}{\sigma_k^2} \varphi_j(t_k) \widetilde{y}_k, \qquad j = 1, \dots, m,$$
 (11.10)

где

$$\widetilde{y}_k = \left\{egin{array}{ll} y_k, & ext{если } k \in I_0, \ \sum\limits_{j=1}^m \widehat{u}_{ ext{po6}j} arphi_j(t_k) + K \sigma_k, & ext{если } k \in I_+, \ \sum\limits_{j=1}^m \widehat{u}_{ ext{po6}j} arphi_j(t_k) - K \sigma_k, & ext{если } k \in I_-. \end{array}
ight.$$

Здесь

$$\begin{split} I_0 &= \Big\{ k: \ |y_k - \sum_{j=1}^m \widehat{u}_{\text{po6}j} \varphi_j(t_k)| \leqslant K \sigma_k \Big\}, \\ I_+ &= \Big\{ k: \ y_k - \sum_{j=1}^m \widehat{u}_{\text{po6}j} \varphi_j(t_k) > K \sigma_k \Big\}, \\ I_- &= \Big\{ k: \ y_k - \sum_{j=1}^m \widehat{u}_{\text{po6}j} \varphi_j(t_k) < -K \sigma_k \Big\}, \end{split}$$

K — параметр Хьюбера.

Следовательно, итерационный процесс (8.15) для численного нахождения оценок (11.10), (11.11) принимает вид

$$y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0^{(l)}, \\ \sum\limits_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k) + K\sigma_k, & \text{если } k \in I_+^{(l)}, \\ \sum\limits_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k) - K\sigma_k, & \text{если } k \in I_-^{(l)}, \end{cases}$$
 
$$u_j^{(l+1)} = \sum\limits_{k=1}^n \frac{1}{\sigma_k^2} \varphi_j(t_k) y_k^{(l)}, \qquad j = 1, \dots, m,$$
 
$$I_0^{(l)} = \left\{ k: \ |y_k - \sum\limits_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k)| \leqslant K\sigma_k \right\},$$
 
$$I_+^{(l)} = \left\{ k: \ y_k - \sum\limits_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k) > K\sigma_k \right\},$$
 
$$I_-^{(l)} = \left\{ k: \ y_k - \sum\limits_{j=1}^m \widehat{u}_j^{(l)} \varphi_j(t_k) < -K\sigma_k \right\}.$$

В качестве начального приближения  $u_j^{(0)}$ ,  $j=1,\ldots,m$ , итерационного процесса (11.12) можно брать неробастные оценки (11.4), т.е.

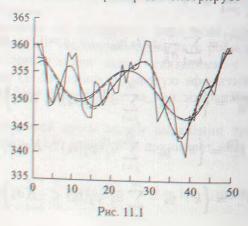
$$u_j^{(0)} = \sum_{k=1}^n \frac{1}{\sigma_k^2} \varphi_j(t_k) y_k, \qquad j = 1, \dots, m.$$
 (11.13)

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

$$y_{\text{мет}}(t_k) = \sum_{j=1}^{m} \widehat{u}_{\text{роб}j} \varphi_j(t_k),$$
 $y_{\text{мет}}(t_k) = \begin{cases} 0, & \text{если } k \in I_0, \\ y_k - y_{\text{дет}}(t_k) - K\sigma_k, & \text{если } k \in I_+, \\ y_k - y_{\text{дет}}(t_k) + K\sigma_k, & \text{если } k \in I_-, \end{cases}$ 
 $y_{\text{меот}}(t_k) = y_k - y_{\text{дет}}(t_k) - y_{\text{выб}}(t_k), \quad k = 1, \dots, n.$ 
(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. Пример демонстрирует влияние мак-



симальной степени полиномов m на результат сглаживания. При увеличении *m* сглаживающий график приближается к исхол-HOMV.

Метод сглаживающих робастных полиномов можно использовать для конструирования алгоритмов, позволяющих оптимальным образом объединять экспериментальные данные при восстановлении функциональных зависимостей [24].

### 2. Метод сглаживающих линейных сплайнов

Проведение операции сглаживания, как правило, априори предполагает, что рассматривается задача восстановления функциональной зависимости y=y(t) с непрерывным изменением aргумента t, т. е. восстановленная функциональная зависимость должна позволять подсчитывать значения функции u(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 у не имеют никакого смысла в рамках рассматриваемой задачи сглаживания.

При рассмотрении такого рода задач по существу речь идет о сглаживании только совокупности «экспериментальных» точек  $(t_k, y_k), k = 1, ..., n$ , и исследователя совершенно не интересуют промежуточные значения аргумента t, принадлежащие открытым интервалам  $t_k, t_{k+1}, k = 1, ..., n-1$ .

Представленные ниже сглаживающие линейные сплайны дают возможность наиболее экономно и в то же время эффективно производить операцию сглаживания дискретной совокупности точек  $(t_k, y_k)$ , k = 1, ..., 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 сглаживаюшими сплайнами (см. § 4).

Напомним, что интерполяционным сплайном степени (линейным интерполяционным сплайном) называется кусочно-линейная функция S(t), проходящая через заданные

точки  $(t_k, y_k)$ , k = 1, ..., n,  $t_1 < t_2 < ... < t_n$ .

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, ..., 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  $h_k = t_{k+1} - t_k$ ,  $k = 1, \dots, n-1$ .

**Лемма 11.1.** Среди всех функций  $f(t) \in W_1$  линейный интерполяционный сплайн S(t) и только он минимизирует функционал

$$J(f) = \int_{t_1}^{t_n} (f'(t))^2 dt.$$
 (11.15)

Доказательство. Пусть f(t) — произвольная функция из класса  $W_1$ . Имеем: J(S-f)=J(f)-J(S)+Q, где

$$Q = 2 \int_{t}^{t} S'(t)(S'(t) - f'(t)) dt.$$

Получаем следующую цепочку равенств:

$$Q = 2\sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} S'(t)(S'(t) - f'(t)) dt =$$

$$= 2\sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} \frac{y_{k+1} - y_k}{h_k} \left(\frac{y_{k+1} - y_k}{h_k} - f'(t)\right) dt =$$

$$= 2\sum_{k=1}^{n-1} \left[\frac{(y_{k+1} - y_k)^2}{h_k} - \frac{y_{k+1} - y_k}{h_k}(y_{k+1} - y_k)\right] = 0.$$

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.

$$\int_{t_h}^{t_{k+1}} (f'^*(t) - S'(t))^2 dt = 0, \quad k = 1, \dots, n-1.$$

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

$$f^*(t) = \arg \min_{f(t)} J_{\alpha}(f(t)) =$$

$$= \arg \min_{f(t)} \left\{ \sum_{k=1}^{n} \frac{(y_k - f(t_k))^2}{\sigma_k^2} + \alpha \int_{t_1}^{t_n} (f'(t))^2 dt \right\}, \quad (11.16)$$

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

$$\int_{t_1}^{t_n} (S'_{\alpha}(t))^2 dt \le \int_{t_1}^{t_n} (f'^*(t))^2 dt.$$

Следовательно,  $J_{lpha}(S_{lpha}(t))\leqslant J_{lpha}(f^{st}(t))$ , 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  $S_k = S_{\alpha}(t_k), \ k=1,\ldots,n$ ,  $S=(S_1,\ldots,S_n)^T$ . 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

$$\frac{1}{2} \frac{\partial J_{\alpha}}{\partial S_{1}} = \frac{S_{1} - y_{1}}{\sigma_{1}^{2}} + \frac{\alpha}{h_{1}} (S_{1} - S_{2}) = 0,$$

$$\frac{1}{2} \frac{\partial J_{\alpha}}{\partial S_{k}} = \frac{S_{k} - y_{k}}{\sigma_{k}^{2}} + \frac{\alpha}{h_{k-1}} (S_{k} - S_{k-1}) + \frac{\alpha}{h_{k}} (S_{k} - S_{k+1}) = 0,$$

$$k = 2, \dots, n - 1,$$

$$\frac{1}{2} \frac{\partial J_{\alpha}}{\partial S_{n}} = \frac{S_{n} - y_{n}}{\sigma_{n}^{2}} + \frac{\alpha}{h_{n-1}} (S_{n} - S_{n-1}) = 0.$$

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

Матрица A — положительно определенная с диагональным преобладанием. Следовательно, существует единственное решение S системы (11.17), которое, в частности, можно найти с помощью метода прогонки [41].

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 
$$\widetilde{J}_{\alpha}(S(t))$$
,  $S(t_1) = y_1$ ,  $S(t_n) = y_n$ , (11.18)

где

$$\widetilde{J}_{\alpha}(S(t)) = \sum_{k=2}^{n-1} \frac{(y_k - S(t_k))^2}{\sigma_k^2} + \alpha \int_{t_1}^{t_n} (S'(t))^2 dt.$$

 $\overline{O}$ бозначим  $\overline{S}_0 = (S_2, \dots, S_{n-1})^T$ . 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

$$\frac{1}{2}\frac{\partial \widetilde{J}_{\alpha}(\overline{S}_{0})}{\partial S_{2}} = \frac{S_{2} - y_{2}}{\sigma_{2}^{2}} + \frac{\alpha}{h_{1}}(S_{2} - y_{1}) + \frac{\alpha}{h_{2}}(S_{2} - S_{3}) = 0,$$

$$\frac{1}{2}\frac{\partial \widetilde{J}_{\alpha}(\overline{S}_0)}{\partial S_k} = \frac{S_k - y_k}{\sigma_k^2} + \frac{\alpha}{h_{k-1}}(S_k - S_{k-1}) + \frac{\alpha}{h_k}(S_k - S_{k+1}) = 0,$$

$$k=3,\ldots,n-2,$$

$$\frac{1}{2} \frac{\partial \widetilde{J}_{\alpha}(\overline{S}_{0})}{\partial S_{n-1}} = 
= \frac{S_{n-1} - y_{n-1}}{\sigma_{n-1}^{2}} + \frac{\alpha}{h_{n-2}} (S_{n-1} - S_{n-2}) + \frac{\alpha}{h_{n-1}} (S_{n-1} - y_{n}) = \mathbf{0}.$$

Следовательно, для определения вектора  $S_0$  получаем систему

 $A_0 \overline{S}_0 = \overline{F}_0, \tag{11.19}$ 

где симметричная трехдиагональная (n-2) imes (n-2)-матрица  $A_0$  и вектор  $\overline{F}_0$  задаются равенствами

$$A_{0} = \begin{bmatrix} \frac{1}{\sigma_{2}^{2}} + \frac{\alpha}{h_{1}} + \frac{\alpha}{h_{2}} & -\frac{\alpha}{h_{2}} & 0 & \dots & 0 & 0 \\ -\frac{\alpha}{h_{2}} & \frac{1}{\sigma_{3}^{2}} + \frac{\alpha}{h_{2}} + \frac{\alpha}{h_{3}} & -\frac{\alpha}{h_{3}} & \dots & 0 & 0 \\ 0 & 0 & \dots & \frac{-\alpha}{h_{n-2}} & \frac{1}{\sigma_{n-1}^{2}} + \frac{\alpha}{h_{n-2}} + \frac{\alpha}{h_{n-1}} \end{bmatrix}$$

$$(11.20)$$

$$\overline{F}_0 = \left(\frac{y_2}{\sigma_2^2} + \frac{\alpha}{h_1} y_1, \frac{y_3}{\sigma_3^2}, \dots, \frac{y_{n-2}}{\sigma_{n-2}^2}, \frac{y_{n-1}}{\sigma_{n-1}^2} + \frac{\alpha}{h_{n-1}} y_n\right)^T.$$

А<sub>0</sub> — 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  $S_{lpha}(t)$  в этом случае определяется равенствами

$$S_{\alpha}(t_1) = y_1$$
,  $S_{\alpha}(t_k) = S_k$ ,  $k = 2, ..., n - 1$ ,  $S_{\alpha}(t_n) = y_n$ .

# 3. Робастные линейные сглаживающие сплайны

Для построения линейных робастных сглаживающих сплайнов применим схему получения робастных М-оценок.

Рассмотрим робастный функционал

$$J_{\alpha}(S) = \sum_{k=1}^{n} \rho\left(\frac{y_k - S(t_k)}{\sigma_k}\right) + \alpha \int_{t_k}^{t_n} (S'(t))^2 dt, \tag{11.21}$$

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 (11.21) в классе кусочнонепрерывных 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

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

Введем три множества индексов:

$$I_{0}(\overline{S}) = \{k : |y_{k} - S_{k}| \leq K\sigma_{k}\},$$

$$I_{+}(\overline{S}) = \{k : y_{k} - S_{k} > K\sigma_{k}\},$$

$$I_{-}(\overline{S}) = \{k : y_{k} - S_{k} < -K\sigma_{k}\}.$$

$$(11.22)$$

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

$$\frac{1}{2} \frac{\partial J_{\alpha}(\overline{S})}{\partial S_{1}} = \frac{S_{1} - \widetilde{y}_{1}}{\sigma_{1}^{2}} + \frac{\alpha}{h_{1}} (S_{1} - S_{2}) = 0,$$

$$\frac{1}{2} \frac{\partial J_{\alpha}(\overline{S})}{\partial S_{k}} = \frac{S_{k} - \widetilde{y}_{k}}{\sigma_{k}^{2}} + \frac{\alpha}{h_{k-1}} (S_{k} - S_{k-1}) + \frac{\alpha}{h_{k}} (S_{k} - S_{k+1}) = 0,$$

$$k = 2, \dots, n - 1,$$

$$\frac{1}{2} \frac{\partial J_{\alpha}(\overline{S})}{\partial S_{n}} = \frac{S_{n} - \widetilde{y}_{n}}{\sigma^{2}} + \frac{\alpha}{h_{n-1}}, (S_{n} - S_{n-1}) = 0,$$
(11.23)

где

$$\widetilde{y}_k = \left\{ egin{array}{ll} y_k, & ext{если } k \in I_0(\overline{S}), \ S_k + K\sigma_k, & ext{если } k \in I_+(\overline{S}), \ S_k - K\sigma_k, & ext{если } k \in I_-(\overline{S}). \end{array} 
ight.$$

Следовательно, для определения искомого вектора  $\overline{S}$  получаем систему  $A\overline{S} = \widetilde{F}$ 

где матрица A определена равенством (11.17), а

$$\widetilde{F} = (\widetilde{y}_1/\sigma_1^2, \dots, \widetilde{y}_n/\sigma_n^2)^T.$$

Система (11.24) нелинейна, поскольку компоненты вектора  $\widehat{F}$ зависят от искомого решения S.

Для решения системы (11.24) используем эффективный ите-

рационный процесс

$$A\overline{S}^{(l+1)} = \overline{F}^{(l)}, \tag{11.25}$$

где

$$\widetilde{F}^{(l)} = \left(rac{y_1^{(l)}}{\sigma_1^2}, \ldots, rac{y_n^{(l)}}{\sigma_n^2}
ight)^T, \ y_k^{(l)} = \left\{ egin{array}{ll} y_k, & ext{если } k \in I_0(\overline{S}^{(l)}), \\ S_k^{(l)} + K\sigma_k, & ext{если } k \in I_+(\overline{S}^{(l)}), \\ S_k^{(l)} - K\sigma_k, & ext{если } k \in I_-(\overline{S}^{(l)}), \end{array} 
ight.$$

а множества  $I_0(\overline{S}^{(l)}),\ I_+(\overline{S}^{(l)}),\ I_-(\overline{S}^{(l)})$  определены равенствами (11.22).

B качестве начального  $\widetilde{F}^{(0)}$  рекомендуется брать  $\widetilde{F}^{(0)}=\overline{F}=$ 

 $= (y_1/\sigma_1^2, \dots, y_n/\sigma_n^2)^T.$ 

Аналогично строится итерационная процедура нахождения линейного робастного сглаживающего сплайна с фиксацией крайних точек  $S_{\alpha}(t_1)=y_1,\ S_{\alpha}(t_n)=y_n$ :

$$A_0 \, \overline{S}_0^{(l+1)} = \overline{F}_0^{(l)}, \tag{11.26}$$

гле

$$\overline{S}_0^{(l)} = (S_2^{(l)}, \dots, S_{n-1}^{(l)})^T,$$
 
$$\overline{F}_0^{(l)} = \left(\frac{y_2^{(l)}}{\sigma_2^2} + \frac{\alpha}{h_1} y_1, \frac{y_3^{(l)}}{\sigma_3^2}, \dots, \frac{y_{n-2}^{(l)}}{\sigma_{n-2}^2}, \frac{y_{n-1}^{(l)}}{\sigma_{n-1}^2} + \frac{\alpha}{h_{n-1}} y_n\right)^T,$$
 
$$y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}_0^{(l)}), \\ S_k^{(l)} + K\sigma_k, & \text{если } k \in I_+(\overline{S}_0^{(l)}), \end{cases} \qquad k = 2, \dots, n-1.$$
 
$$S_k^{(l)} - K\sigma_k, & \text{если } k \in I_-(\overline{S}_0^{(l)}), \end{cases}$$

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 [95, 104]

$$\rho(t) = \begin{cases} t^2, & |t| \leqslant R, \\ R^2, & |t| > R. \end{cases}$$
 (11.27)

Введем два множества индексов:

$$I_{\overline{G}}(\overline{S}) = \{k : |y_k - S_k| \le R\sigma_k\}, \quad I_{\overline{R}}(\overline{S}) = \{k : |y_k - S_k| > R\sigma_k\},\$$

$$(11.28)$$

и функционал  $J_{\alpha}(\overline{S})$  вида (11.21), где  $\rho(t)$  определена равенством (11.27). Каждый индекс  $k=1,\ldots,n$  принадлежит одному и только одному из множеств  $I_0(\overline{S}),\,I_R(\overline{S}).$ 

Для производных  $\frac{\partial J_{\alpha}(\overline{S})}{\partial S_k}$ ,  $k=1,\ldots,n$ , имеем систему равенств (11.23), где

$$\overline{y}_k = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}), \\ S_k, & \text{если } k \in I_R(\overline{S}), \end{cases}$$
  $k = 2, \dots, n-1.$  (11.29)

Следовательно, для определения вектора  $\overline{S}$  получаем систему (11.24), где компоненты вектора  $\overline{F}=(\widetilde{y}_1,\ldots,\widetilde{y}_n)^T$  даются равенствами (11.29).

Для решения системы (11.24), (11.29) используем эффективный итерационный процесс

$$A\overline{S}^{(l+1)} = \overline{F}^{(l)},$$

где

$$\overline{F}^{(l)} = \left(rac{y_1^{(l)}}{\sigma_1^2}, \dots, rac{y_n^{(l)}}{\sigma_n^2}
ight)^T, \quad y_k^{(l)} = \left\{egin{array}{l} y_k, & ext{ если } k \in I_0(\overline{S}^{(l)}), \\ S_k, & ext{ если } k \in I_R(\overline{S}^{(l)}), \end{array}
ight.$$

а множества  $I_0(\overline{S}^{(l)}),\ I_R(\overline{S}^{(l)})$  определены равенствами (11.28). Аналогично строится итерационная процедура нахождения

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

$$\overline{S}_0^{(l)} = (S_2^{(l)}, \dots, S_{n-1}^{(l)})^T,$$
 
$$\overline{F}_0^{(l)} = \left(\frac{y_2^{(l)}}{\sigma_2^2} + \frac{\alpha}{h_1}y_1, \frac{y_3^{(l)}}{\sigma_3^2}, \dots, \frac{y_{n-2}^{(l)}}{\sigma_{n-2}^2}, \frac{y_{n-1}^{(l)}}{\sigma_{n-1}^2} + \frac{\alpha}{h_{n-1}}y_n\right)^T,$$
 
$$y_k^{(l)} = \begin{cases} y_k, & \text{если } k \in I_0(\overline{S}^{(l)}), \\ S_k^{(l)}, & \text{если } k \in I_R(\overline{S}^{(l)}), \end{cases}$$
  $k = 2, \dots, n-1,$ 

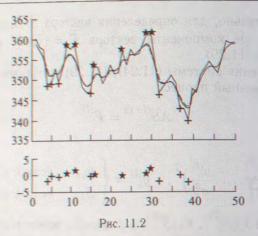
а множества  $I_0(\overline{S})$ ,  $I_R(\overline{S})$  определены равенствами (11.28).

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  $(\star, +)$  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 (11.25).

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  $P_k$ ,



 $k=1,\dots,n$ , цену k-го выпуска (в процентах от номинальной стоимости), имеющего на рассматриваемую дату срок до погашения  $\tau_k$  дней. Тогда совокупность точек  $(\tau_k,P_k),\ k=1,\dots,n$ , 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-то выпуска на текущую дату.

Выпуски, для которых  $P_k > P_{{
m crn}k}$  ( $P_k < P_{{
m crn}k}$ ) называются

переоцененными (недооцененными) рынком.

Динамическое управление портфелем выпусков облигаций состоит в перевложении средств из переоцененных выпусков облигаций в недооцененные. Аналогично может решаться задача формирования портфеля акций [50, 103].

## 4. Метод сглаживающих кубических сплайнов

В настоящем параграфе рассматривается задача восстановления функциональной зависимости, когда независимая перемен-

ная 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 ... n. Задача осложняется тем, что измерения  $y_i$  производятся с ошибками  $\xi_i$  имеющим 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  $\widehat{y}=f(x)$ , 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 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 интерполяцион-

ных сплайнов.

В дальнейшем, не умаляя общности, предполагаем, что узлы  $\{x_i, i=1,\ldots,n\}$ , упорядочены, т. е.

$$a = x_1 < x_2 < \dots < x_n = b. (11.30)$$

Функция  $S_n(x)$  называется кубическим сплайном с узлами (11.30), если: 1° на каждом интервале  $[x_i,x_{i+1}]$   $S_n(x)$  — кубический полином, т. е.

$$S_n(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3,$$
 (11.31)

 $2^{\circ} S_n(x)$  — дважды непрерывно дифференцируема на [a,b]. Кубический сплайн называется интерполяционным, если

$$S_n(x_i) = y_i, i = 1, ..., n.$$
 (11.32)

Из (11.31) следует, что  $y_i = a_i, i = 1, \ldots, n-1$ .

Для построения кубического интерполяционного сплайна необходимо знать 4(n-1) коэффициентов  $\{a_i,b_i,c_i,d_i,i=1,\dots,n-1\}$ . Из условий непрерывности  $S_n(x)$ ,  $S'_n(x)$ ,  $S''_n(x)$  во всех внутренних узлах  $x_i$ ,  $i=2,\dots,n-1$ , получаем 3(n-2) равенств. Кроме того, имеем n равенств интерполяции (11.32). 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 условилими. Обычно используют следующие типы краевых условий:

 $1^{\circ} S'_n(x_1) = m_1, S'_n(x_n) = m_2;$ 

$$2^{s} S_{n}^{u}(x_{1}) = m_{1}, S_{n}^{u}(x_{n}) = m_{2};$$

$$3^{s} S_{n}^{m}(x_{1}) = m_{1}, S_{n}^{m}(x_{2}) = m_{2};$$

$$4^{s} S_{n}^{(k)}(x_{1}) = S_{n}^{(k)}(x_{n}), k = 0, 1.$$

$$S_n^{(k)}(x_1) = S_n^{(k)}(x_n), k = 0, 1.$$

Часто используют краевое условие  $2^{\circ}$  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 (11.30)  $M_i = S_n''(x_i)$ , i = 1, ..., n, а затем восстанавливаются коэффициенты  $b_i$ ,  $c_i$ ,  $d_i$ , i = 1, ..., n-1, по формулам

$$b_i = \frac{y_{i+1} - y_i}{h_i} - h_i \frac{M_{i+1} + 2M_i}{6}, \quad c_i = \frac{M_i}{2}, \quad d_i = \frac{M_{i+1} - M_i}{6h_i},$$
(11.33)

rae  $h_i = x_{i+1} - x_i$ , i = 1, ..., 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 интерполяшни (11.32). Получаем п - 2 равенства:

$$\frac{h_{i-1}}{6}M_{i-1} + \frac{h_{i-1} + h_i}{3}M_i + \frac{h_i}{6}M_{i+1} = \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}},$$

$$i = 2, \dots, n-1.$$
(11.34)

Два недостающих уравнения получаем из краевых условий, которые относительно  $M_i, i = 1, \dots, n$ , имеют вид

$$u_1M_1 + u_2M_2 = D_1, \quad v_1M_{n-1} + v_2M_n = D_2,$$
 (11.35)

гле  $u_1, u_2, v_1, v_2, D_1, D_2$  в зависимости от типа краевых условий даются следующими равенствами:

1) для краевых условий типа 1°  $u_1 = h_1/3$ ,  $u_2 = h_1/6$ ,  $v_1 =$  $=h_{n-1}/6$ ,  $v_2=h_{n-1}/3$ ,  $D_1=(y_2-y_1)/h_1-m_1$ ,  $D_2=(y_{n-1}-y_1)/h_1$  $-y_n/h_{n-1}+m_2$ ;

2) для краевых условий типа  $2^\circ u_1 = 1$ ,  $u_2 = 0$ ,  $v_1 = 0$ ,  $v_2 = 1$ ,  $D_1 = m_1, D_2 = m_2$ 

3) для краевых условий типа 3°  $u_1 = 1$ ,  $u_2 = -1$ ,  $v_1 = -1$ ,  $v_2=1$ ,  $D_1=-m_1h_1$ ,  $D_2=m_2h_{n-1}$ .

Совокупность n равенств (11.34), (11.35) — квадратная система п линейных алгебраических неоднородных уравнений с п неизвестными  $M_i$ , i = 1, ..., n.

Систему (11.34), (11.35) относительно искомого n-мерного вектора  $M = (M_1, \dots, M_n)^T$  запишем в векторно-матричной

$$AM = Hy + f, \tag{11.36}$$

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  $1^{\circ} \ f_1 = -m_1$ ,  $f_n = -m_2$ , для остальных краевых условий  $f_1 = D_1$ ,  $f_n = D_2$ . Матрица А представима в виде

$$A = \begin{pmatrix} A_1 \\ A_0 \\ A_2 \end{pmatrix},$$

где 
$$A_0 = \begin{pmatrix} \frac{h_1}{6} & \frac{h_1 + h_2}{3} & \frac{h_2}{6} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{h_2}{6} & \frac{h_2 + h_3}{3} & \frac{h_3}{6} & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & \frac{h_{n-2}}{6} & \frac{h_{n-2} + h_{n-1}}{3} & \frac{h_{n-1}}{6} \end{pmatrix}$$

 $-((n-2)\times n)$ -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 Н представима в виде

$$H = \begin{pmatrix} H_1 \\ H_0 \\ H_2 \end{pmatrix}.$$

$$H_0 = \begin{pmatrix} \frac{1}{h_1} & -\left(\frac{1}{h_1} + \frac{1}{h_2}\right) & \frac{1}{h_2} & 0 & \dots \\ 0 & \frac{1}{h_2} & -\left(\frac{1}{h_2} + \frac{1}{h_3}\right) & \frac{1}{h_3} & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots \end{pmatrix}$$

- ((n-2) imes 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 (11.36) 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 (11.36). 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 узлами (11.30), наиболее компактной является схема Горнера:

$$S_n(x) = y_i + t(b_i + t(c_i + td_i)),$$
  

$$S_n'(x) = b_i + t(2c_i + 3td_i),$$
  

$$x_i \le x \le x_{i+1}, \quad t = x - x_i.$$
(11.37)

Для подсчета  $S_n(x)$  и  $S_n'(x)$  по формулам (11.37) необходимо предварительно подсчитать  $\{b_i,c_i,d_i,\ i=1,\ldots,n-1\}$  по формулам (11.33).

Если число вычисляемых значений кубического интерполяционного сплайна меньше 1.1n, то целесообразно вместо формул (11.37) использовать формулы

$$S_{n}(x) = y_{i} + t \left\{ \frac{y_{i+1} - y_{i}}{h_{i}} - \frac{(x_{i+1} - x + h_{i})M_{i} + (t + h_{i})M_{i+1}}{6} \right\}.$$

$$S'_{n}(x) = \frac{y_{i+1} - y_{i}}{h_{i}} + \frac{t(x - x_{i+1})\frac{(x_{i+1} - x + h_{i})M_{i} + t(t + h_{i})M_{i+1}}{6} + t(x - x_{i+1})\frac{(2t + h_{i})M_{i+1} - M_{i}}{6}, \quad (11.38)$$

$$x_{i} \leqslant x \leqslant x_{i+1}.$$

Интерполяционные кубические сплайны обладают следующим свойством сходимости. Обозначим  $h_{\max} = \max h_i$ . 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

$$\sup_{x \in [a,b]} |f^{(k)}(x) - S_n^{(k)}(x)| \le C_k h_{\text{max}}^{3-k}, \quad k = 0, 1, 2, \tag{11.39}$$

где  $C_k$  не зависят от узлов (11.30), 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

$$J(\varphi) = \int_{a}^{b} (\varphi''(x))^2 dx. \tag{11.40}$$

Поставим задачу на минимум функционала (11.40) 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  $1^\circ-4^\circ$ . Обозначим класс таких интерполяционных функций через  $V_{2k}(x),\ k=1,2,3,4,$  в зависимости от выполнения краевых условий  $K^\circ$ .

**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  $K^{\circ}$ .

В частности, интерполяционный кубический сплайн  $S_n(x)$ , удовлетворяющий нулевым краевым условиям  $2^\circ$  ( $m_1=m_2=0$ ), доставляет минимум функционалу  $J(\varphi)$  в классе дважды непре-

рывно дифференцируемых интерполяционных функций.

Функционал (11.40) 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, ..., n, с шарнирным закреплением в этих точках.

Тогда функционал (11.40) 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 сплай-

HOB.

Пусть случайные величины  $\xi_i$  (ошибки в измерении  $y_i$ ),  $i=1,\dots,n$ , 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

$$J_{\alpha}(\varphi) = \sum_{i=1}^{n} \frac{(y_i - \varphi(x_i))^2}{\sigma_i^2} + \alpha \int_{a}^{b} (\varphi''(x))^2 dx,$$
 (11.41)

где  $\alpha > 0$  — параметр сглаживания.

Поставим задачу на минимум функционала (11.41) в классе  $W_k$  дважды непрерывно дифференцируемых функций, удовлетворяющих условиям  $K^{\circ}$ . Подчеркнем, что в отличие от экстремальной задачи на минимум функционала (11.40) не требуется выполнения интерполяционных равенств  $\varphi(x_i) = y_i, i = 1, ..., n$ .

**Теорема 11.2.** Решением задачи на минимум функционана (11.41) в классе  $W_k$  является кубический сплайн, удовлетворяющий краевым условиям  $K^{\circ}$ .

Доказательство. Пусть функция  $\varphi^*(x) \in W_k$  доставляет минимум функционалу (11.41). Обозначим  $y_i^* = \varphi^*(x_i), \ i=1,\dots,n,$  и построим интерполяционный кубический сплайн  $S_n^*(x)$ , удовлетворяющий интерполяционным условиям  $S_n^*(x_i) = y_i^*, \ i=1,\dots,n$  и краевым условиям  $K^\circ$ . Из теоремы 11.1 следует, что

 $\int_{a} (S_n^*(x))^2 dx \le \int_{a} (\varphi^{*n}(x))^2 dx. \text{ Ho}$ 

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2} (y_i - \varphi^*(x_i)^a)^2 = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} (y_i^a - S_n^*(x_i))^2.$$

Поэтому  $J_{\alpha}(S_n^*) \leqslant J_{\alpha}(\varphi)$ . Поскольку функционал (11.41) в классе функций  $W_k$  сильно выпуклый, то решение задачи на минимум функционала (11.41) 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 (11.41) приводит к следующему замечательному результату относительно искомого сглаживающего сплайна  $S_{n\alpha}(x)$  [37].

**Теорема 11.3.** Кубический сплайн  $S_{n\alpha}(x)$  является сглаживающим тогда и только тогда, когда выполнена система равенств

$$S_{n\alpha}(x_i) + \alpha \sigma_i^2 Z_i = y_i, \qquad i = 1, \dots, n, \tag{11.42}$$

ede 
$$Z_1 = S_{n\alpha}^{\prime\prime\prime}(a), Z_i = S_{n\alpha}^{\prime\prime\prime}(x_i + 0) - S_{n\alpha}^{\prime\prime\prime}(x_i - 0), i = 1, 2, \dots, n - 1, Z_n = S_{n\alpha}^{\prime\prime\prime}(b).$$

Система равенств (11.42) дает возможность построить эффективный численный алгоритм нахождения сглаживающего кубического сплайна. Обозначим

$$Z = (Z_1, ..., Z_n)^T, \quad y = (y_1, ..., y_n)^T,$$
  
 $S_{\alpha} = (S_{\alpha 1}, ..., S_{\alpha n})^T, \quad M = (M_1, ..., M_n)^T,$ 

где  $S_{\alpha i} = S_{n\alpha}(x_i)$ ,  $M_i = S''_{n\alpha}(x_i)$ , i = 1, ..., n. Введем матрицу  $K_{\xi} = \mathrm{diag}\,(\sigma_1^2, \ldots, \sigma_n^2)$ .

Тогда систему уравнений (11.42) можно записать в векторноматричной форме:  $S_{\alpha} + \alpha K_{\mathcal{E}} Z = y. \tag{11.43}$ 

Векторы  $S_{\alpha}$  и Z можно выразить через вектор M. Действительно, для кубических сплайнов справедливы равенства

$$S_n^m(x_i+0) = \frac{M_{i+1} - M_i}{h_i}, \quad i = 1, \dots, n-1,$$
  
$$S_n^m(x_i-0) = \frac{M_i - M_{i-1}}{h_{i-1}}, \quad i = 2, \dots, n,$$

из которых следует представление

$$Z_1 = \frac{M_2 - M_1}{h_1}, \quad Z_i = \frac{M_{i+1} - M_i}{h_i} - \frac{M_i - M_{i-1}}{h_{i-1}}, \quad i = 2, ..., n-1,$$

$$Z_n = \frac{M_{n-1} - M_n}{h_{n-1}}.$$

Тогда систему предыдущих равенств можно записать в векторно-матричной форме:

$$Z = H^T M, (11.44)$$

где матрица Н определена в формуле (11.36).

Из системы (11.36), справедливой для любого кубического сплайна. для сплайна  $S_{n\alpha}(x)$  следует соотношение

$$HS_{\alpha} = AM - f. \tag{11.45}$$

Подействуем на обе части системы (11.43) матрицей H и выразим Z и  $HS_{\alpha}$  через M согласно равенствам (11.44) и (11.45). Получаем систему линейных алгебраических уравнений относительно неизвестного вектора M:

$$(A + \alpha H K_{\xi} H^T)M = Hy + f. \tag{11.46}$$

Поскольку матрица A трехдиагональна, а матрица  $HK_{\xi}H^T$  пятидиагональна, матрица  $A+\alpha HK_{\xi}H^T$  пятидиагональна.

Для всех типов 1°-4° краевых условий матрицы A и  $HK_\xi H^T$  положительно определены. Поэтому матрица  $A + \alpha HK_\xi H^T$  положительно определена для любого  $\alpha \geqslant 0$ .

Итак, система (11.46) для любого  $\alpha\geqslant 0$  имеет единственное

решение.

Поскольку матрица  $A + \alpha H K_\xi H^T$  пятидиагональна, система (11.46) эффективно решается методом обобщенной прогонки, требующим для своей реализации  $\sim 30n$  арифметических операции [41].

После решения системы (11.46) значения сглаживающего кубического сплайна  $S_{n\alpha}(x)$  в узлах  $x_i$  можно подсчитать по формуле

$$S_{\alpha} = y - \alpha K_{\xi} H^{T} M. \tag{11.47}$$

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 (11.37) 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  $\alpha$  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  $\alpha$  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 [66], выбирают такое значение  $\alpha_0$ , для которого  $l_n(\alpha_0)=\beta$ , (11.48)

где  $\beta$  рекомендуется брать из промежутка [0,8;1].

Фактическое нахождение подходящего значения  $\alpha_0$  согласно способу невязки сводится к последовательному нахождению  $S_{n\alpha}(x)$  для различных  $\alpha$  и выбору из них такого  $S_{n\alpha_0}(x)$ , для которого наиболее точно выполняется равенство (11.48).

В частности, если все измерения равноточны, т.е.  $\sigma_i^2 = \sigma^2$ ,  $i=1,\dots,n$ , то среднеарифметическая

$$\widehat{\sigma}^2 = rac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2,$$

где  $f(x_i)$  — значения искомой функции в узлах, является состоятельной оценкой  $\sigma^2$  (если бы  $\sigma^2$  была неизвестной) и при достаточно большом n  $\widehat{\sigma}^2$  близка к  $\sigma^2$ . 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

 $\widehat{\sigma}^2 pprox rac{1}{n} \sum_{i=1}^n (S_{nlpha}(x_i) - y_i)^2.$ 

Приравняв  $\widehat{\sigma}^2$  известной  $\sigma^2$ , получаем равенство для определения подходящего значения  $\alpha$ :

$$\frac{1}{n} \sum_{i=1}^{n} (S_{n\alpha}(x_i) - y_i)^2 = \sigma^2$$
 (11.49)

(сравните с формулой (11.48)).

Существуют другие методы определения подходящего значения параметра сглаживания, в том числе для случая, когда дисперсия ошибок  $\sigma^2$  неизвестна [25].

Пусть  $\sigma^2$  неизвестна. Тогда для определения подходящего значения  $\alpha_0$  используют способ перекрестной значимости, согласно которому

$$\alpha_0 = \arg\min_{\alpha} M \|y - S_{\alpha}\|,$$

гае 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

$$u(\alpha) = \sum_{k=1}^{n} (y_k - S_k(x_k))^2, \tag{11.50}$$

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 (11.50). Численная реализация способа перекрестной значимости требует построения семейства сглаживающих кубических сплайнов  $\{S_k(x),\ k=1,\ldots,n\}$  для различных значений  $\alpha$ .

В заключение параграфа приведем схему восстановления функциональной зависимости с помощью робастных сглаживающих кубических сплайнов.

Вместо функционала (11.41) рассмотрим робастный сглаживающий функционал

$$J_{\text{po6}} = \sum_{i=1}^{n} \rho\left(\frac{y_i - \varphi(x_i)}{\sigma_i}\right) + \alpha \int_{0}^{b} (\varphi''(x))^2 dx, \qquad (11.51)$$

где  $\rho(s)$  — функция Хьюбера.

Следующая теорема доказывается аналогично теореме 11.2.

**Теорема 11.4.** Решением задачи на минимум функционала (11.51) в классе  $w_k$  является кубический сплайн  $S_n^{**}(x)$ , удовлетворяющий краевым условиям  $K^{\circ}$ .

В дальнейшем  $S_n^{**}(x)$  называется робастным сглаживающим кубическим сплайном и обозначается  $S_{\text{po6}}(x)$ . Обозначим  $S_{\text{po6}}=(S_{\text{po6}}(x_1),\dots,S_{\text{po6}}(x_n))^T, M=(M_1,\dots,M_n)^T, M_i=S_{\text{po6}}''(x_i)$ .

Можно показать, что вектор  $S_{\mathsf{pof}}$  связан с M соотношением

$$S_{\text{po6}} = y' - \alpha K_{\xi} H^T M, \qquad (11.52)$$

а вектор вторых производных находится из системы нелинейных уравнений

 $(A + \alpha H K_{\xi} H^T)M = Hy' + f, \qquad (11.53)$ 

где матрицы  $A,\ H,\ K_\xi$  и векторы  $y,\ f$  такие же, как и в (11.46), а вектор  $y'=(y_1',\dots,y_n')^T$  определяется равенствами

$$y_i' = \left\{egin{array}{ll} y_i, & ext{если } i \in I_0, \ S_{ ext{po6}}(x_i) + K\sigma_i, & ext{если } i \in I_+, \ S_{ ext{po6}}(x_i) - K\sigma_i, & ext{если } i \in I_-, \ I_0 = \{i: \ |y_i - S_{ ext{po6}}(x_i)| \leqslant K\sigma_i\}, \end{array}
ight.$$

$$I_{+} = \{i: y_{i} - S_{\text{po6}}(x_{i}) > K\sigma_{i}\}, \quad I_{-} = \{i: y_{i} - S_{\text{po6}}(x_{i}) < -K\sigma_{i}\}.$$

Для решения системы нелинейных уравнений (11.53) можно использовать нижеследующий итерационный процесс:

$$(A + \alpha H K_{\xi} H^{T}) M^{(l+1)} = H y^{(l)} + f, \qquad (11.54)$$

гле

$$\begin{split} y_i^{(l)} &= \left\{ \begin{array}{l} y_i, & \text{если } i \in I_0^{(l)}, \\ S_i^{(l)} + K\sigma_i, & \text{если } i \in I_+^{(l)}, \\ S_i^{(l)} - K\sigma_i, & \text{если } i \in I_-^{(l)}, \\ S_i^{(l)} - K\sigma_i, & \text{если } i \in I_-^{(l)}, \\ S_i^{(l)} &= (S_1^{(l)}, \dots, S_n^{(l)})^T = y^{(l-1)} - \alpha \, K_\xi H^T M^{(l)}, \\ I_0^{(l)} &= \{i: \ |y_i - S_i^{(l)}| \leqslant K\sigma_i\}, \\ I_+^{(l)} &= \{i: \ y_i - S_i^{(l)} > K\sigma_i\}, \quad I_-^{(l)} &= \{i: \ y_i - S_i^{(l)} < -K\sigma_i\}. \end{split}$$

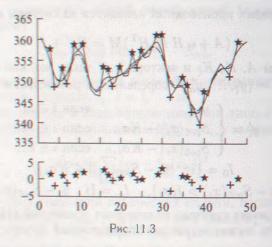
В качестве начального приближения  $S^{(0)}$  можно взять сглаживающий кубический сплайн, задаваемый равенствами (11.46), (11.47).

Итерационный процесс (11.54) сходится за конечное число

шагов к решению системы (11.52), (11.53).

Заметим, что итерационный процесс (11.54) обладает высокой вычислительной эффективностью, в частности из-за того, что на каждом шаге процесса решается система (11.54) 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 (\*, +) отмечены значения с  $i \in I_+$  и  $i \in I_-$ . На нижнем графике по-



казаны величины корректировок значений исходного временного ряда в результате применения итерационного процесса минимизации робастного сглаживающего функционала (11.51).

Робастные сглаживающие кубические сплайны были успешно использованы для решения ряда некорректных задач при обработке данных физических экспериментов [8, 9].

### 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 направлени-

ях усилиями многих ученых (Jean Morlet, Y. Meyer, Stephane Mallat, Ingrid Daubechies) [112, 116, 119, 121, 123, 126].

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:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{ikt},$$
(11.55)

где 
$$c_k = rac{1}{2\pi}\int\limits_{-\pi}^{\pi}f(t)e^{-ikt}\,dt.$$

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  $\int |f(t)|^2 \, dt < \infty$  и спра-

ведливо разложение в интеграл Фурье:

$$f(t) = \int_{-\infty}^{\infty} c(\omega)e^{i\omega t} d\omega, \quad c(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt.$$

В то же время, поскольку  $e^{ikt} 
eq 0$  при  $t 
ightarrow \infty$ ,  $e^{ikt}$  не может быть счетным базисом в пространстве  $L_2(-\infty,\infty)$ .

Поставим задачу о сохранении представления типа (11.55)

для  $f(t) \in L_2(-\infty, \infty)$ :

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \psi_k(t), \quad c_k = \frac{1}{C} \int_{-\infty}^{\infty} f(t) \widetilde{\psi}_k(t) dt, \quad (11.56)$$

 $\psi_k(t)$  образуют ортогональный базис пространства august in the same of the grant grown, in

 $L_2(-\infty,\infty)$ .

Оказывается, можно построить ортогональный 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  $\psi_{nk}(t) = 2^{n/2}\psi(2^nt - k)$ .

Для того чтобы  $\psi(t) \in L_2(-\infty,\infty)$  была базисным вейвлетом,

необходимо выполнение двух условий:

 $1^{\circ}$   $\psi(t)=0$  за пределами некоторого конечного промежутка либо  $\psi(t) o 0$  при  $t o \infty;$ 

$$2^{\circ} \int_{-\infty}^{\infty} \psi(t) dt = 0.$$

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).

$$\psi(t) = \begin{cases} 1, & 0 \le t \le 0.5, \\ -1, & 0.5 < t \le 1, \\ 0, & t \notin [0, 1]. \end{cases}$$
 (11.57)

Базисный вейвлет Хаара имеет конечный носитель  $t\in[0,1]$ . Функции  $\psi_{nk}(t)=2^{n/2}\psi(2^nt-k),\ n,k=0,\pm 1,\pm 2,\ldots$ , образуют ортонормированный базис пространства  $L_2(-\infty,\infty)$ .

Определим семейство функций

$$\widetilde{\psi}_{nk}(t) = \sum_{j=-\infty}^{\infty} 2^{n/2} \psi(2^n(t+j) - k),$$

$$n = 0, 1, 2, \dots, \quad k = 0, 1, \dots, 2^n - 1.$$
(11.58)

Можно показать, что  $\psi_{nk}(t)$  — периодические функции с периодом T=1. Их называют периодическими вейвлетами Хаара. Периодические вейвлеты (11.58) образуют ортогональный ба-

зис пространства  $L_2(0,1)$ .

Пример 11.2 (Французская шляпа — FHAT-wavelet).

$$\psi(t) = \begin{cases} -1/2, & t \in [-1, -1/3), \\ 1, & t \in [-1/3, 1/3], \\ 1/2, & t \in (1/3, 1], \\ 0, & t \notin [-1, 1]. \end{cases}$$
(11.59)

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

$$\psi(t;m) = (-1)^m \frac{d^m}{dt^m} e^{-t^2/2}, \qquad m = 1, 2, \dots$$
 (11.60)

Вейвлет  $\psi(t;1)$  называют волновым вейвлетом (wavewavelet), 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 $N,\,N=2,3,\dots$  Наиболее часто используется вейвлет Daub4 четвертого порядка. Для  $\psi_{\mathrm{Daub4}}$  справедливо равен-

ство 
$$\int\limits_{-\infty}^{\infty}t\psi_{\mathrm{Daub4}}(t)\,dt=0$$
 и его носитель  $t\in[0,3].$ 

Базисные вейвлеты  $\psi_{\mathrm{Daub}N}(t)$  могут быть получены из равенства

$$\psi_{\text{Daub}N}(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} (-1)^k C_{1-k}(N) \Phi(N, 2t - k), \qquad (11.61)$$

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  $\psi_{\mathrm{Daub4}}$  и определяется как предел

$$\Phi(N,t) = \lim_{n \to \infty} \Phi_n(N,t), \tag{11.62}$$

где  $\Phi_0(N,t)=\eta(x)-\eta(x-1),\ \eta(x)$  — функция Хевисайда,  $\Phi_n(N,t)=\sqrt{2}\,\sum^\infty_{}\,C_k(N)\Phi_{n-1}(N,2t-k).$ 

Все функции  $\Phi(N,t)$ , N=2,3,..., непрерывны и имеют носитель  $t\in[0,3]$ . Из равенства (11.61) следует, что базисные вейвлеты  $\psi_{\mathrm{Daub}N}$  также непрерывны и имеют носитель  $t\in[0,3]$ .

Аналогично (11.58) определим семейство функций

$$\widetilde{\psi}_{nk}(N,t) = \sum_{j=-\infty}^{\infty} 2^{n/2} \psi_{\text{Daub}N}(2^n(t+j)-k),$$

$$n = 0, 1, 2, \dots, k = 0, 1, \dots, 2^{n-1}.$$
(11.63)

Можно показать, что  $\psi_{nk}(N,t)$  — периодические функции с периодом T=1. Периодические вейвлеты Добеши (11.63) образуют ортонормированный базис пространства  $L_2(0,1)$ . Для любой  $f(t)\in L_2(0,1)$  справедливо разложение

$$f(t) = c_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n - 1} \alpha_{nk}(N) \widetilde{\psi}_{nk}(N, t).$$
 (11.64)

FRE

$$c_0 = \int_0^1 f(t) dt$$
,  $\alpha_{nk}(N) = \int_0^1 f(t) \widetilde{\psi}_{nk}(N, t) dt$ .

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 Кофманом (R. Coiman), 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 (11.61) через сканирующую функцию (разную для разных кофлетов) с конечным числом отличных от нуля коэффициентов  $C_i$ .

Пусть базисный вейвлет  $\psi(t)$  удовлетворяет условию нормировки  $|\psi(t)|^2\,dt=1.$  Покажем, что все вейвлеты семейства

$$\psi_{nk}(t) = 2^{n/2}\psi(2^nt - k), \quad n, k = 0, \pm 1, \dots,$$

также нормированы. Действительно.

$$\int_{-\infty}^{\infty} \psi_{nk}^{2}(t) dt = 2^{n} \int_{-\infty}^{\infty} \psi^{2}(2^{n}t - k) dt = \int_{-\infty}^{\infty} \psi^{2}(t') dt' = 1,$$

где  $t'=2^nt-k$ ,  $\mathrm{d}t=2^{-n}dt'$ . Таким образом, для всех вышеперечисленных семейств вейвлетов  $\psi_{nk}(t),\,n,k=0,\pm1,\ldots$ , выполнено

условие ортонормированности

$$\int_{-\infty}^{\infty} \psi_{nk}(t)\psi_{n'k'}(t) dt = \delta_{nn'}\delta_{kk'},$$

 $_{nn'}$  — 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 [119, 123].

#### 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, 42, 60, 73, 74, 88, 115]. 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.

## 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

для будущих временных дат

$$t_1, \dots, t_L$$

и задаются уровни погрешности этих оценок в виде числовых  $\sigma_1^2,\ldots,\sigma_L^2$ . значений дисперсий: (12.2)

Предполагается также, что в распоряжении исследователя имеется реализованная часть исследуемого временного ряда для прошедших временных дат  $t_{-N}, t_{-N+1}, \dots, t_{-1}$  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

$$w_k = 1/\sigma_k^2, \qquad k = -N, -N+1, \dots, 0, 1, \dots, L.$$
 (12.3)

Согласно результатам, представленным в § 11.1, детерминированная компонента временного ряда дается равенством

$$y_{\text{дет}}(t_k) = \sum_{j=1}^{m} \widehat{u}_j \, \varphi_j(t_k), \qquad k = -N, -N+1, \dots, 0, 1, 2, \dots, L,$$

$$\rho_j(t_k), \qquad k = -N, -N+1, \dots, 0, 1, 2, \dots, L,$$

$$\widehat{u}_j = \sum_{k=-N}^{L} w_k \varphi_j(t_k) y_k, \qquad (12.5)$$

 $y_{-N}, y_{-N+1}, \dots, y_0$  — значения реализованной части исследуемого временного ряда,  $y_1, \dots, y_L$  — экспертные оценки (12.1).

Окончательно прогнозные оценки  $y_{\mathsf{npor}}(t_k)$  значений временного ряда на будущие даты  $t_1, t_2, \dots, t_L$  даются равенствами

$$y_{\text{npor}}(t_k) = y_{\text{ger}}(t_k), \qquad k = 1, \dots, L,$$
 (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  $\sigma_k^2 o 0$ ,  $y_{\rm npor}(t_k) o y_k$ ,  $k=1,\ldots,L$  и информация в реализованной части временного ряда становится ненужной.

Замечание 12.2. Если экспертные оценки (12.1) отсутствуют, то по умолчанию будем полагать

$$y_1 = y_2, \dots, y_L = y_0. (12.7)$$

В этом случае достаточно надежный прогноз можно получить лишь для небольшого числа временных шагов вперед (в предположении выполнения равенства  $h_k = t_{k+1} - t_k = \mathrm{const}$ ).

Замечание 12.3. Если дисперсии  $\sigma_{-N}^2, \sigma_{-N+1}^2, \dots, \sigma_0^2$  априори неизвестны, то положим  $\sigma_{-N}^2 = \sigma_{-N+1}^2 = \dots = \sigma_0^2 = \sigma^2$ , где  $\sigma^2$  предварительно определяется из равенства

$$\sigma^2 = \frac{1}{N} \sum_{k=-N}^{0} (y_k - \hat{y}_k)^2, \tag{12.8}$$

где 
$$\widehat{y_k} = \sum_{j=1}^m u_j' \varphi_j'(t_k)$$
,  $k = -N, -N+1, \ldots, 0$ ,

$$u'_{j} = \sum_{k=-N}^{0} \varphi'_{j}(t_{k})y_{k},$$
 (12.9)

а  $\{\varphi_t(t)\}$  — 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 1, детерминированная компонента временного ряда дается равенством

$$y_{\text{Mer}}(t_k) = \sum_{j=1}^{m} u_{\text{po6}j} \varphi_j(t_k), \quad k = -N, -N+1, \dots, 0, 1, \dots, L,$$

$$u_{\text{po6}j} = \sum_{k=-N}^{L} w_k \varphi_j(t_k) \widetilde{y}_k, \qquad j = 1, \dots, m,$$
(12.10)

$$\widetilde{y}_k = \begin{cases} y_k, & \text{если } k \in I_0, \\ \sum\limits_{j=1}^m u_{\text{роб}j} \varphi_j(t_k) + K \sigma_k, & \text{если } k \in I_+, \\ \sum\limits_{j=1}^m u_{\text{роб}j} \varphi_j(t_k) - K \sigma_k, & \text{если } k \in I_-, \end{cases}$$
 
$$I_0 = \{k: \ |y_k - \sum\limits_{j=1}^m u_{\text{роб}j} \varphi_j(t_k)| \leqslant K \sigma_k \},$$
 
$$I_+ = \{k: \ y_k - \sum\limits_{j=1}^m u_{\text{роб}j} \varphi_j(t_k) > K \sigma_k \},$$
 
$$I_- = \{k: \ y_k - \sum\limits_{j=1}^m u_{\text{роб}j} \varphi_j(t_k) < -K \sigma_k \}.$$

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).

### 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

$$\sum_{k=-N}^{L} w_k (y_k - y(t_k))^2 + \alpha \sum_{k=-N}^{L-1} \frac{1}{h_k} (y(t_{k+1}) - y(t_k))^2, \quad (12.12)$$

где  $w_k$  определяются равенством (12.3),  $h_k = t_{k+1} - t_k$ ,  $\alpha > 0$  —

параметр сглаживания.

Тем самым, согласно § 2, вектор детерминированной компоненты  $Y_{\text{дет}} = (y_{\text{дет}}(t_{-N}), \dots, y_{\text{дет}}(t_{L}))^{T}$  определяется из системы линейных алгебраических уравнений

$$A Y_{\text{get}} = F, \tag{12.13}$$

$$A = \begin{pmatrix} w_{-N} + \frac{\alpha}{h_{-N+1}} & -\frac{\alpha}{h_{-N+1}} & 0 & 0 & \dots \\ -\frac{\alpha}{h_{-N+1}} & w_{-N+1} + \frac{\alpha}{h_{-N+1}} + \frac{\alpha}{h_{-N}} & -\frac{\alpha}{h_{-N+1}} & 0 & \dots \\ 0 & 0 & 0 & \dots & \dots \\ & \dots & 0 & 0 & \dots \\ & \dots & 0 & 0 & \dots \\ & \dots & 0 & 0 & \dots \\ & \dots & 0 & -\frac{\alpha}{h_{L-1}} & w_L + \frac{\alpha}{h_{L-1}} \end{pmatrix}$$

$$F = (y_{-N}w_{-N}, \dots, y_L w_L)^T.$$

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  $Y_{
m ner}$ 

Для предлагаемой в этом параграфе схемы остаются справедливыми все замечания предыдущего параграфа с заменой в замечании 12.5 параметров (m,N,L) на параметры  $(\alpha,N,L)$ .

Если в реализованной части  $y_{-N},\dots,y_0$  временного ряда присутствуют большие выбросы, то для выделения детерминированной компоненты временного ряда  $y_{-N}, y_{-N+1}, \dots, y_0, y_1, \dots, y_L$ вместо сглаживающих линейных сплайнов должны использо ваться робастные линейные сглаживающие сплайны.

Согласно § 11.3 в этом случае вместо сглаживающего функционала (12.12) используем робастный сглаживающий функцио-

$$\sum_{k=-N}^{L} \rho(\sqrt{w_k} (y_k - y(t_k)) + \alpha \sum_{k=-N}^{L-1} \frac{1}{h_k} (y(t_{k+1}) - y(t_k))^2, \quad (12.14)$$
The  $\rho(s)$  — vertical supplementary expresses the vertical supplementary  $\rho(s)$  and  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  are  $\rho(s)$  and  $\rho(s)$  are  $\rho(s)$  are

где  $\rho(s)$  — четная дифференцируемая функция,  $\rho(0)=0$ ,  $\rho(s)\sim 2$  для малых s,  $\rho(s)/s^2\to 0$  при  $s\to \infty$ . Если в качестве  $\rho(s)$  взять функцию Хьюбера, то детерминированная компонента  $Y_{\rm act}=(y_{\rm det}(t_{-N}),\dots,y_{\rm det}(t_{L}))^T$  временного ряда  $y_{-N},\dots,y_L$ , определяемая как вектор, минимизирующий робастный сглаживающий функционал (12.14), является решением системы нелинейных уравнений

$$A Y_{\text{дет}} = F', \tag{12.15}$$

гле

$$F' = (w_{-N}y'_{-N}, \ldots, w_L y'_L)^T, \quad y'_k = \left\{ egin{array}{ll} y_k, & ext{если } k \in I_0, \ y_{ ext{дет}} + K \sigma_k, & ext{если } k \in I_+, \ y_{ ext{дет}} - K \sigma_k, & ext{если } k \in I_-, \end{array} 
ight.$$

$$I_{0} = \{k: |y_{k} - y_{\text{MeT}}(t_{k})| \leq K\sigma_{k}\}, I_{+} = \{k: |y_{k} - y_{\text{MeT}}(t_{k})| > K\sigma_{k}\}, I_{-} = \{k: |y_{k} - y_{\text{MeT}}(t_{k})| < -K\sigma_{k}\}.$$

Для фактического нахождения  $Y_{\text{дет}}$  можно использовать итерационную процедуру, изложенную в § 11.3.

Если в качестве  $\rho(s)$  взять функцию Тьюки

$$\rho(s) = \left\{ \begin{array}{ll} s^2, & |s| \leqslant R, \\ R^2, & |s| > R, \end{array} \right.$$

то  $Y_{\text{дет}}$  как решение задачи на минимум функционала (12.14) является решением системы нелинейных уравнений (12.15), где

$$y_k' = \left\{ egin{array}{ll} y_k, & ext{если } k \in I_0, \\ y_{ ext{дет}}(t_k), & ext{если } k \in I_R, \end{array} 
ight.$$

$$I_0 = \{k: \ |y_k - y_{\text{det}}(t_k)| \leqslant R\sigma_k\}, \quad I_R = \{k: \ |y_k - y_{\text{det}}(t_k)| > R\sigma_k\}.$$

Для фактического нахождения  $Y_{\text{дет}}$  как решения нелинейной системы (12.16) можно использовать описанную в § 3 итерационную процедуру, сходящуюся к искомому решению за конечное число шагов.

Как и прежде, прогнозные значения для дат  $t_1, \dots, t_L$  определяются равенствами

$$y_{\mathrm{npor}}(t_k) = y_{\mathtt{Дет}}(t_k), \qquad k = 1, \dots, L.$$

## 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 [27, 109–111, 114, 115, 117, 130], хотя основы этого метода были заложены в более ранних работах [14, 44]. В настоящее время SSA успешно используется для выявления скрытых периодичностей и выделения детерминированных

где симметричная трехдиагональная матрица A и вектор F даются равенствами

$$A = \begin{pmatrix} w_{-N} + \frac{\alpha}{h_{-N+1}} & -\frac{\alpha}{h_{-N+1}} & 0 & 0 & \dots \\ -\frac{h_{-N+1}}{\alpha} & w_{-N+1} + \frac{\alpha}{h_{-N+1}} + \frac{\alpha}{h_{-N}} & -\frac{\alpha}{h_{-N+1}} & 0 & \dots \\ 0 & 0 & 0 & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots \\ \dots & 0 & 0 & \dots & \dots \\ \dots & 0 & 0 & \dots & \dots \\ \dots & 0 & -\frac{\alpha}{h_{L-1}} & w_L + \frac{\alpha}{h_{L-1}} \end{pmatrix}$$

$$F = (y_{-N}w_{-N}, \dots, y_L w_L)^T.$$

Поскольку А — 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  $Y_{\text{лет}}$ .

Для предлагаемой в этом параграфе схемы остаются справедливыми все замечания предыдущего параграфа с заменой в замечании 12.5 параметров (m,N,L) на параметры (lpha,N,L).

Если в реализованной части  $y_{-N},\dots,y_0$  временного ряда присутствуют большие выбросы, то для выделения детерминированной компоненты временного ряда  $y_{-N}, y_{-N+1}, \dots, y_0, y_1, \dots, y_L$ вместо сглаживающих линейных сплайнов должны использоваться робастные линейные сглаживающие сплайны.

Согласно § 11.3 в этом случае вместо сглаживающего функционала (12.12) используем робастный сглаживающий функцио-

нал

$$\sum_{k=-N}^{L} \rho(\sqrt{w_k}(y_k - y(t_k)) + \alpha \sum_{k=-N}^{L-1} \frac{1}{h_k} (y(t_{k+1}) - y(t_k))^2, \quad (12.14)$$

где ho(s) — четная дифференцируемая функция,  $ho(0)=0,\ 
ho(s)\sim$  $\sim s^2$  для малых  $|s|, \, 
ho(s)/s^2 
ightarrow 0$  при  $s 
ightarrow \infty$ .

Ести в качестве ho(s) взять функцию Хьюбера, то детерминированная компонента  $Y_{\mathtt{Aer}} = (y_{\mathtt{Aer}}(t_{-N}), \dots, y_{\mathtt{Aer}}(t_L))^T$  временного ряда  $y_{-N},\dots,y_L$ , определяемая как вектор, минимизирующий робастный сглаживающий функционал (12.14), является решением системы нелинейных уравнений

$$AY_{\text{MeT}} = F', \tag{12.15}$$

где

$$F' = (w_{-N}y'_{-N}, \ldots, w_Ly'_L)^T, \quad y'_k = \left\{egin{array}{ll} y_k, & ext{если } k \in I_0, \\ y_{ ext{дет}} + K\sigma_k, & ext{если } k \in I_+, \\ y_{ ext{дет}} - K\sigma_k, & ext{если } k \in I_-, \end{array}
ight.$$

$$\begin{array}{l} I_0 = \{k: \ |y_k - y_{\text{MET}}(t_k)| \leqslant K\sigma_k\}, \ I_+ = \{k: \ y_k - y_{\text{MET}}(t_k) > K\sigma_k\}, \\ I_- = \{k: \ y_k - y_{\text{MET}}(t_k) < -K\sigma_k\}. \end{array}$$

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 Удет как решение задачи на минимум функционала (12.14) является решением системы нелинейных уравнений (12.15), где

$$y_k' = \left\{ egin{array}{ll} y_k, & \mbox{если } k \in I_0, \\ y_{ ext{\tiny Дет}}(t_k), & \mbox{если } k \in I_R, \end{array} 
ight. \ I_0 = \left\{ k: \, |y_k - y_{ ext{\tiny Дет}}(t_k)| \leqslant R\sigma_k 
ight\}, \quad I_R = \left\{ k: \, |y_k - y_{ ext{\tiny Дет}}(t_k)| > R\sigma_k 
ight\}. \end{array}$$

$$I_0 = \{k : |y_k - y_{\text{net}}(t_k)| \le R\sigma_k\}, \quad I_R = \{k : |y_k - y_{\text{net}}(t_k)| > R\sigma_k\}.$$

Для фактического нахождения  $Y_{\text{дет}}$  как решения нелинейной системы (12.16) можно использовать описанную в § 3 итерационную процедуру, сходящуюся к искомому решению за конечное число шагов.

 $\mathsf{Kak}$  и прежде, прогнозные значения для дат  $t_1,\dots,t_L$  определяются равенствами

$$y_{\mathsf{npor}}(t_k) = y_{\mathsf{дer}}(t_k), \qquad k = 1, \dots, L.$$

#### 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 [27, 109-111, 114, 115, 117, 130], хотя основы этого метода были заложены в более ранних работах [14, 44]. 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 [27, 115].

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:

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_M \\ y_2 & y_3 & \dots & y_{M+1} \\ \dots & \dots & \dots & \dots \\ y_L & y_{L+1} & \dots & y_{L+M-1} \end{pmatrix}.$$
(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  $Y^TY$  называются сингулярными числами матрицы Y (что и дало название SSA).

Произведем ранжирование сингулярных чисел в порядке убывания:

$$\mu_1 \geqslant \mu_2 \geqslant \dots \geqslant \mu_M \geqslant 0,$$
 (12.20)

обозначив через  $\psi_j = (\psi_{j1}, \dots, \psi_{jM})^T$ ,  $j = 1, \dots, M$ , соответствующую им ортонормированную систему собственных векторов.

Сформируем ортогональную матрицу  $\Omega = (\psi_1, \dots, \psi_M)$  и определим  $(L \times M)$ -матрицу главных компонент V равенством

$$V = Y \Omega = (V_1, \dots, V_M),$$
 (12.21)

гае  $V_j$ ,  $j=1,\ldots,M$ , 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  $V_1,\ j=1,\ldots,M$ .

Числовые значения сингулярных чисел  $\mu_j$  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

$$Y' = \sum_{j=1}^{m} V_j \psi_j^T.$$
 (12.23)

Сглаженные детерминированные значения  $y_{\text{дет}}(t_k)$  исходного временного ряда получаются путем усреднения значений элементов на побочных диагоналях матрицы  $Y'=\{y'_{ij}\},$   $i=1,\ldots,L,\ j=1,\ldots,M\colon y_{\text{дет}}(t_1)=y'_{11},\ y_{\text{дет}}(t_2)=(y'_{12}+y'_{21})/2,$   $y_{\text{дет}}(t_3)=(y'_{13}+y'_{22}+y'_{31})/3,\ldots$ 

Одна из возможных схем прогнозирования с помощью SSA основана на рассмотрении системы линейных алгебраических

уравнений

где  $\psi_{ij}-i$ -я компонента вектора  $\psi_j,\ y_i$  — члены исходного или сглаженного рядов (12.17).

В системе (12.24) m неизвестных  $x_1, ..., x_m$  и M-1 уравне-

ние

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 коэффициентами:

$$y_{m+i} + a_1 y_{m+i-1} + \dots + a_m y_i = 0, \qquad i = 1, 2, \dots$$
 (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

и прогноз

$$y_{n+k+1} = \sum_{i=1}^{m} x_i^{(k)} \psi_{Mi}, \qquad k = 1, 2, \dots$$
 (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, ...

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:

$$AX = F - \varepsilon, \tag{12.29}$$

где  $F = (y_{n-M+2}, \dots, y_n)^T$  — вектор правых частей уравнений системы (12.24),  $\varepsilon$  — вектор «погрешностей» по отношению к неизвестному вектору  $F^*$ , для которого система  $AX = F^*$  совместна.

В условиях несовместности системы (12.24) определим вектор X как МНК-оценку вектора X линейной регрессионной модели (12.29).

Имеем

$$X^* = (A^T A)^{-1} A^T F. (12.30)$$

где  $X^* = (x_1^*, \dots, x_m^*)^T$ .

Прогнозируемое значение  $y_{n+1}$  определяется равенством

$$y_{n+1}^* = \sum_{i=1}^m x_i^* \psi_{Mi}. \tag{12.31}$$

Определим вектор  $Y^*=(y_{n-M+2}^*,\dots,y_n^*,y_{n+1}^*)^T$ , где  $y_{n+1}^*$  дается равенством (12.31),  $(y_{n-M+2}^*,\dots,y_n^*)^T=A\,X^*$ .

<sup>)</sup> Если система (12.24) совместна, то  $\varepsilon=0$ .

Вместо рекуррентных систем (12.27) рассмотрим рекуррентные системы

Все системы (12.32) совместны и каждая из них имеет единственное решение  $X^{(k)*}=(x_1^{(k)*},\dots,x_m^{(k)*})^T.$ 

Определим прогноз для последующих временных дат равен-

СТВОМ

$$y_{n+k+1} = \sum_{i=1}^{m} x_i^{(k)*} \psi_{Mi}, \qquad k = 1, 2, \dots$$
 (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  $\mu_j$ . Ширина окна M должна быть несколько больше числа m. M, 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

$$Y_i = u_0 + \sum_{j=1}^{m-1} u_j X_{ji}, (12.34)$$

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

$$y_{m+i} + a_1 y_{m+i-1} + \ldots + a_m y_i + \varepsilon_i = 0, \tag{12.3}$$

которая в обозначениях уравнения (12.34) запишется в виде

$$Y_i = u_0 + \sum_{j=1}^{m-1} u_j X_{ji} + \varepsilon_i, \qquad i = 1, 2, \dots$$
 (12.36)

Система равенств (12.36) полностью соответствует многофакторной линейной регрессионной модели (7.30) и тем самым между SSA и линейным регрессионным анализом существует очевидная взаимосвязь.

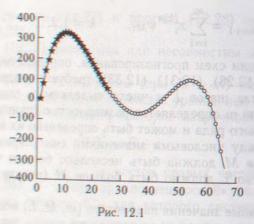


Рисунок 12.1 демонстрирует применение метода сингулярноспектрального анализа для прогнозирования ряда. Для примера использован полином 4-й степени (сплошная линия). Символом (\*) обозначены значения, использованные при обучении, а символом (О) — спрогнозированные значения.

Известно, что если временной ряд (12.17) является решением конечно-разностного уравнения (12.26), то SSA обеспечивает точный прогноз на любое число шагов вперед.

Если временной ряд (12.17) является реализацией линейной однородной регрессионной модели (12.35), где  $\varepsilon_i$  — хаотическая компонента, то SSA обеспечивает прогноз, точность которого определяется степенью вариации хаотической компоненты  $\varepsilon_i$ .

Рассмотрим теперь совокупность взаимосвязанных временных рядов

$$y_{j1}, \dots, y_{jn}, j = 1, \dots, p.$$
 (12.37)

В соответствии с многомерным сингулярно-спектральным анализом (MSSA) построим многомерную матрицу пошагового движения

где L = n - M + 1.

Аналогично одномерному SSA рассмотрим задачу на собственные значения  $(M\cdot p\times M\cdot p)$ -матрицы  $Y^TY$ :

$$Y^T Y \Psi = \mu \Psi, \tag{12.38}$$

множество ранжированных собственных значений задачи (12.38)

$$\mu_1 \geqslant \mu_2 \geqslant \dots \geqslant \mu_{M-p} \geqslant 0 \tag{12.39}$$

и соответствующую ортонормированную систему собственных векторов

 $\Psi_j = (\Psi_{1,j} \dots \Psi_{Mp,j})^T, j = 1, \dots, M \cdot p.$ 

Рассмотрим систему линейных алгебраических уравнений:

$$\sum_{i=1}^{m} x_i \cdot \Psi_{1,i} = y_{1,L+1};$$

$$\sum_{i=1}^{m} x_i \cdot \Psi_{M-1,i} = y_{1,n};$$

$$\sum_{i=1}^{m} x_i \cdot \Psi_{M+1,i} = y_{2,L+1};$$

$$\sum_{i=1}^{m} x_{i} \cdot \Psi_{2M-1,i} = y_{2,n};$$

$$\sum_{i=1}^{m} x_{i} \cdot \Psi_{(p-1)M+1,i} = y_{p,L+1};$$

$$\sum_{i=1}^{m} x_{i} \cdot \Psi_{pM-1,i} = y_{p,n},$$
(12.40)

где  $m \leq M$ .

В системе (12.40) m неизвестных  $\mathbf{x}=(x_1,\ldots,x_m)^T$  и  $(M-1)\cdot p$  уравнений. Систему (12.40) запишем в векторно-матричной форме

$$B\mathbf{x} = \mathbf{y},\tag{12.41}$$

где  $\mathbf{y} = (y_{1,L+1}, \dots, y_{1,n}, \dots, y_{p,L+1}, \dots, y_{p,n})^T$ . Система (12.41), как правило, несовместна. Определим вектор  $\mathbf{x}^*$  как МНК-оценку системы (12.41):

$$\mathbf{x}^{\bullet} = (B^T B)^{-1} \cdot B^T \cdot \mathbf{y}. \tag{12.42}$$

Прогнозируемые значения  $y_{1,n+1}^*,\dots,y_{p,n+1}^*$  определяются равенствами

$$y_{1,n+1}^* = \sum_{i=1}^m x_i^* \cdot \Psi_{M,i},$$

$$y_{p,n+1}^* = \sum_{i=1}^m x_i^* \cdot \Psi_{p,M,i}.$$
(12.43)

Если исследователя интересует прогноз только одного временного ряда, например  $y_{1i}$ ,  $i=1,\ldots,n$ , то подходящее значение можно взять из условия минимума квадратов отклонений  $\sum_{i=1}^{n} (y_{1,n-k}^i - y_{1,n-k})^2$ . Если же исследователя интересует прогноз всех временных рядов (12.37), то подходящее значение m можно

взять из условия минимума суммы квадратов отклонений для всех рядов  $\sum_{i=1}^p \sum_{k=1}^N (y_{j,n-k}^* - y_{j,n-k})^2$ .

Вернемся к случаю одного временного ряда  $y_1, \ldots, y_n$ .

Пусть наряду с исследуемым временным рядом известны реализации временных процессов  $f_{ji}, j=1,\ldots,q, i=1,\ldots,n$ , причем каждая реализация  $f_{ji}, j=1,\ldots,q$  оказывает аддитивное влияние на значение  $y_i$  временного ряда, а предыдущие реализации  $f_{i,i-k}, k=1,\ldots$ , влияния на реализацию  $y_i$  не оказывают.

Для того чтобы учесть влияние нестационарных факторов  $\{f_{ji}\}, j=1,\ldots,q$ , на исследуемый временной процесс, образуем

матрицу пошагового движения вида

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_M & f_{1,M} & \dots & f_{q,M} \\ y_2 & y_3 & \dots & y_{M+1} & f_{1,M+1} & \dots & f_{q,M+1} \\ \dots & \dots & \dots & \dots & \dots \\ y_L & y_{L+1} & \dots & y_n & f_{1,n} & \dots & f_{q,n} \end{pmatrix}$$
(12.44)

размерности  $(L \times (M+q))$ .

Рассмотрим задачу на собственные значения симметричной неотрицательной матрицы  $Y^TY$  размерности  $(M+q)\times (M+q)$ :

$$Y^T Y \Psi_j = \mu_j \Psi_j, \tag{12.45}$$

где  $\mu_1\geqslant \mu_2\geqslant \ldots\geqslant \mu_{M+q}\geqslant 0$  — ранжированная система собственных значений и  $\Psi_j=(\Psi_{1j}\dots\Psi_{M+q,j})^T, j=1,\dots,M+q,$  ортонормированная система собственных векторов матрицы  $Y^TY$ .

Обозначим через  $\Psi = (\Psi_1 \dots \Psi_{M+q}) - (M+q) imes (M+q)$ 

ортогональную матрицу системы собственных векторов. Введем L imes (M+q)-матрицу главных компонент:

$$V = (v_1, \dots, v_{M+q}) = Y \cdot \Psi.$$

Имеем

$$Y = V^T \cdot \Psi = \sum_{j=1}^{M+q} v_j \cdot \Psi_j.$$

Если выделить первые r < M + q главных компонент, то равенство

$$Y_{smoo} = \sum_{j=1}^{r} v_j \cdot \Psi_j \tag{12.46}$$

определяет значения сглаженного временного ряда. Чем меньше в равенстве (12.46)  $r \leqslant M+q$ , тем больше степень сглаживания исходного временного ряда.

Поскольку на побочных диагоналях матрицы наблюдений

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_M \\ y_2 & y_3 & \dots & y_{M+1} \\ \dots & \dots & \dots & \dots \\ y_L & y_{L+1} & \dots & y_n \end{pmatrix}$$

находятся одинаковые значения исходного временного ряда, рекомендуется значения сглаженного временного ряда  $y_{smooi}$  получать путем усреднения элементов побочных диагоналей минора  $Y_{smool}$  матрицы  $Y_{smoo}$ .

Предположим, что исходный временной ряд порожден конечно-разностным уравнением

$$y_i = \sum_{k=1}^{l} a_k \cdot y_{i-k} + \sum_{j=1}^{q} u_j \cdot f_{j,i}, \qquad (12.47)$$

где  $a_k, u_j, k = 1, \dots, l, j = 1, \dots, q$ , — постоянные, и уравнение (12.47) не может быть сведено к уравнению вида (12.47) с меньшими значениями l и q.

В этом случае будем говорить, что временной ряд  $y_i, i=1,...$ 

..., n, имеет ранг (l, q).

Из определения ранга нестационарного ряда следует, что временные ряды  $\{f_{ji}, i=1,\dots,n\}, j=1,\dots,q$  (12.47), линейно-независимы. Заметим, что непрерывным аналогом конечно-разностного уравнения является неоднородное нестационарное дифференциальное уравнение

$$\frac{d^{l}y}{dt^{l}} + \sum_{k=1}^{l} b_{k} \frac{d^{l-k}y}{dt^{l-k}} = \sum_{j=1}^{q} u_{j} \cdot f_{j}(t).$$

Определим матрицу пошагового движения (12.44) с M=l. Рассмотрим совокупность точек пространства  $E^{l+q}$ :

$$Y^{(1)} = (y_1, y_2, \dots, y_l, f_{1,l}, \dots, f_{q,l})^T;$$

$$Y^{(2)} = (y_2, y_3, \dots, y_{l+1}, f_{1,l+1}, \dots, f_{q,l+1})^T;$$

$$\dots$$

$$Y^{(L)} = (y_L, y_{L+1}, \dots, y_n, f_{1,n}, \dots, f_{q,n})^T;$$

$$Y^{(L+1)} = (y_{L+1}, y_{L+2}, \dots, y_{n+1}, f_{1,n+1}, \dots, f_{q,n+1})^T;$$

$$L > l + q.$$

Поскольку ортонормированная система собственных векторов задачи (12.45)  $\Psi_j, j=1,\ldots,l+q$ , образует базис пространства  $E^{l+q}$ , для элемента  $Y^{(L+1)}$  имеем представление

$$Y^{(L+1)} = \sum_{j=1}^{l+q} z_j \cdot \Psi_j, \tag{12.48}$$

или в покоординатном представлении:

$$y_{L+1} = \sum_{j=1}^{l+q} z_{j} \cdot \Psi_{1,j};$$

$$y_{L+2} = \sum_{j=1}^{l+q} z_{j} \cdot \Psi_{2,j};$$

$$y_{n} = \sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L,j};$$

$$f_{1,n+1} = \sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L+2,j};$$

$$f_{q,n+1} = \sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n+q-L,j};$$

$$y_{n+1} = \sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L+1,j}.$$

$$(12.49)$$

Рассмотрим подсистему системы (12.48) относительно неизвестных  $z_1, \ldots, z_{l+q}$ :

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,1} = y_{L+1};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,2} = y_{L+2};$$
...
$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,n-L} = y_{n};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,n-L+2} = f_{1,n+1};$$
...
$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,l+q} = f_{q,n+1}.$$
(12.50)

Правые части системы (12.50),  $y_{L+1},\dots,y_n,f_{1,n+1},\dots,f_{q,n+1}$ , известны. Число уравнений системы (12.50) равно l+q-1.

Очевидно, что если исходный ряд порожден конечноразностным уравнением (12.47) и тем самым ортонормированная система  $\Psi_j$ ,  $j=1,\ldots,l+q$ , образует базис пространства  $E^{l+q}$ , то система (12.50) совместна.

Наоборот, если система (12.50) для произвольного временного ряда несовместна, то ортонормированная система  $\Psi_j, j=1,\ldots$  1+q, не образует базис пространства  $E^{I+q}$  и рассматриваемый временной ряд не порожден конечно-разностным уравнением (12.47).

Утверждение 12.1. Если система (12.50) совместна и (l,q) — ранг временного ряда, то решение системы (12.50) единственно.

Действительно, если решение системы (12.50) единственно, то соответствующая системе (12.50) система однородных уравнений

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{1,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{2,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L+2,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{l+q,j} = 0$$
(12.51)

имеет только тривиальное решение  $z_j=0, j=1,\dots,l+q.$ 

Если же система (12.50) имеет несколько решений, то система (12.51) имеет нетривиальное решение  $\mathbf{z}^{(o)} \neq \mathbf{0}$ . Но последнее означает, что система векторов  $\Psi_j, j=1,\dots,l+q$ , линейно зависима. Полученное противоречие доказывает утверждение 12.1.

Пусть исследуемый нестационарный ряд порожден конечноразностным уравнением (12.47) и тем самым имеет ранг (l,q). Тогда система (12.50) совместна и имеет единственное решение  $\mathbf{z}^* = (z_1, \dots, z_{l+q})^T$ . Подставляя  $\mathbf{z}^*$  в последнее уравнение системы (12.49), полу-

чим прогнозируемое значение

$$y_{n+1} = \sum_{j=1}^{l+q} z_j^* \cdot \Psi_{n-L+1,j}.$$
 (12.52)

Замечание 12.6. Пусть временной ряд имеет ранг (l,q). Тогда можно определить матрицу пошагового движения для любого  $M \geqslant l$ , и все утверждения для систем (12.47)-(12.51) остаются в силе.

Если система (12.50) несовместна, то можно ввести понятие обобщенного прогноза.

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,1} = y_{L+1};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,2} = y_{L+2};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,n-L} = y_{n};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,n-L+2} = f_{1,n+1};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,n-L+2} = f_{1,n+1};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,l+q} = f_{q,n+1}.$$
(12.50)

Правые части системы (12.50),  $y_{L+1},\dots,y_n,f_{1,n+1},\dots,f_{q,n+1}$ , известны. Число уравнений системы (12.50) равно l+q-1.

Очевидно, что если исходный ряд порожден конечноразностным уравнением (12.47) и тем самым ортонормированная система  $\Psi_{j} = 1, \ldots, l+q$ , образует базис пространства  $E^{l+q}$ , то система (12.50) совместна.

Наоборот, если система (12.50) для произвольного временного ряда несовместна, то ортонормированная система  $\Psi_j, j=1,\ldots,l+q$ , не образует базис пространства  $E^{l+q}$  и рассматриваемый временной ряд не порожден конечно-разностным уравнением (12.47).

Утверждение 12.1. Если система (12.50) совместна и (l,q) — ранг временного ряда, то решение системы (12.50) единственно.

Действительно, если решение системы (12.50) единственно, то соответствующая системе (12.50) система однородных уравнений

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{1,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{2,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L+2,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{l+q,j} = 0$$
(12.51)

имеет только тривиальное решение  $z_{j} = 0, j = 1, ..., l + q$ .

Если же система (12.50) имеет несколько решений, то система (12.51) имеет нетривиальное решение  $\mathbf{z}^{(a)} \neq \mathbf{0}$ . Но последнее означает, что система векторов  $\Psi_j, j=1,\ldots,l+q$ , линейно зависима. Полученное противоречие доказывает утверждение 12.1.

Пусть исследуемый нестационарный ряд порожден конечноразностным уравнением (12.47) и тем самым имеет ранг (l,q). Тогда система (12.50) совместна и имеет единственное решение

 $\mathbf{z}^* = (z_1, \dots, z_{l+q})^T$ . Подставляя  $\mathbf{z}^*$  в последнее уравнение системы (12.49), получим прогнозируемое значение

$$y_{n+1} = \sum_{j=1}^{l+q} z_j^* \cdot \Psi_{n-L+1,j}.$$
 (12.52)

Замечание 12.6. Пусть временной ряд имеет ранг (l,q). Тогда можно определить матрицу пошагового движения для любого  $M \geqslant l$ , и все утверждения для систем (12.47)–(12.51) остаются в силе.

Если система (12.50) несовместна, то можно ввести понятие обобщенного прогноза.

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,1} = y_{L+1};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,2} = y_{L+2};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,n-L} = y_{n};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,n-L+2} = f_{1,n+1};$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{j,l+q} = f_{q,n+1}.$$
(12.50)

Правые части системы (12.50),  $y_{L+1},\dots,y_n,f_{1,n+1},\dots,f_{q,n+1}$ , известны. Число уравнений системы (12.50) равно l+q-1.

Очевидно, что если исходный ряд порожден конечноразностным уравнением (12.47) и тем самым ортонормированная система  $\Psi_{j}$ ,  $j=1,\ldots,l+q$ , образует базис пространства  $E^{l+q}$ , то система (12.50) совместна.

Наоборот, если система (12.50) для произвольного временного ряда несовместна, то ортонормированная система  $\Psi_j, j=1,\ldots,1+q$ , не образует базис пространства  $E^{l+q}$  и рассматриваемый временной ряд не порожден конечно-разностным уравнением (12.47).

Утверждение 12.1. Если система (12.50) совместна и (l,q) – ранг временного ряда, то решение системы (12.50) единственно. Действительно, если решение системы (12.50) единственно. то соответствующая системе (12.50) система однородных уравнений

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{1,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{2,j} = 0;$$

$$\vdots$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L,j} = 0;$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{n-L+2,j} = 0;$$

$$\vdots$$

$$\vdots$$

$$\sum_{j=1}^{l+q} z_{j} \cdot \Psi_{l+q,j} = 0$$

$$\vdots$$

**имеет** только тривиальное решение  $z_j = 0, j = 1, \dots, l + q$ .

Если же система (12.50) имеет несколько решений, то система (12.51) имеет нетривиальное решение  $\mathbf{z}^{(o)} \neq \mathbf{0}$ . Но последнее означает, что система векторов  $\Psi_j, j=1,\dots,l+q$ , линейно зависима. Полученное противоречие доказывает утверждение 12.1.

Пусть исследуемый нестационарный ряд порожден конечноразностным уравнением (12.47) и тем самым имеет ранг (l,q). Тогда система (12.50) совместна и имеет единственное решение  $\mathbf{z}^* = (z_1, \dots, z_{l+q})^T$ . Подставляя  $\mathbf{z}^*$  в последнее уравнение системы (12.49), полу-

чим прогнозируемое значение

$$y_{n+1} = \sum_{j=1}^{l+q} z_j^* \cdot \Psi_{n-L+1,j}.$$
 (12.52)

**Замечание 12.6.** Пусть временной ряд имеет ранг (l,q). Тогда можно определить матрицу пошагового движения для любого  $M \geqslant l$ , и все утверждения для систем (12.47)-(12.51) остаются в силе.

Если система (12.50) несовместна, то можно ввести понятие обобщенного прогноза.

Запишем систему (12.50) в матрично-векторном виде

$$Az = F, (12.53)$$

$$\text{rime } A = \begin{pmatrix} \Psi_{1,1} & \Psi_{1,2} & \dots & \Psi_{1,l+q} \\ \vdots & & & & \ddots \\ \Psi_{n-L,1} & \Psi_{nL,2} & \dots & \Psi_{n-L,l+q} \\ \Psi_{n-L+2,1} & \Psi_{n-L+2,2} & \dots & \Psi_{n-L+2,l+q} \\ \vdots & & & & \ddots \\ \Psi_{l+q,1} & \Psi_{l+q,2} & \dots & \Psi_{l+q,l+q} \end{pmatrix}$$

—  $[(l+q-1)\times(l+q)]$ -матрица,  $\mathbf{F}=(y_{L+1},\ldots,y_n,f_{1,n+1},\ldots,f_{q,n+1})^T$  — вектор размерности l+q.

Определим МНК-решение системы (12.53) равенством

$$\mathbf{z}_{MHK} = (A^T A)^{-1} A^T \mathbf{F}.$$
 (12.54)

Замечание 12.7. Если симметричная неотрицательная матрица  $A^TA$  плохо обусловлена, то вместо МНК-решения (12.54) возьмем регуляризированное решение

$$\mathbf{z}_{\text{PMHK}} = (A^T A + \alpha \cdot \operatorname{diag}(A^T A))^{-1} A^T \mathbf{F}, \tag{12.55}$$

где  $\alpha>0$  — параметр регуляризации,  $\mathrm{diag}(A^TA)$  — диагональная матрица, на главной диагонали которой стоят положительные числа  $(A^TA)_{l,1},\ldots,(A^TA)_{l+q,l+q}$ .

Обобщенный прогноз  $y_{n+1}$  определяется равенством

$$y_{n+1} = \sum_{j=1}^{l+q} z_{\text{MHK}j} \cdot \Psi_{n-L+1,j}, \qquad (12.56)$$

если используется МНК-решение, либо равенством

$$y_{n+1} = \sum_{j=1}^{l+q} z_{\text{PMHK}j} \cdot \Psi_{n-L+1,j}, \qquad (12.57)$$

если используется РМНК-решение.

Пусть теперь исследуемый ряд содержит хаотическую компоненту, например, порожден хаотическим конечно-разностным уравнением

$$y_i = \sum_{k=1}^{l} a_k \cdot y_{i-k} + \sum_{j=1}^{q} u_j \cdot f_{j,i+1} + \varepsilon_i, \qquad (12.58)$$

где  $\varepsilon_i$  — хаотическая компонента.

Для прогнозирования временного ряда при наличии хаотической компоненты можно использовать две схемы.

Согласно первой схеме исходный временной ряд исследуется и прогнозируется на основе равенств (12.50), (12.52) (или (12.54), (12.55)).

Согласно второй схеме производится предварительное сглаживание исходного временного ряда с выделением детерминированной компоненты  $y_{\text{det}i}$ ,  $i=1,\ldots,n$ , а затем используется схема прогнозирования (12.50), (12.52) (или (12.54), (12.55)), в которой роль исходного временного ряда  $y_i$ ,  $i=1,\ldots,n$ , играет детерминированная компонента  $y_{\text{det}i}$ ,  $i=1,\ldots,n$ .

Сглаживание исходного нестационарного временного ряда можно производить в рамках самого NSSA на основе равенства (12.46). Заметим, что точность прогнозирования при наличии хаотической компоненты определяется уровнем ее вариации.

# интерполяция функций одной и многих переменных с помощью схем, основанных на метрическом анализе

Задачи интерполяции и экстраполяции значений функций являются одними из основных задач математики, в том числе для решения различных прикладных задач, и прежде всего задач

прогнозирования.

Для функций одной переменной задачи интерполирования ставились и решались со времен Ньютона и Лагранжа, и к настоящему времени здесь получены достаточно завершенные результаты по разработке различных методов интерполяции и выявлению свойств интерполяционных значений, включая проблемы погрешности интерполяций и сходимости интерполяционных значений к точным значениям [12, 28].

Напомним, что классическая схема интерполяции функции одной переменной г основана на представлении исходной функ-

ции у(х) в виде линейной комбинации:

$$y(x) \approx L_n(x) = \sum_{j=0}^{N} c_{(j)} \cdot \varphi_j(x),$$

где  $\varphi_j(x), j=0,\ldots,m$ . — система заданных базисных функций,  $c_j$  — искомые параметры.

Необходимо определить  $c_j$  так, чтобы функция y(x) в заданных точках  $x_i$ ,  $i=1,\ldots,n$ , принимала известные значения  $Y_i$ .

Так, в системе Лагранжа в качестве базисной системы берут-

ся мономы  $\varphi_j(x) = x^j$ .

Однако выяснилось, что интерполяция Лагранжа обеспечивает равномерную сходимость интерполяционных полиномов к исследуемой функции лишь для определенного класса гладких функций, например целых функций.

Фабер показал, что для любого варианта сгущения точек  $x_i, i=1,\ldots,n$ , можно найти такую непрерывную функцию y(x), принимающую значения  $Y_i$  в точках  $x_i, i=1,\ldots,n$ , что

при  $n o\infty$  последовательность  $L_n(x)$  не стремится к y(x) рав-

номерно.

Более того, С. Н. Бернштейн показал, что последовательность интерполяционных полиномов Лагранжа  $L_n(x)$  для функции y=|x|, построенной на равномерной сетке  $i=-1+ih,\ i=0,\dots$ 

 $\dots, n, h = \frac{2}{n!}$ , расходится в любой точке  $x \neq -1, 0, 1$ . Причиной расходимости является наличие у рассматриваемой

Причиной расходимости является наличие у рассматриваемой функции угловых точек, в которых первая производная терпит разрыв. Пример Бернштейна показывает, что даже наличие одной угловой точки может привести к расходимости последовательности интерполяционных полиномов на всем рассматриваемом промежутке.

В 60-70-е годы как альтернатива интерполяционной схеме Лагранжа была предложена схема сплайн-интерполяций, которая позволяла локализовать влияние угловых точек и обеспечивала равномерную сходимость интерполяционных сплайнприближений для любой непрерывной функции [19, 37, 45, 66].

Подчеркнем, что интерполяционные функции Лагранжа мож-

но представить в виде

$$y(x) \approx L_n(x) = \sum_{k=1}^n z_k(x) \cdot Y_k,$$

где коэффициенты г удовлетворяют условию нормировки

$$\sum_{k=1}^{n} z_k(x) = 1$$

для любого x из промежутка интерполяции и, кроме того,  $z_k(x_i) = \delta_{ki}, \ k, i = 1, \dots, n, \ \delta_{ki}$  — символ Кронекера.

Схемы представления функций в виде линейных комбинаций базисных функций, в том числе полиномов и сплайн-аппроксимаций, в принципе, могут быть обобщены на функции многих переменных, но практически такие схемы являются работоспособными только для функций двух, максимум трех-четырех переменных. Для функций большого числа переменных эффективных общих схем интерполяции и экстраполяции до сих пор нет. Имеются лишь приближенные схемы интерполяции типа кусочно-линейных, которые, с одной стороны, требуют для своей реализации большого числа данных, с другой стороны, даже при большом числе данных не обеспечивают требуемой точности. Примером таких схем являются нейронные сети, с помощью которых

производится интерполирование и экстраполирование функций многих переменных [34].

В настоящее время для решения многих прикладных задач (прежде всего с помощью компьютерной техники) имеется острая необходимость в разработке универсального метода интерполяции и экстраполяции функций многих переменных y=f(x), в том числе в условиях наличия хаотических погрешностей в известных значениях функции для точек  $X_1,\ldots,X_n\in E^m$ , по которым восстанавливаются значения функции в других точках пространства  $E^m$ .

В гл. 13–15 представлен универсальный подход, позволяющий с помощью компьютерной техники эффективно решать задачи интерполирования, восстановления и прогнозирования функций одной и многих переменных без фиксации априори вида функциональной зависимости от переменных, а используя лишь информацию, имеющуюся в реализованных значениях функции  $Y_1, \ldots, Y_n$  в точках  $X_1, \ldots, X_n$ , в том числе и в условиях наличия хаотической погрешности в значениях  $Y_1, \ldots, Y_n$  [51, 120].

В главе рассматривается интерполяция функций одной и многих переменных с помощью схем, основанных на метрическом анализе. Предлагаются новые схемы и алгоритмы интерполяции функций одной и многих переменных на основе метрического анализа [53]. Показано, что предложенные схемы с высокой степенью точности интерполируют функции многих переменных даже при сравнительно небольшом числе точек, в которых значения функции известны.

# 1. Интерполяция функций одной и многих переменных с помощью схем, основанных на метрическом анализе

Рассмотрим задачу, связанную с функциональной зависимостью:

$$Y = F(X_1, ..., X_m) = F(X),$$
 (13.1)

где функция  $F(\mathbf{X})$  неизвестна, и подлежат восстановлению значения функции либо в одной точке  $\mathbf{X}$ , либо в совокупности заданных точек, при наличии информации о значениях этой функции в точках  $\mathbf{X}_1,\dots,\mathbf{X}_n$  (узлы интерполяции):

$$Y_k = F(\mathbf{X_k}), \qquad \mathbf{X_k} = (X_{k1}, \dots, X_{km})^T \in E^m.$$

В пространстве  $E^m$  зададим метрику, порожденную нормой:

$$\|\mathbf{X}\| = \sum_{j=1}^{m} \omega_j X_j^2,$$
 (13.2)

где  $\omega_j$  — метрические веса, которые удовлетворяют следующим соотношениям:

$$\omega_j\geqslant 0, \qquad \sum_{j=1}^m \omega_j=m.$$

Если задать метрические веса  $\omega_k=1,\ k=1,\dots,m$ , то получим метрику, порожденную евклидовой нормой. В общем случае метрические веса  $\omega_1,\dots,\omega_m$  учитывают характер изменения исследуемой функции при изменении ее аргументов и вычисляются с учетом взаимного расположения точек  $\mathbf{X_1},\dots,\mathbf{X_n}$  и значений функции  $Y_1,\dots,Y_n$  в этих точках. Одна из возможных схем такого задания метрических весов будет приведена ниже, пока же предположим что они заданы как функции от  $\mathbf{X}^*$ .

Рассмотрим задачу восстановления значения функции в заданной точке  $\mathbf{X}^*$ . Начнем со случая 1  $n\leqslant m$ , т.е. число точек, в которых заданы значения функции, меньше, чем число аргументов.

Составим  $(n \times n)$ -матрицу метрической неопределенности W для точки  $\mathbf{X}^*$  относительно совокупности точек  $\mathbf{X}_1,\dots,\mathbf{X}_n$ :

$$W = \begin{pmatrix} \rho^2(\mathbf{X}_1, \mathbf{X}^*)_{\boldsymbol{\omega}} & (\mathbf{X}_1, \mathbf{X}_2)_{\boldsymbol{\omega}} & \dots & (\mathbf{X}_1, \mathbf{X}_n)_{\boldsymbol{\omega}} \\ (\mathbf{X}_2, \mathbf{X}_1)_{\boldsymbol{\omega}} & \rho^2(\mathbf{X}_2, \mathbf{X}^*)_{\boldsymbol{\omega}} & \dots & (\mathbf{X}_2, \mathbf{X}_n)_{\boldsymbol{\omega}} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{X}_n, \mathbf{X}_1)_{\boldsymbol{\omega}} & (\mathbf{X}_n, \mathbf{X}_2)_{\boldsymbol{\omega}} & \dots & \rho^2(\mathbf{X}_1, \mathbf{X}^*)_{\boldsymbol{\omega}} \end{pmatrix}, (13.3)$$

гле

$$\rho^{2}(\mathbf{X}_{i}, \mathbf{X}^{*})_{\omega} = \sum_{k=1}^{m} \omega_{k} \cdot (X_{ik} - X_{k}^{*})^{2}.$$

$$(\mathbf{X}_{i}, \mathbf{X}_{j})_{\omega} = \sum_{k=1}^{m} \omega_{k} \cdot (X_{ik} - X_{k}^{*}) \cdot (X_{jk} - X_{k}^{*}), \quad i, j = 1, \dots, n.$$
(13.4)

Матрица метрической неопределенности W определяется лишь местоположением точек  $\mathbf{X}_i,\dots,\mathbf{X}_n,\mathbf{X}^*$  и метрическими весами  $\omega_1,\dots,\omega_m$ .

Из (13.4) следует, что матрицу метрической неопределенности W можно представить в следующем виде:

$$W = X \operatorname{diag}(\omega_1, ..., \omega_m) X^T = Y Y^T,$$
  

$$Y = X \operatorname{diag}(\sqrt{\omega_1}, ..., \sqrt{\omega_m}),$$
(13.5)

LIG

$$X = \begin{pmatrix} X_{11} - X_1^* & X_{12} - X_2^* & \dots & X_{1m} - X_m^* \\ X_{21} - X_1^* & X_{22} - X_2^* & \dots & X_{2m} - X_m^* \\ \dots & \dots & \dots & \dots \\ X_{n1} - X_1^* & X_{n2} - X_2^* & \dots & X_{nm} - X_m^* \end{pmatrix}, \quad (13.6)$$

 $\operatorname{diag}(\omega_1, ..., \omega_m)$  — диагональная  $(m \times m)$ -матрица, на главной диагонали которой стоят метрические веса.

Из представления (13.5) и вида матрицы (13.6) следует, что при n > m матрица метрической неопределенности W является заведомо вырожденной, поэтому этот случай будет рассмотрен отдельно.

В рассматриваемом случае  $n\leqslant m$  предполагается, что матрица W не является вырожденной. Интерполяционная формула для значения  $Y^*$  функции  $Y=F(\mathbf{X})$  в точке  $\mathbf{X}^*$  имеет следующий вид:

$$Y^* = \sum_{i=1}^n z_i Y_i = (\mathbf{z}, \mathbf{Y}), \tag{13.7}$$

где  $\mathbf{z}=(z_1,\ldots,z_n)^T,\,\mathbf{Y}=(Y_1,\ldots,Y_n)^T,\,$ а интерполяционные веса  $z_1,\ldots,z_n$  удовлетворяют, как это имеет место во многих интерполяционных формулах, условию нормировки:  $\sum_{i=1}^n z_i=1$ . Заметим, что интерполяционные веса  $z_i$  могут быть и отрицательными.

Определим числовую характеристику  $\sigma_{ND}^2$  метрической неопределенности восстановленного значения  $Y^*$  в точке  $X^*$  по значениям функции в точках  $\mathbf{X}_1,\dots,\mathbf{X}_n$  с заданными интерполяционными весами  $\mathbf{z}=(z_1,\dots,z_n)^T$  с помощью равенства

$$\sigma_{ND}^2 = W(\mathbf{X}_1, \dots, \mathbf{X}_n, \mathbf{X}^*) \mathbf{z}, \mathbf{z}. \tag{13.8}$$

Ставится задача подбора таких значений весов  $z_i$ ,  $i=1,\ldots,n$ , удовлетворяющих условию нормировки  $\sum_{i=1}^{n} z_i = 1$ , чтобы числовое значение характеристики метрической неопределенности (13.8) было минимальным:

$$\begin{cases} (W\mathbf{z}, \mathbf{z}) - \min \mathbf{z}, \\ (\mathbf{z}, \mathbf{1}) = 1, \mathbf{1} = (1, \dots, 1)^T. \end{cases}$$
(13.9)

Задача (13.9) решается методом Лагранжа:

$$L(\mathbf{z}, u) = \frac{1}{2}(W\mathbf{z}, \mathbf{z}) - u(\mathbf{z}, 1) - 1, \quad \nabla_{\mathbf{z}}L = W\mathbf{z} - u\mathbf{1} = 0,$$
$$\mathbf{z} = uW^{-1}\mathbf{1}, \quad (\mathbf{z}, \mathbf{1}) = u(W^{-1}\mathbf{1}, \mathbf{1}) = 1, \quad u = \frac{1}{(W^{-1}\mathbf{1}, \mathbf{1})}.$$

Отсюда искомый вектор  $\mathbf{z}^*$  и искомое интерполяционное значение  $Y^*$  даются равенствами

$$\mathbf{z}^* = \frac{W^{-1}\mathbf{1}}{(W^{-1}\mathbf{1},\mathbf{1})}, \quad Y^* = \frac{W^{-1}\mathbf{1},\mathbf{Y}}{(W^{-1}\mathbf{1},\mathbf{1})}.$$
 (13.10)

Из равенства (13.5) следует, что матрица W неотрицательна, а с учетом невырожденности матрицы W получаем, что она положительна:  $\forall \mathfrak{d} \in E^n, \, \mathfrak{d} \neq 0 \Rightarrow (W\mathfrak{d}, \mathfrak{d}) > 0$ .

**Теорема 13.1 (о сходимости).** Интерполяционное значение функции  $Y(\mathbf{X}^*)$  сходится к точному значению  $Y_k$ ,  $k=1,\ldots,n$ , когда  $\mathbf{X}^* \to \mathbf{X}_k$ .

Доказательство. Положим  $\mathbf{X}^* = \mathbf{X}^1 - \boldsymbol{\varepsilon}$  (знак минус взят для удобства) и вектор  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$  устремим к нулю. В этом случае матрица X, определенная по формуле (13.6), имеет вид

$$X = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_m \\ X_{21} + \varepsilon_1 & X_{22} + \varepsilon_2 & \dots & X_{2m} + \varepsilon_m \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} + \varepsilon_1 & X_{n2} + \varepsilon_2 & \dots & X_{nm} + \varepsilon_m \end{pmatrix}.$$

Здесь  $\mathbf{X}_{ij} = X_{ij} - X_{1j}, \ i=2,\ldots,n, \ j=1,\ldots,m.$  Обозначим

$$\rho_{\omega} = \sqrt{\sum_{k=1}^{m} \omega_k \varepsilon_k^2}.$$

Тогда матрица метрической неопределенности будет иметь вид

$$W = \begin{pmatrix} \rho_{\omega}^{2} & \sum_{k=1}^{m} \omega_{k} \varepsilon_{k} \overline{X}_{2k} + \rho_{\omega}^{2} \\ \sum_{k=1}^{m} \omega_{k} \varepsilon_{k} \overline{X}_{2k} + \rho_{\omega}^{2} & \sum_{k=1}^{m} \omega_{k} (\overline{X}_{2k} + \varepsilon_{k})^{2} \\ \sum_{k=1}^{m} \omega_{k} \varepsilon_{k} \overline{X}_{nk} + \rho_{\omega}^{2} & \sum_{k=1}^{m} \omega_{k} (\overline{X}_{nk} + \varepsilon_{k}) (\overline{X}_{2k} + \varepsilon_{k}) & \dots \\ \sum_{k=1}^{m} \omega_{k} \varepsilon_{k} \overline{X}_{nk} + \rho_{\omega}^{2} & \dots & \sum_{k=1}^{m} \omega_{k} (\overline{X}_{2k} + \varepsilon_{k}) (\overline{X}_{nk} + \varepsilon_{k}) \\ \sum_{k=1}^{m} \omega_{k} (\overline{X}_{nk} + \varepsilon_{k})^{2} \end{pmatrix}$$
(13.11)

Для обратной матрицы имеем

$$W^{-1} = \frac{1}{\det(W)} A^{T} = \frac{1}{\det(W)} A. \tag{13.12}$$

Последнее равенство учитывает симметричность матрицы W, а элементы матрицы A представляют собой соответствующие алгебраические дополнения элементов матрицы W:

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix}.$$

Для матрицы метрической неопределенности (13.11) получаем соответствующие алгебраические дополнения как функции от є:

$$A_{11}(\mathbf{\epsilon}) \xrightarrow{\boldsymbol{\epsilon} \to \mathbf{0}} \begin{vmatrix} \sum_{k=1}^{m} \omega_k \widetilde{X}_{2k}^2 & \dots & \sum_{k=1}^{m} \omega_k \widetilde{X}_{2k} \widetilde{X}_{nk} \\ \sum_{k=1}^{m} \omega_k \widetilde{X}_{nk} \widetilde{X}_{2k} & \dots & \sum_{k=1}^{m} \omega_k \widetilde{X}_{nk}^2 \end{vmatrix}$$

$$A_{1j}(\mathbf{\epsilon}) = \rho_{\omega} C_{1j}(\mathbf{\epsilon}), \quad j = 2, \dots, n,$$

$$A_{i1}(\mathbf{\epsilon}) = \rho_{\omega} C_{i1}(\mathbf{\epsilon}), \quad i = 2, \dots, n,$$

$$A_{ij}(\mathbf{\epsilon}) = \rho_{\omega}^2 C_{ij}(\mathbf{\epsilon}), \quad i, j = 2, \dots, n,$$

гле  $C_{ij}(\epsilon)$  — ограниченные функции при  $\epsilon \to 0$ . Для интерполяционных весов имеем

$$z = \frac{W^{-1}1}{W^{-1}1, 1} = \frac{A1}{A1, 1},$$

$$z_1(\varepsilon) = \frac{C_{11}(\varepsilon) + \rho_{\omega} \sum_{k=2}^{n} C_{1k}(\varepsilon)}{C_{11}(\varepsilon) + 2\rho_{\omega} \sum_{k=2}^{n} C_{1k}(\varepsilon) + \rho_{\omega}^2 \sum_{i,j=2}^{n} C_{ij}(\varepsilon)} = 1 + \rho_{\omega} \widetilde{C}_1(\varepsilon),$$

$$z_1(\varepsilon) = \frac{\rho_{\omega} C_{ii}(\varepsilon) + \rho_{\omega}^2 \sum_{k=2}^{n} C_{ik}(\varepsilon)}{C_{11}(\varepsilon) + 2\rho_{\omega} \sum_{k=2}^{n} C_{1k}(\varepsilon) + \rho_{\omega}^2 \sum_{i,j=2}^{n} C_{ij}(\varepsilon)} = \rho_{\omega} \widetilde{C}_1(\varepsilon), \quad i = 2, ..., n$$

 $\widetilde{C}_1(f \epsilon),\dots,\widetilde{C}_n(f \epsilon)$  — ограниченные функции при  $f \epsilon o 0$ . Для интерполяционного значения в точке  ${f X}^*$  получаем

$$Y(\mathbf{X}_1 + \boldsymbol{\varepsilon}) = \sum_{k=2}^{n} z_i(\boldsymbol{\varepsilon}) Y_i = Y_1 + \rho_{\omega} \sum_{k=2}^{n} \widetilde{C}_1(\boldsymbol{\varepsilon}) Y_i = Y_1 + \rho_{\omega} \widetilde{C}(\boldsymbol{\varepsilon}),$$

 $\widetilde{C}(arepsilon)$  — ограниченная функция при arepsilon o 0;

$$ho_{\omega} = \sqrt{\sum_{k=1}^{m} \omega_{k} arepsilon_{k}^{2}} \leqslant \parallel \mathbf{\omega} \parallel \parallel \mathbf{\epsilon}^{2} \parallel = \parallel \mathbf{\omega} \parallel 
ho \quad$$
 при  $\mathbf{\epsilon} 
ightarrow \mathbf{0}$ ,

где  $\omega = (\omega_1, \dots, \omega_n)^T$ ,  $\varepsilon^2 = (\varepsilon_1^2, \dots, \varepsilon_n^2)^T$ ,  $\rho = \parallel \varepsilon^2 \parallel$ .

Из ограниченности  $\| \omega \|$  следует, что интерполяционное значение  $Y(\mathbf{X}^*)$  сходится к значению функции  $Y_i$  в интерполяционном узле, а скорость сходимости имеет порядок  $\rho$ . Теорема доказана.

Замечание 13.1. При формировании матрицы метрической неопределенности (13.3) ее элементы можно определить и другими способами, в частности, скалярное произведение  $W_{ij} = (\mathbf{X}_i, \mathbf{X}_j)_{\mathbf{0}}$  можно вводить по-разному. В результате можно получить разные скорости сходимости интерполяционной функции к значениям функции в узлах интерполяции. Можно, например, ввести скалярное произведение как

$$(\mathbf{X}_i, \mathbf{X}_j)_{\mathbf{\omega}} = \sum_{k=1}^m \omega_k \cdot (X_{ik} - X_k^*)^{\alpha} \cdot (X_{jk} - X_k^*)^{\alpha},$$

где  $\alpha\geqslant 1$  — параметр. В этом случае скорость сходимости будет иметь порядок  $\rho^{\alpha}$ . Такой подход часто является удобным, когда в рассматриваемой прикладной задаче имеется требование

на скорость сходимости.

Теперь определим метрические веса  $\omega_i$ ,  $i=l,\ldots,m$ . Смысл этих весов состоит в том, что они учитывают характер изменения исследуемой функции при изменении ее аргументов. В этом случае матрица метрической неопределенности будет учитывать не только геометрическое расположение точек в исходном геометрическом пространстве, но и разный уровень изменения функции относительно разных аргументов функции.

Одна из возможных схем определения числовых значений метрических весов основана на исследовании степени влия-

ния каждого аргумента на изменение функции путем сравнения восстановленных значений функции в точке  $\mathbf{X}^*$  с включением аргумента и без его включения в усеченной точке  $\mathbf{X}^*_{(l)} = (\mathbf{X}^*_{(1)}, \dots, \mathbf{X}^*_{(l-1)}, \mathbf{X}^*_{(l+1)}, \dots, \mathbf{X}^*_{(m)})^T$  размерности m-l:

$$\widetilde{Y}^* = \frac{(\widetilde{W}^{-1}\mathbf{1},\mathbf{Y})}{(\widetilde{W}^{-1}\mathbf{1},\mathbf{1})}, \qquad \widetilde{Y}_l^* = \frac{(\widetilde{W}_l^{-1}\mathbf{1},\mathbf{Y})}{(\widetilde{W}_l^{-1}\mathbf{1},\mathbf{1})},$$

где элементы матриц  $\widetilde{W}$  и  $\widetilde{W}_{(l)}$  даются формулами

$$\begin{split} (\widetilde{W})_{ij} &= \sum_{k=1}^{m} (X_{ik} - X_k^*) \cdot (X_{jk} - X_k^*), \\ (\widetilde{W}_{(l)})_{ij} &= \sum_{k=1, k \neq l}^{m} (X_{ik} - X_k^*) \cdot (X_{jk} - X_k^*). \end{split}$$

В этом случае вместо весовой нормы (13.2) берем евклидову норму, и интерполяционные значения  $\widetilde{Y}$ ,  $\widetilde{Y}_{(l)}$  будут линейными относительно значений функций в узлах интерполяции Y. Далее определяется степень влияния l-го аргумента и метрические веса:

$$\widetilde{\omega}_1 = (\widetilde{Y}^{\bullet} - \widetilde{Y}^{\bullet}_{(l)})^2, \qquad \omega_1 = \frac{m\widetilde{\omega}_1}{\sum_{k=1}^m \widetilde{\omega}_k}, l = 1, \dots, m.$$

Замечание 13.2. Из  $\sum_{l=1}^{m} \omega_l = m$  следует, что если  $\omega_l > 1(\omega_l < 1)$ , то это указывает на больший (меньший) уровень изменения функции при изменении l-го аргумента по отношению к одинаковой степени чувствительности функции к изменению аргументов.

Замечание 13.3. Так как метрические веса зависят от значений функции в узлах интерполяции, то интерполяционные значения, получаемые методом метрического анализа, в общем случае будут нелинейно зависеть от заданных значений функции:  $Y^* = (\mathbf{z}(Y), Y)$ .

Таким образом, вышеуказанная схема выбора весовой метрики дает возможность выявить степень влияния каждого из аргументов и учесть различную степень влияния путем перехода к новой метрике с соответствующими неравными весами.

Рассмотрим вырожденный случай, когда в исследуемой функции введен аргумент (фактор), от которого функция не зависит. Тогда при реализации схемы нахождения весов новой метрики будет получен однозначный результат: метрический вес, соответствующий вышеуказанному фактору, будет равен нулю и этот фактор будет автоматически исключен из дальнейшего рассмот-

рения. Поэтому вышеприведенная схема перехода к метрике с весами дает возможность учесть, подобно тому, как это делается в факторном анализе, влияние аргументов на изменение функции и исключать несущественные из них, понижая размерность факторного пространства.

Перепишем задачу на минимум метрической неопределенности (13.9) в терминах собственных значений и собственных векторов матрицы W. Пусть  $\lambda_1,\ldots,\lambda_n$  — собственные значения, а  $\varphi_1,\ldots,\varphi_i$  — соответствующая им ортонормированная система собственных векторов матрицы W. Разложим вектор интерполяционных весов z по этой системе. Имеем

$$\mathbf{z} = \sum_{i=1}^{n} c_i \mathbf{\varphi}_i, \qquad W \mathbf{z} = \sum_{i=1}^{n} c_i \lambda_i \mathbf{\varphi}_i,$$

$$\sigma_{ND}^{2} = (W\mathbf{z}, \mathbf{z}) = \left(\sum_{i=1}^{n} c_{i} \lambda_{i} \mathbf{\varphi}_{i}, \sum_{j=1}^{n} c_{j} \mathbf{\varphi}_{j}\right) =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} c_{i} \lambda_{i} c_{j}(\mathbf{\varphi}_{j}, \mathbf{\varphi}_{j}) = \sum_{i=1}^{n} \lambda_{i} c_{i}^{2}, \quad (13.13)$$

$$(\mathbf{z},\mathbf{1}) = \left(\sum_{i=1}^n c_i \mathbf{\varphi}_i,\mathbf{1}\right) = \sum_{i=1}^n c_i(\mathbf{\varphi}_i,\mathbf{1}) = 1.$$

Задача (13.9) эквивалентна задаче

$$\begin{cases} \sum_{i=1}^{n} \lambda_{i} c_{i}^{2} - \min \mathbf{c}, \\ \sum_{i=1}^{n} c_{i}(\mathbf{\phi}_{i}, \mathbf{1}) = 1, \end{cases}$$
 (13.14)

где  $\mathbf{c} = (c_1, \dots, c_n)^T$ . Решим задачу (13.14) методом Лагранжа:

$$L(\mathbf{c}, u) = \frac{1}{2} \sum_{k=0}^{n} \lambda_i c_i^2 + u \left( 1 - \sum_{i=1}^{n} c_i(\mathbf{\phi}_i, \mathbf{1}) \right).$$

$$\frac{dL}{dc_1} = \lambda_i c_i - u(\mathbf{\phi}_i, \mathbf{1}) = 0 \Rightarrow c_i = \frac{u(\mathbf{\phi}_i, \mathbf{1})}{\lambda_i}, \qquad i = 1, \dots, n,$$

$$\sum_{i=1}^{n} c_i(\mathbf{\phi}_i, \mathbf{1}) = u \sum_{i=1}^{n} \frac{(\mathbf{\phi}_i, \mathbf{1})^2}{\lambda_i} = 1 \Rightarrow$$

$$\Rightarrow u = \frac{1}{\sum_{i=1}^{n} \frac{(\phi_{i}, 1)^{2}}{\lambda_{i}}}, \qquad c_{i} = \frac{\frac{(\phi_{i}, 1)}{\lambda_{i}}}{\sum_{i=1}^{n} \frac{(\phi_{i}, 1)^{2}}{\lambda_{i}}}.$$

Следовательно, вектор интерполяционных весов  $\mathbf{z}^*$  и интерполяционное значение  $Y^*$  для точки  $\mathbf{X}^*$  даются равенствами

$$\mathbf{z}^* = \frac{\sum_{i=1}^{n} \frac{(\boldsymbol{\varphi}_i, \mathbf{1})}{\lambda_i} \boldsymbol{\varphi}_i}{\sum_{i=1}^{n} \frac{(\boldsymbol{\varphi}_i, \mathbf{1})^2}{\lambda_i}}, \qquad Y^* = \frac{\sum_{i=1}^{n} \frac{(\boldsymbol{\varphi}_i, \mathbf{1})}{\lambda_i} (\boldsymbol{\varphi}_i, \mathbf{1})}{\sum_{i=1}^{n} \frac{(\boldsymbol{\varphi}_i, \mathbf{1})^2}{\lambda_i}}. \tag{13.15}$$

Таким образом, получили вектор интерполяционных весов и интерполяционное значение (13.10), выраженные через собственные векторы и собственные значения матрицы метрической неопределенности.

Пример 13.1. Рассмотрим случай, когда значение функции известно лишь в одной точке,  $Y_1 = F(\mathbf{X}_1)$ , и найдем значение в любой произвольной точке  $\mathbf{X}^*$ . Матрица метрической неопределенности  $W = \|\mathbf{X}_1 - \mathbf{X}^*\|^2$ , по формулам (13.10) получаем

$$W^{-1} = \frac{1}{\|\mathbf{X}_1 - \mathbf{X}^*\|^2}, \qquad = \frac{W^{-1}\mathbf{1}}{(W^{-1}\mathbf{1}, \mathbf{1})} = \mathbf{1} = 1, \qquad Y^* = Y_1,$$

или же, используя формулы (13.15):

$$\varphi_1 = 1,$$
  $\lambda_1 = ||\mathbf{X}_1 - \mathbf{X}^*||^2,$   $\mathbf{z} = \varphi_1 = 1,$   $Y^* = Y_1.$ 

Получаем, что при наличии информации о значении функции лишь в одной точке значение в любой другой точке, даваемое методом метрического анализа, совпадает со значением функции в этой заданной точке.

В случае n>m матрица метрической неопределенности W вырождена. Покажем, что при n>m+1 задача (13.9) имеет бесконечно много решений. Действительно, имеется бесконечно много функций  $z_i(\mathbf{X}^*),\ i=1,\dots,n$ , удовлетворяющих условию  $\sum_{i=1}^n z_i(\mathbf{X}^*) = 1$ , для которых мера метрической неопределенности равна нулю:

$$\sigma_{ND}^2 = (W\mathbf{z}, \mathbf{z}) = (YY^T\mathbf{z}, \mathbf{z}) = (Y^T\mathbf{z}, Y^T\mathbf{z}) = \mathbf{0} \Leftrightarrow Y^T\mathbf{z} = \mathbf{0},$$

где матрица Y определена в (13.5). Матрица  $Y^T$  имеет размер  $m \times n$ , отсюда система  $Y^T\mathbf{z} = \mathbf{0}$  вместе с уравнением  $\sum_{i=1}^n z_i = 1$ 

образуют систему из m+1 уравнений с n неизвестными. Поэтому при n>m+1 имеется бесконечно много интерполяционных функций вида

$$Y(\mathbf{X}^*) = \sum_{i=1}^n z_i(\mathbf{X}^*) Y_i, \quad \sum_{i=1}^n z_i(\mathbf{X}^*) = 1,$$
  
$$\sigma_{ND}^2 = (W\mathbf{z}(\mathbf{X}^*), \ \mathbf{z}(\mathbf{X}^*)) = 0,$$
 (13.16)

с нулевой мерой метрической неопределенности.

Для случая n=m+1 задача (13.9), так же как и в случае  $n \leq m$ , имеет единственное решение. Этот случай разобран ниже в качестве примера.

Покажем, что в одномерном случае m=1 для интерполяционных полиномов Лагранжа выполняются условия (13.16). Имеем:  $Y_1=f(x_1),\ldots,Y_n=f(x_n)$ . Построим интерполяционный полином Лагранжа  $L_n(x)$  из следующих соображений:

$$L_n(x) = \sum_{k=1}^{n} z_k(x) Y_k,$$
(13.17)

где  $Y_k$  представляют собой полином степени не выше чем n-1, тогда  $L_n(x)$  — также полином степени не выше чем n-1. Из требования  $L_n=Y_i$  и представления (13.17) получаем:  $z_k(x_i)=\sigma_{ik}$ , поэтому все n-1  $x_i, i\neq k$ , являются корнями полинома  $z_k(x)$ . Отсюда  $z_k(x)=C_k(x-x_k),\ldots,(x-x_{k-1})(x-x_{k+1}),\ldots,(x-x_n)$ . Константы  $C_k$  находим из условий  $z_k(x_k)=1$ :

$$C_k = \frac{1}{(x - x_k) \dots (x - x_{k-1})(x - x_{k+1}) \dots (x - x_n)},$$

$$z_k(x) = \frac{(x - x_k) \dots (x - x_{k-1})(x - x_{k+1}) \dots (x - x_n)}{(x_k - x_1) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n)}.$$

Функция  $\delta_n(x) = \sum_{k=1}^n z_k(x) - 1$  является полиномом степени не выше n-1 и обращается в нуль в n различных точках  $\delta_n(x_i) = 0$ , следовательно, она тождественно равна нулю и  $\sum_{k=1}^n z_k(x) \equiv 1$ :

$$\sigma_{ND}^{2}(x) = (W\mathbf{z}, \mathbf{z}) = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} z_{i}(x) z_{j}(x) =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} (x_{i} - x)(x_{j} - x) z_{i}(x) z_{j}(x) = \left(\sum_{i=1}^{n} (x - x_{i}) z_{i}(x)\right)^{2}.$$

Условие 
$$\sigma_{ND}^2(x) = 0 \Leftrightarrow \sum_{i=1}^n (x-x_i)z_i(x) = 0.$$

$$\sum_{i=1}^{n} (x - x_i) z_i(x) = \sum_{i=1}^{n} \frac{\prod_{k=1}^{n} (x - x_k)}{\prod_{k=1, k \neq i}^{n} (x_i - x_k)} = \prod_{k=1}^{n} (x - x_k) \sum_{i=1}^{n} \frac{1}{\prod_{k=1, k \neq i}^{n} (x_i - x_k)}$$

Покажем, что

$$\sum_{i=1}^{n} \frac{1}{\prod_{k=1, k \neq i}^{n} (x_i - x_k)} = 0,$$
(13.18)

тогда получим:  $\sigma_{ND}^2(x)\equiv 0$ . Равенство (13.18) доказывается методом математической индукции. Для n=2 получаем

$$\frac{1}{x_1-x_2}-\frac{1}{x_2-x_1}=0.$$

Предполагая, что равенство (13.18) верно для n=p, докажем, что оно верно для n=p+1:

$$\sum_{i=1}^{p+1} \frac{1}{\prod_{k=1}^{p+1} k \pm i} (x_i - x_k)} = \sum_{i=2}^{p+1} \frac{1}{\prod_{k=2, k \neq i}^{p} (x_i - x_k)} \times \frac{1}{(x_i - x_1)(x_i - x_{p+1})} + \frac{1}{\prod_{k=2}^{p+1} (x_1 - x_k)} + \frac{1}{\prod_{k=1}^{p} (x_{p+1} - x_k)} = \frac{1}{x_1 - x_{p+1}} \sum_{i=1}^{n} \frac{1}{\prod_{k=2, k \neq i}^{p} (x_i - x_k)} \left[ \frac{1}{x_i - x_1} - \frac{1}{x_i - x_{p+1}} \right] + \frac{1}{x_1 - x_{p+1}} \cdot \frac{1}{\prod_{k=2}^{p} (x_1 - x_k)} = \frac{1}{x_1 - x_{p+1}} \left[ \sum_{i=1}^{p} \frac{1}{\prod_{k=2, k \neq i}^{p} (x_i - x_k)} - \sum_{i=2}^{p+1} \frac{1}{\prod_{k=2, k \neq i}^{p} (x_i - x_k)} \right] = \frac{1}{x_1 - x_{p+1}} [0 - 0] = 0.$$

Таким образом, равенство (13.18) верно для любого n и полиномы Лагранжа удовлетворяют всем условиям (13.16).

Чтобы обойти сложности, связанные с вырожденностью матрицы метрической неопределенности W, можно использовать различные подходы.

Одним из подходов является использование вместо матриц W регуляризованной матрицы  $W_r = W + \alpha M$ , где M — матрица-регуляризатор,  $\alpha$  — параметр регуляризации. Например, в качестве матрицы M может быть использована диагональная матрица, диагональ которой состоит из диагонали матрицы W:

$$W_r = W + \alpha \operatorname{diag}(W_{11}, \dots, W_{nn}) =$$
  
=  $W + \alpha \operatorname{diag}(\rho^2(\mathbf{X}_1, \mathbf{X}^*)_{\mathbf{\omega}}, \dots, \rho^2(\mathbf{X}_n, \mathbf{X}^*)_{\mathbf{\omega}}).$  (13.19)

Как известно, если для произвольной  $(n \times n)$ -матрицы  $A = (\alpha_{ij})$  выполняются n неравенств

$$|\alpha_{ij}| - \sum_{j=1, j \neq i}^{n} |\alpha_{ij}| > 0, \quad i = 1, \dots, n,$$
 (13.20)

то матрица A является невырожденной. Для регуляризованной матрицы  $W_r$  условия (13.20) имеют вид

$$(1+\alpha)W_{ii} - \sum_{j=1, j\neq i}^{n} |W_{ij}| > 0, \quad i=1,\ldots,n.$$

Поэтому при достаточно больших  $\alpha$  матрица  $W_r$  невырожденная. В частности, из (13.4) и неравенства Коши-Буняковского следует

$$(1+\alpha)W_{ii} - \sum_{j=1,j\neq i}^{n} |W_{ij}| = (1+\alpha)\rho^{2}(\mathbf{X}_{i}, \mathbf{X}^{*})_{\omega} -$$

$$- \sum_{j=1,j\neq i}^{n} |(\mathbf{X}_{i}, \mathbf{X}^{*})_{\omega}| \ge (1+\alpha)\rho^{2}(\mathbf{X}_{i}, \mathbf{X}^{*})_{\omega} -$$

$$- \sum_{j=1,j\neq i}^{n} ||(\mathbf{X}_{i} ||_{\omega}|| \mathbf{X}^{*} ||_{\omega} = (1+\alpha)\rho^{2}(\mathbf{X}_{i}, \mathbf{X}^{*})_{\omega} -$$

$$- \sum_{j=1,j\neq i}^{n} \rho(\mathbf{X}_{i}, \mathbf{X}^{*})_{\omega} \rho(\mathbf{X}_{j}, \mathbf{X}^{*})_{\omega}, i = 1, \dots, n.$$

Следовательно, при

$$\alpha > \max \left[ \frac{\sum_{j=1, j \neq i}^{n} \rho(\mathbf{X}_{j}, \mathbf{X}^{*})_{\mathbf{o}}}{\rho(\mathbf{X}_{j}, \mathbf{X}^{*})_{\mathbf{o}}} - 1 \right]$$
(13.21)

матрица  $W_r$  будет не вырождена. Условие (13.21) является достаточным условием невырожденности матрицы  $W_r$ . При практической же реализации значения  $\alpha$ , которые подбираются в автоматическом режиме для обращения матрицы  $W_r$ , являются достаточно малыми. Значение параметра  $\alpha$  подбирается настолько возможно малым, чтобы число обусловленности матрицы  $W_r$  не превосходило некоторого порогового значения, обеспечивающего точность вычислений.

Другой подход основан на представлениях (13.13), где интерполяционные веса и мера метрической неопределенности представлены через собственные значения и собственные векторы матрицы W. Матрица W имеет n-m собственных значений, равных нулю:  $\lambda_i=0, i=1,\ldots,n-m,\ 0<\lambda_{n-m+1}<\ldots<\lambda_n.$  Положив коэффициенты  $c_i$  при нулевых собственных значениях равными нулю, из формул (13.13) получим

$$\mathbf{z} = \sum_{i=1}^{n-m} c_i \mathbf{\varphi}_i, \qquad \sigma_{ND}^2 = 0, \qquad \sum_{i=1}^{n-m} c_i (\mathbf{\varphi}_i, \mathbf{1}) = 1.$$

Следовательно, в этом случае ищем решение в подпространстве нулевого собственного значения. Рассмотрим следующую задачу на минимум метрической неопределенности:

$$\begin{cases}
\sum_{i=1}^{n-m} \mu_i c_i^2 - \min \mathbf{c}, \\
\sum_{i=1}^{n-m} c_i(\mathbf{\varphi}_i, \mathbf{1}) = 1,
\end{cases}$$
(13.22)

здесь  $\mu_{\parallel}>0$  — заданные числа для фиксированной точки  $X^*$ , в которой ищется интерполяционное значение.

Задача (13.22) решается так же, как и задача (13.14), с помощью метода Лагранжа. Вектор интерполяционных весов и интерполяционное значение в точке  $\mathbf{X}^*$  даются равенствами

$$\mathbf{z}^* = C \sum_{i=1}^{n-m} \frac{(\mathbf{\phi}_i, \mathbf{1})}{\mu_i} \mathbf{\phi}_i, \qquad Y^* = C \sum_{i=1}^{n-m} \frac{(\mathbf{\phi}_i, \mathbf{1})}{\mu_i} \frac{(\mathbf{\phi}_i, \mathbf{1})}{\mu_i} (\mathbf{\phi}_i, \mathbf{Y}),$$

$$C = \frac{1}{\sum_{i=1}^{n-m} \frac{(\mathbf{\phi}_i, \mathbf{1})^2}{\mu_i}}.$$
(13.23)

 ${\sf B}$ ыбор набора  ${\sf \mu}=(\mu_1,\ldots,\mu_{n-m},\mu_i>0)$  соответствуют выбору конкретной интерполяционной схемы в рамках метрического ана-

Замечание 13.4. Если в задаче (13.22) часть функций  $\mu_{i1}(\mathbf{X}^*),\ldots,\mu_{ik}(\mathbf{X}^*)$  ) положить равными нулю, то формулы для интерполяционных весов и интерполяционного значения будут иметь вид

$$\mathbf{z}^* = C \sum_{i \in I} \frac{(\mathbf{\varphi}_i, \mathbf{1})}{\mu_i} \mathbf{\varphi}_i, \qquad Y^* = C \sum_{i \in I} \frac{(\mathbf{\varphi}_i, \mathbf{1})}{\mu_i} \frac{(\mathbf{\varphi}_i, \mathbf{1})}{\mu_i} (\mathbf{\varphi}_i, \mathbf{Y}),$$

$$C = \frac{1}{\sum_{i \in I} \frac{(\mathbf{\varphi}_i, \mathbf{1})^2}{\mu_i}}$$
(13.24)

Замечание 13.5. Функции  $\mu_{i1}(\mathbf{X}^*)$  должны определяться таким образом, чтобы обеспечивать сходимость функции  $Y^*(\mathbf{X}^*)$  к значениям  $Y_i$  в интерполяционных узлах.

Пример 13.2.  $m=1, n=2.Y_1=\dot{F}(x_1), Y_2=F(x_2)$ . Найдем

интерполяционное значение  $Y^st$  в точке  $x^st$ . Имеем

$$W = \begin{pmatrix} (x_1 - x^*)^2 & (x_1 - x^*)(x_2 - x^*) \\ (x_2 - x^*)(x_1 - x^*) & (x_2 - x^*)^2 \end{pmatrix},$$

$$\lambda_1 = 0, \quad \varphi_1 = \frac{1}{\sqrt{(x_1 - x^*)^2 + (x_2 - x^*)^2}} (-(x_2 - x^*), x_1 - x^*)^T,$$

$$\mathbf{z}(x^*) = \frac{1}{x_1 - x_2} (-(x_2 - x^*), x_1 - x^*)^T,$$

$$Y(x^*) = (\mathbf{z}^*, \mathbf{Y}) = Y_1 \frac{x_2 - x^*}{x_2 - x_1} + Y_2 \frac{x^* - x_1}{x_2 - x_1}.$$

График функции  $Y(x^st)$  является прямой, соединяющей точки

 $(x_2, Y_2)$  и  $(x_1, Y_1)$ .

**Пример 13.3**. m > 1, n = m + 1. В этом случае задача (13.14) имеет единственное решение, в отличие от случая n>m+1, когда имеется множество различных решений в рамках метрического анализа, в зависимости от выбора набора  $\mu=(\mu_1,\dots,\mu_{n-m},\mu_i>0)$ . Дано  $Y_1=f(\mathbf{X}_1),\dots,Y_n=f(\mathbf{X}_n)$ , най-дем интерполяционное значение  $Y^*$  в точке  $\mathbf{X}^*$ . Имеем:

$$\lambda_1 = 0 < \lambda_2 < \lambda < \ldots < \lambda_n; \quad W \varphi_1 = 0, \quad || \varphi_1 || = 1.$$

Из формул (13.23) получаем

$$C = \frac{1}{\frac{(\varphi_1, 1)^2}{\mu_1}}, \qquad \mathbf{z}^* = C \frac{(\varphi_1, 1)}{\mu_1} \varphi_1 = \frac{\varphi_1}{(\varphi_1, 1)}, \qquad Y^* = \frac{(\varphi_1, Y)}{(\varphi_1, 1)}.$$

Для случая m=2 эта схема интерполяции (по трем ближайшим точкам) подобна известной схеме триангуляции. Поэтому такую схему можно назвать схемой «мультиангуляции».

Еще один способ обойти сложности, связанные с существованием бесконечного числа решений в случае n>m+1, заключается в использовании модифицированной матрицы метрической неопределенности. Рассмотрим матрицу, элементы которой определяются по формулам

$$W_{ij} = (\mathbf{X}_1 - \mathbf{X}^*, \mathbf{X}_j - \mathbf{X}^*) \cdot \rho^{\alpha}(\mathbf{X}_i, \mathbf{X}_j), \tag{13.25}$$

где параметр  $\alpha > 1$ .

Такая матрица, в отличие от матрицы метрической неопределенности (13.3), учитывает не только расположение точки  $\mathbf{X}^*$  относительно пары точек  $\mathbf{X}_i, \mathbf{X}_j$ , но и расстояние между точками  $\mathbf{X}_i, \mathbf{X}_j$ . Эта матрица в общем случае не является вырожденной, вне зависимости от соотношения между n и m. Такой метод интерполяции дает хорошие результаты в случае размерностей m=1 и m=2 с параметром  $\alpha=3$ . Параметр  $\alpha$ , при условии его допустимости, влияет на гладкость кривых и поверхностей. В общем случае рекомендуется также использовать значение  $\alpha=3$ .

## 2. Примеры интерполяции

Рассмотрим случай одномерной интерполяции, m=1. Определим следующую регуляризованную матрицу  $W_r$  для матрицы W:

$$W_r = W + \alpha(\operatorname{diag}(W_{11}, \dots, W_{nn}) + 0, 5 \operatorname{diag}(W_{12}, \dots, W_{n-1,n}) + 0, 5 \operatorname{diag}(W_{21}, \dots, W_{nn-1})),$$
 (13.26)

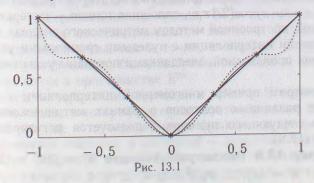
где  $\operatorname{diag}(W_{12},\dots,W_{n-1n})$  и  $\operatorname{diag}(W_{21},\dots,W_{nn-1})$  — матрицы с ненулевыми верхней и нижней побочными диагоналями соот-

ветственно. В формулах (13.23) в качестве функций  $\mu_1, \ldots, \mu_{n-1}$  возьмем упорядоченные по возрастанию собственные значения  $\lambda_1, \ldots, \lambda_{n-1}$  матрицы  $W_r$ , а вместо собственных векторов матрицы W возьмем собственные векторы  $\mathbf{\phi}_i$  матрицы  $W_r$ , соответствующие собственным значениям  $\lambda_i$ .

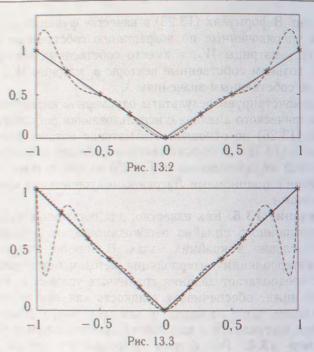
Продемонстрируем результаты одномерной интерполяции методом метрического анализа с использованием регуляризованной матрицы (13.26) по отношению к матрице метрической неопределенности (13.3) с использованием модифицированной матрицы метрической неопределенности (13.25) на конкретных примерах в сравнении с полиномами Лагранжа и кубическими одномерными сплайнами.

Замечание 13.5. Как известно, для построения кубического интерполяционного сплайна необходимо задать дополнительно краевые условия в крайних узлах. В отличие от одномерной сплайн-интерполяции интерполяция методом метрического анализа не предполагает задания граничных условий и, как будет показано ниже, обеспечивает гладкость как внутри интервала интерполяции, так и на краях, и, как следствие, хорошо отражает участки выпуклости и вогнутости интерполируемой функции.

Пример 13.4. Рассмотрим функцию y=|x| на отрезке [-1,1]. Значения функции заданы на равномерной сетке с n узлами интерполяции. На рис. 13.1-13.3 показаны результаты для  $n=7,\ n=9$  и n=11 соответственно. Сплошная жирная кривая соответствует графику функции y=|x|; сплошная тонкая — интерполяционной кривой, построенной методом метрического анализа; пунктирная — полиному Лагранжа. Звездочками показаны узлы интерполяции.



Известно, что последовательность интерполяционных полиномов Лагранжа  $L_n(x)$  для функции y=|x|, построенных на равномерной сетке, расходится в любой точке  $x\neq -1,0,1$ .



**Пример** 13.5. Рассмотрим функцию y(x) = 4

 $-\exp(x)\cos(2.1\pi x)$  на отрезке [-1,1].

Значения функции заданы на равномерной сетке с n узлами интерполяции. Вычисляются интерполяционные значения N=100 точек, заданных равномерно на отрезке  $[-1,\ 1]$ . На рис. 13.4 показан результат интерполирования для n=10. Сплошная жирная кривая соответствует графику функции  $y(x)=4-\exp(x)\cos(2.1\pi x)$ , сплошная тонкая — интерполяционной кривой, построенной методом метрического анализа, штриховая — сплайн интерполяции с нулевыми граничными условиями для второй производной, звездочками показаны узлы интерполяции.

Рассмотрим пример многомерной интерполяции с использованием различных подходов в рамках метрического анализа. Для следующего примера используется регуляризованная матрица (13.9).

Пример 13.6 Рассмотрим функцию  $Y = f(\mathbf{x}) = (V\mathbf{x}, \mathbf{x}) + (\mathbf{c}, \mathbf{x}), \quad \mathbf{x} = (x_1, \dots, x_m)^T, \quad 0 \leqslant x_i \leqslant 1, \quad i = 1, \dots, m, \quad \text{где}$   $V = \text{постоянная } (m \times m)$ -матрица,  $\mathbf{c} = (c_1, \dots, c_m)^T = \text{постоян-ный вектор-столбец, } m = 12 - \text{мерность пространства.}$   $Y(\mathbf{X}_k) = Y_k, \quad k = 1, \dots, n$ , где  $\mathbf{X}_k = \mathbf{y}$  узлы интерполяции.

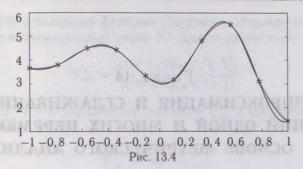


Таблица 1

N	Точное значение	Интерпол. значение	Отн. ошибка (%)	N	Точное значение	Интерпол. значение	Отн. ошибка (%)
1	38,00	38,59	1,54	11	35,18	35,89	2,02
2	44,38	45,81	3,20	12	44,57	45,28	1,59
3	44,40	46,44	4,58	13	42,63	44,04	3,30
4	49,94	49,42	1,04	14	49,50	49,89	0,79
5	36,30	37,66	3,74	15	25,84	25,22	2,38
6	44,76	45,42	1,48	16	41,00	41,38	0,92
7	43,19	42,45	1,70	17	55,39	53,83	2,83
8	43,46	42,82	1,47	18	42,05	44,93	6,86
9	37,43	36,51	2,45	19	38,95	39,86	2,35
10	30,78	29,06	5,60	20	47,37	47,22	0,32

 $Y_k$  — значения функции в узлах интерполяции, n=25 количество интерполяционных узлов, N=20 — количество точек, в которых восстанавливаются значения функции. Как интерполяционные узлы, так и точки, в которых интерполируются значения функции, были выбраны с помощью генератора случайных точек в пространстве  $E^{\hat{m}}$ .

Результат интерполяции приведен в табл. 1. В первой колонке — номер точки, в которой интерполируется значение функции, во второй колонке — точное значение функции, в третьей — интерполированное значение, в четвертой — относительная ошибка в процентах. Несмотря на небольшое количество узлов интерполяции, мы получили достаточно высокую точность интерполированных значений.

#### Глава 14

## АППРОКСИМАЦИЯ И СГЛАЖИВАНИЕ ФУНКЦИЙ ОДНОЙ И МНОГИХ ПЕРЕМЕННЫХ НА ОСНОВЕ МЕТРИЧЕСКОГО АНАЛИЗА

# 1. Схемы аппроксимации и сглаживания методом метрического анализа

Рассмотрим задачу сглаживания и аппроксимации функциональной зависимости

$$Y = F(X_1, \dots, X_m) = F(\mathbf{X})$$

при наличии хаотических отклонений от точных значений.

Заданные в точках  $\mathbf{X}_1,\dots,\mathbf{X}_n$  значения функции  $Y=F(\mathbf{X}_1),\dots,Y_n=F(\mathbf{X}_n)$  известны с погрешностями. Пусть  $K_{\mathbf{Y}}$  — ковариационная матрица вектора значений  $\mathbf{Y}=(Y_1,\dots,Y_n)^T$ .

Случай 1. Матрица метрической неопределенности невырожлена

Для произвольной точки  $X^*$  восстановленное значение  $Y^*$  ищем в представлении

$$Y^* = (\mathbf{z}, \mathbf{Y}), \tag{14.1}$$

где вектор весов  ${\bf z}$  находится как решение следующей задачи на минимум суммарной детерминированной и стохастической неопределенности  $K_Y$ :

$$\begin{cases} (W\mathbf{z}, \mathbf{z}) + \alpha(K_Y\mathbf{z}, \mathbf{z}) - \min \mathbf{z}, \\ (\mathbf{z}, \mathbf{1}) = 1, \qquad \mathbf{1} = (1, \dots, 1)^T, \end{cases}$$
(14.2)

0 > 0 — параметр сглаживания.

В то время как выражение  $(W\mathbf{z}, \mathbf{z})$  отвечает за метрическую неопределенность, выражение  $(K_Y\mathbf{z}, \mathbf{z})$  отвечает за стохастическую неопределенность. Задача (14.2) решается так же, как за-

дача (13.9) с помощью функции Лагранжа. Искомое сглаженное значение в произвольной точке  $\mathbf{X}^*$  дается равенством

$$Y_{sm} = \frac{((W + \alpha K_{\mathbf{Y}})^{-1} \mathbf{1}, \mathbf{Y})}{((W + \alpha K_{\mathbf{Y}})^{-1} \mathbf{1}, \mathbf{1})}.$$
 (14.3)

При  $\alpha \to +\infty$  сглаженное значение  $Y^*_{*m}$  в произвольной точке  $\mathbf{X}^*$  одно и то же:

$$Y_{sm}^* = \frac{(K_{\mathbf{Y}}^{-1}\mathbf{1}, \mathbf{Y})}{(K_{\mathbf{Y}}^{-1}\mathbf{1}, \mathbf{1})}.$$
 (14.4)

При  $\alpha \to +0$  решение задачи (14.2) для  $\mathbf{X}^* \neq \mathbf{X}_i, i=1,\dots,n,$  дает интерполяционную функцию

$$Y^* = \frac{(W^{-1}\mathbf{1}, \mathbf{Y})}{(W^{-1}\mathbf{1}, \mathbf{1})}. (14.5)$$

**Теорема 14.1.** Для точки  $\mathbf{X}^* = \mathbf{X}_i$  при  $\alpha \to +0$  решение задачи (14.2) стремится к значению функции  $Y_i$ .

Доказательство. Действительно, имеем

$$\begin{split} L(\mathbf{z}, u) &= \frac{1}{2} \left( (W + \alpha K_{\mathbf{Y}}) \mathbf{z}, \mathbf{z} \right) - u \left( (\mathbf{z}, \mathbf{1}) - 1 \right), \\ \nabla_{\mathbf{z}} L &= (W + \alpha K_{\mathbf{Y}}) \mathbf{z} - u \mathbf{1} = 0. \end{split}$$

Так как i-я строка и i-й столбец матрицы W состоят из нулей, то из системы уравнений

$$(W + \alpha K_{\mathbf{Y}})\mathbf{z} = u\mathbf{1} \tag{14.6}$$

получаем

$$\alpha \sum_{j=1}^{n} (K_{\mathbf{Y}})_{ij} z_j = u z_i = \frac{u}{\alpha (K_{\mathbf{Y}}) u} - \frac{(\mathbf{K}_i, \widetilde{\mathbf{z}})}{K_{\mathbf{Y}})_{ii}}, \tag{14.7}$$

где введены обозначения  $\mathbf{\bar{z}} = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n)^T$ ,  $\mathbf{K}_i = ((K_{\mathbf{Y}})_{i1}, \dots, (K_{\mathbf{Y}})_{i,i-1}, (K_{\mathbf{Y}})_{i,i+1}, \dots, (K_{\mathbf{Y}})_{in})^T$ ; r-е уравнение системы (14.6),  $r \neq i$ , перепишем в виде

$$\sum_{j=1,j\neq i}^{n} \left( W_{rj} + \alpha(K_{\mathbf{Y}})_{rj} - \alpha \frac{(K_{\mathbf{Y}})_{ri}}{(K_{\mathbf{Y}})_{ii}} (K_{\mathbf{Y}})_{ij} \right) z_j = u \left( 1 - \frac{(K_{\mathbf{Y}})_{rj}}{(K_{\mathbf{Y}})_{ii}} \right).$$

Oбозначим через  $M_{lpha}$  матрицу размера (n-1) imes (n-1) с эле ментами

$$(M_{\alpha})_{rj} = W_{rj} + \alpha (K_{\mathbf{Y}})_{rj} - \alpha \frac{(K_{\mathbf{Y}})_{ri}}{(K_{\mathbf{Y}})_{ii}} (K_{\mathbf{Y}})_{ij},$$
  
 $r, j = 1, \dots, i - 1, i + 1, \dots, n, \quad (14.8)$ 

а через  ${\bf f}$  — вектор-столбец размера  $(n-1) \times 1$  с элементами

$$f_r = \left(1 - \frac{(K_Y)_{rj}}{(K_Y)_{ii}}\right), \qquad r = 1, \dots, i-1, i+1, \dots, n.$$

Тогда получаем 
$$\widetilde{\mathbf{z}} = u M_{\alpha}^{-1} \mathbf{f}, \tag{14.9}$$

а из (14.7)

$$z_i = u \left( \frac{1}{\alpha(K_{\mathbf{Y}})_{ii}} - \frac{(\mathbf{K}_i, M_{\alpha}^{-1} \mathbf{f})}{(K_{\mathbf{Y}})_{ii}} \right). \tag{14.10}$$

Из условия (2.1) = 1 получаем

$$u(\alpha) = \frac{\alpha(K_{\mathbf{Y}})_{ii}}{1 + \alpha(K_{\mathbf{Y}})_{ii}(M_{\alpha}^{-1}\mathbf{f}, \mathbf{1}) - \alpha(\mathbf{K}_{i}, M_{\alpha}^{-1}\mathbf{f})}.$$
 (14.11)

Подставляя выражение (14.11) в (14.9) и в (14.10), имеем

$$\widetilde{\mathbf{z}} = \frac{\alpha(K_{\mathbf{Y}})_{ii} M_{\alpha}^{-1} \mathbf{f}}{1 + \alpha(K_{\mathbf{Y}})_{ii} (M_{\alpha}^{-1} \mathbf{f}, 1) - \alpha(K_{\tau}, M_{\alpha}^{-1} \mathbf{f})},$$
(14.12)

$$z_{i} = \frac{1 - (\mathbf{K}_{i}, M_{\alpha}^{-1} \mathbf{f})}{1 + \alpha(\mathbf{K}_{Y})_{ii}(M_{\alpha}^{-1} \mathbf{f}, 1) - \alpha(\mathbf{K}_{i}, M_{\alpha}^{-1} \mathbf{f})}.$$
 (14.13)

Обозначим через W матрицу, которая получается из матрицы И удалением і-го столбца и і-й строки (которые состоят из нулей).

Тогда из представления (14.8) следует, что

$$M_{\alpha} \xrightarrow[\alpha \to 0]{} \widetilde{W},$$

н из (14.14), (14.13) имеем

$$\overline{z} \xrightarrow[\alpha \to 0]{} 0, \qquad z_i \xrightarrow[\alpha \to 0]{} 1.$$

Из (14.1) для точек  $X^* = X_i$  окончательно получаем

$$Y^* \xrightarrow[\alpha \to 0]{} Y_i$$

Теорема доказана.

Задачу (14.2) можно также решить с помощью собственных значений и собственных векторов матрицы W. Пусть  $\lambda_1,\ldots,\lambda_n$  — собственные значения, а  $\varphi_1,\ldots,\varphi_n$  — соответствующая им ортонормированная система собственных векторов матрицы W.

Разложим вектор весов z по этой системе:

$$\mathbf{z} = \sum_{i=1}^n c_i \mathbf{\varphi}_i.$$

Обозначим через  $F=[\mathbf{\phi}_1,\ldots,\mathbf{\phi}_n]$  матрицу, столбцы которой образованы собственными векторами матрицы  $W,\ \mathbf{c}=(c_1,\ldots,c_n)^T.$  Тогда

$$\mathbf{z} = F\mathbf{c}.\tag{14.14}$$

Для суммарной неопределенности имеем

$$(W\mathbf{z}, \mathbf{z}) + \alpha(K_{\mathbf{Y}}\mathbf{z}, \mathbf{z}) = \sum_{i=1}^{n} \lambda_{i} c_{i}^{2} + \sum_{i,j=1}^{n} c_{i} c_{j} (K_{\mathbf{Y}} \varphi_{i}, \varphi_{j}) =$$
$$= (D\mathbf{c}, \mathbf{c}) + \alpha(B\mathbf{c}, \mathbf{c}),$$

где

$$D = \operatorname{diag}(\lambda_1, \dots, \lambda_n), \qquad B_{ij} = (K_{\mathbf{Y}} \mathbf{\varphi}_i, \mathbf{\varphi}_j), \quad i, j = 1, \dots, n,$$
$$(\mathbf{z}, \mathbf{1}) = \sum_{i=1}^n c_i(\mathbf{\varphi}_j, \mathbf{1}) = (\mathbf{c}, \mathbf{\phi}), \qquad \mathbf{\phi} = ((\mathbf{\varphi}_1, \mathbf{1}), \dots, (\mathbf{\varphi}_n, \mathbf{1}))^T.$$

Тогда задача (14.2) для нахождения вектора  ${\bf z}$  сводится к следующей задаче на нахождение вектора  ${\bf c}$ :

$$\begin{cases} (D\mathbf{c}, \mathbf{c}) + \alpha \cdot (B\mathbf{c}, \mathbf{c}) - \min \mathbf{c}, \\ (\mathbf{c}, \phi) = 1. \end{cases}$$
 (14.15)

Задача (14.15) решается с помощью функции Лагранжа:

$$L(\mathbf{c}, u) = \frac{1}{2} ((D + \alpha B)\mathbf{c}, \mathbf{c}) + u (1 - (\mathbf{c}, \phi)),$$

$$\nabla_{\mathbf{c}} L = (D + \alpha B)\mathbf{c} - u\phi = \mathbf{0} \Rightarrow u(D + \alpha B)^{-1}\phi,$$

$$(\mathbf{c}, \phi) = 1 \Rightarrow u = \frac{1}{((D + \alpha B)^{-1}\phi, \phi)}.$$

Отсюда имеем вектор с\*:

$$\mathbf{c}^* = \frac{(D + \alpha B)^{-1} \phi}{((D + \alpha B)^{-1} \phi, \phi)}.$$
 (14.16)

Из (14.14) получаем искомый вектор весов z\*

$$\mathbf{z}^* = F \frac{(D + \alpha B)^{-1} \mathbf{\phi}}{((D + \alpha B)^{-1} \mathbf{\phi}, \mathbf{\phi})},$$
 (14.17)

сглаженное значение

$$Y_{sm}^{\bullet} = \frac{\left(F(D+\alpha B)^{-1}\phi, \mathbf{Y}\right)}{\left((D+\alpha B)^{-1}\phi, \phi\right)}.$$
 (14.18)

Замечание 14.1. Из доказательства теоремы (14.1) следует, что сглаженные значения в самих точках  $\mathbf{X}_i$  находятся с помощью формул (14.12) и (14.13), которые определяют вектор весов  $\mathbf{z}$ .

Замечание 14.2. В качестве суммарной неопределенности вместо

$$(W\mathbf{z},\mathbf{z}) + \alpha(K_{\mathbf{Y}}\mathbf{z},\mathbf{z})$$

можно взять

$$(1-\alpha)(Wz,z) + \alpha(K_Yz,z),$$

и тогда параметр  $\alpha$  будет меняться в пределах [0,1] вместо  $[0,+\infty]$ . Такое рассмотрение более удобно для вычислений.

Случай 2. Матрица метрической неопределенности вырождена.

В этом случае, так же как и для задачи интерполяции. применяется один из описанных в гл. 13 подходов.

В первом подходе вместо вырожденной матрицы метрической неопределенности используется регуляризованная матрица  $W_r = W + \alpha M$ , где M — матрица-регуляризатор,  $\alpha$  — параметр регуляризации, а вместо задачи (14.2) рассматривается задача с использованием регуляризованной матрицы  $W_r$ :

$$\begin{cases} (W_r \mathbf{z}, \mathbf{z}) + \alpha(K_Y \mathbf{z}, \mathbf{z}) - \min \mathbf{z}, \\ (\mathbf{z}, \mathbf{1}) = 1, \qquad 1 = (1, \dots, 1)^T. \end{cases}$$
(14.19)

Во втором подходе действуем в подпространстве нулевого собственного значения,  $\phi_1, \ldots, \phi_{n-m}$  — собственные векторы матрицы W, соответствующие нулевому собственному

значению  $\lambda_0 = 0$ :

$$\mathbf{z} = \sum_{i=1}^{n-m} c_i \mathbf{\varphi}_i, \qquad W \mathbf{\varphi}_i = \mathbf{0}, \qquad (\mathbf{\varphi}_i, \mathbf{\varphi}_j) = \delta_{ij}, \quad i, j = 1, \dots, n-m.$$

Обозначим  $F_0 = [\varphi_1, \dots, \varphi_{n-m}]$  матрицу, столбцы которой образуют собственные векторы матрицы W, соответствующие нулевому собственному значению  $\mathbf{c} = (c_1, \dots, c_{n-m})^T$ , тогда вектор весов  $\mathbf{z}$  можно представить в виде

$$\mathbf{z} = F_0 \mathbf{c}_0, \phi_0 = ((\mathbf{\phi}_1, \mathbf{1}), \dots, (\mathbf{\phi}_n, \mathbf{1}))^T, \quad (B_0)_{i\bar{j}} = (K_{\mathbf{Y}} \mathbf{\phi}_i, \mathbf{\phi}_j), i, j = 1, \dots, n - m, D_0 = \operatorname{diag}(\mu_1, \dots, \mu_n),$$
 (14.20)

где функции  $\mu_i(\mathbf{X}^*)$  должны удовлетворять условию интерполяции при  $\alpha=0$ , т. е. определять интерполяционную функцию, сходящуюся к значениям функции в точках  $\mathbf{X}_i$ . Искомый вектор  $\mathbf{c}_0$ , вектор весов  $\mathbf{z}$  и сглаженное значение  $Y^*_m$  определяются формулами

$$\mathbf{c}_0 = \frac{(D + \alpha B_0)^{-1} \phi_0}{((D + \alpha B_0)^{-1} \phi_0, \phi_0)},$$
(14.21)

$$\mathbf{z}^* = \frac{F_0(D + \alpha B_0)^{-1} \phi_0}{((D + \alpha B_0)^{-1} \phi_0, \phi_0)}.$$
 (14.22)

$$Y_{sm}^* = \frac{(F_0(D + \alpha B_0)^{-1} \phi_0, \mathbf{Y})}{((D + \alpha B_0)^{-1} \phi_0, \phi_0)}.$$
 (14.23)

В третьем подходе для вырожденного случая вместо матрицы метрической неопределенности (13.3) в задаче (14.2) используется модифицированная матрица метрической неопределенности. Например, для мерностей пространства  $m=1,\ m=2$  эффективно использование матрицы (13.25).

Если сглаживание производится в нескольких точках  $\mathbf{X}_1^*,\dots,\mathbf{X}_N^*$  то, в общем случае, задавая различные  $\alpha$ , получим различные уровни сглаживания. Однако если требуется восстановить (аппроксимировать) исходную функциональную зависимость, где значения функции в каждой заданной точке известны с погрешностями, то сглаживание нужно проводить со значением  $\alpha$ , найденным с помощью принципа невязки.

Пусть в точках  $\mathbf{X}_1,\dots,\mathbf{X}_n$  значения функции заданы с погрешностями, дисперсии которых равны  $\sigma_1^2,\dots,\sigma_n^2$  соответственно. В этом случае ковариационная матрица имеет

диагональный вид:  $K_{\mathbf{Y}} = \operatorname{diag}(\sigma_1^2, \dots, \sigma_n^2)$ . Введем функционал невязки

$$\delta(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \frac{(Y_{sm}^{*}(\alpha, \mathbf{X}_{i}) - Y_{i})^{2}}{\sigma_{i}^{2}} - 1, \qquad (14.24)$$

где  $Y_{sm}^*(\alpha, \mathbf{X}_i)$  — сглаженные значения в узлах  $\mathbf{X}_i$ , в которых заданы значения функции.

Для восстановления исходной функциональной зависимости выбирается такое значение  $\alpha_0$ , для которого

$$\delta(\alpha_0) = 0. \tag{14.25}$$

По найденному значению  $\alpha_0$  для точек  $\mathbf{X}_1^*,\dots,\mathbf{X}_n^*$  находятся восстановленные значения  $Y_{sm}^*(\alpha_0,\mathbf{X}_k^*),\ k=1,\dots,n.$ 

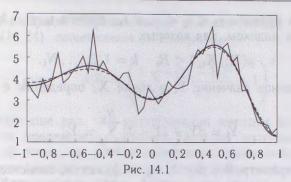
Величина

$$Y_{ch}(\mathbf{X}_i) = Y_i - Y_{sm}^*(\alpha_0, \mathbf{X}_i^*)$$

представляет собой хаотическую компоненту значения функции в точке Х:

Фактическое нахождение подходящего значения  $\alpha_0$  согласно способу невязки сводится к последовательному нахождению сглаженных значений  $Y^*_{sm}(\alpha_0, \mathbf{X}^*_i)$ ,  $k=1,\ldots,n$  для различных  $\alpha$  и выбору таких сглаженных значений, для которых наиболее точно выполняется равенство (14.25). Так как параметр сглаживания  $\alpha$  является непрерывным, то при численной реализации можно найти подходящее значение  $\alpha_0$  таким образом, чтобы равенство (14.25) выполнялось с наперед заданной точностью.

Пример 14.1. Рассмотрим функцию  $f(x) = 4 + \exp(x) \sin(2\pi(x+\pi/4))$ , заданную на отрезке [-1,1] в n=50 точках  $x_1, \ldots, x_n$  на равномерной сетке. Зададим вектор относительных погрешностей  $\sigma_1, \ldots, \sigma_n$  с помощью генератора случайных чисел в интервале от 1% до 25% от значений функции f(x), наложим шум на функцию f(x) с относительными погрешностями  $\sigma_i$  в точках  $x_i$  соответственно и тем самым получим зашумленную функцию  $f_{sh}(x)$ . Далее восстанавливаем исходную функцию f(x) с помощью аппроксимационной схемы метрического анализа с использованием функционала невязки (14.24). На рис. 14.1 представлены графики функции f(x) (сплошная жирная кривая), зашумленной функции  $f_{sh}(x)$  (пломаная кривая) и аппроксимационной функции  $f_{ap}(x)$  (сплошная тонкая кривая).



Подсчитаем средние относительные погрешности:

$$\sigma_{sh} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{f_{sh}(x_i) - f(x_i)}{f(x_i)} \right| = 12, 24\%,$$

$$\delta_{ap} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{f_{ap}(x_i) - f(x_i)}{f(x_i)} \right| = 2, 56\%.$$

В случае когда дисперсия ошибки неизвестна, но она одинакова для всех точек  $\mathbf{X}_i,\ 1,\dots,n$ , для определения подходящего значения  $\alpha_0$  можно использовать способ перекрестной значимости, согласно которому

$$\alpha_0 = \underset{\alpha}{\operatorname{argmin}} \gamma(\alpha),$$
 (14.26)

функционал  $\gamma(\alpha)$ :

$$\gamma(\alpha) = \sum_{i=1}^{n} \left( Y_{sm}^{(k)*}(\alpha, \mathbf{X}_k) - Y_k \right)^2, \tag{14.27}$$

где  $Y_{sm}^{(k)*}(\alpha, \mathbf{X}_k)$  — сглаженное значение в точке  $\mathbf{X}_k$ , построенное методом метрического анализа по n-1 значениям функции  $Y_1,\ldots,Y_{k-1},Y_{k+1},\ldots,Y_n$  в n-1 точках  $\mathbf{X}_1,\ldots,\mathbf{X}_{k-1},\mathbf{X}_{k+1},\ldots,\ldots,\mathbf{X}_n$ . За подходящее значение параметра сглаживания берется то, которое минимизирует функционал перекрестной значимости (14.27).

После аппроксимации функциональной зависимости для значений функции  $Y_1, \ldots, Y_n$  в точках  $\mathbf{X}_1, \ldots, \mathbf{X}_n$  имеем следующее представление:  $Y_i = Y_{sm}(\mathbf{X}_i) + Y_{ch}(\mathbf{X}_i). \tag{14.28}$ 

Можно рассмотреть и другой подход в рамках метрического анализа к задаче сглаживания на основе информации о взаимном расположении точек  $\mathbf{X}_1,\dots,\mathbf{X}_n$ . Пусть  $Y_i$  — значения функции в точках  $\mathbf{X}_i,\ i=1,\dots,n$ . Зададим число R>0. Для точки  $\mathbf{X}_i$ 

определим индексы  $i_1 < i_2 < \ldots < i_n$ , причем  $i_k \neq i$ ,  $k = 1, 2, \ldots$ ,  $N_i$ , как индексы, для которых

$$\rho(\mathbf{X}_i, \mathbf{X}_{i_k}) < R, \quad k = 1, 2, \dots, N_i.$$
 (14.29)

Сглаженное значение  $\widetilde{Y}_i$  в точке  $\mathbf{X}_i$  определим с помощью равенства

 $\widetilde{Y}_{i} = \beta Y_{i} + (1 - \beta) \sum_{k=1}^{N_{i}} \gamma_{i_{k}} Y_{i_{k}},$  (14.30)

где  $\beta$  — параметр.  $0<\beta<1,\ \gamma_{i_k}$  — функции, зависящие от взанимного расположения точек  $\mathbf{X}_{i_1},\dots,\mathbf{X}_{i_N}$ , причем  $\sum_{k=1}^{N_i}\gamma_{i_k}=1$ . В частности, можно взять

$$\gamma_{i_k} = \frac{\frac{1}{\rho^p(\mathbf{X}_i, \mathbf{X}_{i_k})}}{\sum_{k=1}^{N_i} \rho^p(\mathbf{X}_i, \mathbf{X}_{i_k})},$$
(14.31)

р — параметр.

Потребуем, чтобы для параметра R выполнялось условие

$$R > \operatorname{maxmin} \rho^p(\mathbf{X}_i, \mathbf{X}_j),$$

которое гарантирует, что все значения  $Y_i$  будут сглажены с заданным уровнем сглаживания.

Повторяя сглаживание сглаженных значений по формуле (14.30), получаем итерационный процесс

$$Y_i^l = \beta Y_i^{l-1} + (1 - \beta) \sum_{k=1}^{N_i} \gamma_{i_k} Y_{i_k}^{l-1}.$$
 (14.32)

Вектор сглаженных  $Y_1^l,\dots,Y_n^l$  значений на l-й итерации можно представить через значения  $Y_1,\dots,Y_n$  с помощью векторноматричного равенства

 $\mathbf{Y}^l = A^l \mathbf{Y},\tag{14.33}$ 

где элементы  $a_{ij}$  матрицы A даются равенствами

$$\mathbf{a}_{ii} = \boldsymbol{\beta}, \qquad j \neq i \quad a_{ij} = \begin{cases} 0, & \rho(\mathbf{X}_i, \mathbf{X}_j) > R, \\ (1 - \beta)\gamma_j, & \rho(\mathbf{X}_i, \mathbf{X}_j) \leqslant R. \end{cases}$$
(14.34)

Замечание 14.3. При таком подходе параметр сглаживания і является дискретным в отличие от параметра о при сглаживании методом метрического анализа (14.2). Поэтому в случае аппроксимации условие (14.25) для функционала (14.24) заменяется на следующее:

$$\delta(l) \to \min, \qquad \delta(l) = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{(Y_i^l - Y_i)^2}{\sigma_i^2} - 1\right)^2.$$
 (14.35)

Отметим еще раз, что интерполяция методом метрического анализа не предполагает задания базисной системы функций, а в каждой точке, где рассчитывается интерполяционное значение, индивидуально учитывается ее расположение относительно интерполяционных узлов. Аналогичное замечание относится к сглаживанию и аппроксимации методом метрического анализа. В одномерном случае m=1 другим методом сглаживания, не предполагающим априорного задания базисной системы функцией, является метод, основанный на сингулярно-спектральном анализе (см. § 3, гл. 12).

# 2. Оптимальность восстановленных значений функции на основе метрического анализа

Рассмотрим более подробно аспекты неопределенности при восстановлении значений функции с помощью метрического анализа.

Ранее уже отмечалось, что согласно принципу суперпозиции суммарная мера неопределенности  $\sigma_{sum}^2(Y/\mathbf{X}^*)$  значения Y в точке  $\mathbf{X}^*$  определяется равенством

$$\sigma_{sum}^2(Y/\mathbf{X}^*) = \sigma_Y^2(\mathbf{X}^*, \mathbf{z}/\mathbf{X}_1, \dots, \mathbf{X}_n) + \sigma^2(Y, \mathbf{z}), \tag{14.36}$$

которое запишем в виде

$$\sigma_{sum}^2(Y/\mathbf{X}^*) = (V_Y(\mathbf{X}^*/\mathbf{X}_1, \dots, \mathbf{X}_n)\mathbf{z}, \mathbf{z}), \tag{14.37}$$

где  $V_Y(\mathbf{X}^*/\mathbf{X}_1,...,\mathbf{X}_n) = W(\mathbf{X}^*/\mathbf{X}_1,...,\mathbf{X}_n) + \sigma^2 K_Y$  — симметричная неотрицательная  $(n \times n)$ -матрица.

Матрицу  $V_Y(\mathbf{X}^*/\mathbf{X}_1,\ldots,\mathbf{X}_n)$  из (14.37) будем называть матрицей суммарной неопределенности значения функции Y в точке  $\mathbf{X}^*$ .

Замечание 14.4. При решении прикладных задач определение меры неопределенности значения функции в точке  $\mathbf{X}^*$  согласно равенству (14.36) предполагает одинаковую размерность метрической  $\sigma_Y^2$  и стохастической  $\sigma^2(Y,\mathbf{z})$  мер неопределенности. Поэтому при решении прикладных задач необхо-

димо обезразмерить все переменные  $Y, X_1, \dots, X_n$  и провести их масштабирование на одинаковые интервалы измерения.

Замечание 14.5. При решении практических задач уровень стохастической погрешности  $\sigma^2$  значений Y может быть неизвестен. Тогда можно предложить схему оценки неизвестного  $\sigma^2$ , с использованием только известных значений функции  $Y_1, \ldots, Y_n$  в точках  $X_1, \ldots, X_n$  и элементов ковариационной матрицы  $K_Y$  погрешностей в значениях  $Y_1, \ldots, Y_n$ . Такая схема будет дана ниже. Сейчас же предполагается, что подходящее значение  $\sigma^2$  известно.

Рассмотрим обратную матрицу  $V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1,\ldots,\mathbf{X}_n)$ .

Величину  $I(Y/\mathbf{X}) = (V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)\mathbf{1},\mathbf{1})>0$  будем называть информацией о значении функции в точке  $\mathbf{X}^*$  по известным с погрешностями значениям функции  $Y_1,\dots,Y_n$  в точках  $\mathbf{X}_1,\dots,\mathbf{X}_n$ .

Замечание 14.6. В случае вырожденности матриц  $V_Y$ ,  $W_Y$ ,  $K_Y$  вместо обратных матриц используются псевдообратные или регуляризованные матрицы  $V_Y^-$ ,  $W_Y^-$ ,  $K_Y^-$  соответственно [7, 26, 77].

Из неравенства  $V_Y(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)>W(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)$  следует неравенство  $V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)< W_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)$  и. следовательно, неравенство  $(V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n),\mathbf{X}_n)$  ...  $\mathbf{X}_n$  1, 1)  $<(W_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n),\mathbf{X}_n)$  1, 1).

Таким образом, при наличии погрешности в значениях функции  $Y_1, \ldots, Y_n$  в точках  $\mathbf{X}_1, \ldots, \mathbf{X}_n$  итоговая информация о значении функции в точке  $\mathbf{X}^*$  меньше, чем информация о значении функции в точке  $\mathbf{X}^*$  в условиях отсутствия погрешностей в значениях  $Y_1, \ldots, Y_n$ .

Поставим задачу о нахождении значения Y функции в точке  $\mathbf{X}^*$ , для которого суммарная мера неопределенности (14.36) принимает минимальное значение.

Математическая модель указанной задачи имеет вид

$$(V_Y(\mathbf{X}^*/\mathbf{X}_1, \dots, \mathbf{X}_n)\mathbf{z}, \mathbf{z}) - \min_{\mathbf{z}}, \quad (\mathbf{z}, \mathbf{1}) = 1.$$
 (14.38)

Решение задачи (14.38) дается формулой

$$\mathbf{z}^* = \frac{V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1, \dots, \mathbf{X}_n) \cdot \mathbf{1}}{(V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1, \dots, \mathbf{X}_n) \cdot \mathbf{1}, \mathbf{1})}.$$
 (14.39)

Подставляя  $\mathbf{z} = \mathbf{z}^*$  в равенство  $Y^* = (\mathbf{z}, \mathbf{Y})$ , получаем искомое оптимальное восстановленное значение:

$$Y^* = \frac{(V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1, \dots, \mathbf{X}_n) \cdot \mathbf{1}, \mathbf{Y})}{(V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1, \dots, \mathbf{X}_n) \cdot \mathbf{1}, \mathbf{1})}.$$
 (14.40)

Очевидно, что суммарная мера неопределенности  $Y^*$  в точке  $\mathbf{X}^*$  равна

$$(V_Y(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)\mathbf{z}^*,\mathbf{z}^*) = \frac{1}{(V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)\cdot\mathbf{1},\mathbf{1})}.$$
 (14.41)

Из вышесказанного вытекают следующие теоремы.

**Теорема 14.2** (неравенство для суммарной неопределенности). Суммарная мера неопределенности любого восстановленного значения функции в точке по заданным значениям функции с погрешностями в точках  $\mathbf{X}_1,\ldots,\mathbf{X}_n$  удовлетворяет неравенству

 $\sigma_{sum}^2 \geqslant \frac{1}{I(Y/\mathbf{X}^*)},\tag{14.42}$ 

где  $I(Y/\mathbf{X}^*)$  — итоговая информация о значении функции в точке  $\mathbf{X}^*$  по известным (с погрешностями) значениям функции  $Y_1,\ldots,Y_n$  в точках  $\mathbf{X}_1,\ldots,\mathbf{X}_n$ .

Определение. Восстановленное значение функции  $Y^*$  в точке  $X^*$  при наличии погрешностей в значениях функции  $Y_1, \ldots, Y_n$  в точках  $X_1, \ldots, X_n$ , даваемое равенством (14.40), будем называть оптимальным, если для его суммарной меры неопределенности в неравенстве (14.42) достигается равенство.

**Теорема 14.3** (об оптимальности восстановленного значения). Восстановленное значение функции  $Y^*$  в точке  $X^*$ , даваемое равенством (14.40), оптимально.

Доказательство. Рассмотрим вырожденный случай, когда все точки  $\mathbf{X}_1,\dots,\mathbf{X}_n$  совпадают с точкой  $\mathbf{X}^*$ . В этом случае матрица метрической неопределенности  $W_Y(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)$  нулевая. Следовательно, псевдообратная матрица  $W_Y^+(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)$  также нулевая. Поэтому мера метрической неопределенности для значения любой функции в точке  $\mathbf{X}^*$  равна нулю.

Пусть  $\varepsilon$  распределена по нормальному закону и  $K_Y$  не вырождена. Докажем, то в этом случае оптимальное восстановленное значение (14.40) совпадает с оценкой максимального правдоподобия значения функции Y в точке  $\mathbf{X}^*$ .

Действительно, имеем:  $MY_i = Y$ ,  $i = 1, \ldots, n$ , где Y — неизвестное точное значение функции в точке  $\mathbf{X}^*$ . Функция правдоподобия  $L(\mathbf{Y},Y)$  для вектора  $\mathbf{Y}=(Y_1,\ldots,Y_n)^T$  имеет вид

$$L(\mathbf{Y}, Y) = \frac{1}{(2\pi)^{n/2} |K_Y|^{1/2}} \times \exp\left\{-\frac{1}{2} (K_Y^{-1}(\mathbf{Y} - Y \cdot 1), (\mathbf{Y} - Y \cdot 1))\right\}.$$

Следовательно, ММП-оценка  $Y_{\text{ММП}}$  искомого значения Y определяется равенством

$$Y_{\text{MMII}} = \arg\min(K_Y^{-1}(\mathbf{Y} - Y \cdot \mathbf{1}), (\mathbf{Y} - Y \cdot \mathbf{1})),$$

откуда

$$Y_{\text{MM}\Pi} = \frac{(K_Y^{-1}\mathbf{1}, \mathbf{Y})}{(K_Y^{-1}\mathbf{1}, \mathbf{1})}.$$

Но в рассматриваемом случае  $V_Y^{-1}(\mathbf{X}^*/\mathbf{X}_1,\dots,\mathbf{X}_n)=K_Y^{-1},$  т.е.  $Y_{\mathrm{MM\Pi}}=Y^*$  из (14.40), что и требовалось доказать.

Дадим теперь схему выбора подходящего значения  $\sigma^2$ , использующую только известные значения функции  $Y_1,\dots,Y_n$  в точках  $\mathbf{X}_1,\dots,\mathbf{X}_n$  и элементов матрицы  $K_Y$ , определяющей характеристики погрешностей в значениях  $\mathbf{Y}=(Y_1,\dots,Y_n)^T$ . Схема выбора подходящего значения  $\sigma^2$  основана на методе наилучшего восстановления и предполагает последовательное восстановление Y в точках  $\mathbf{X}_i$  с помощью совокупности оставшихся значений  $Y_1,\dots,Y_{i-1},Y_{i+1},\dots,Y_n$  в точках  $\mathbf{X}_1,\dots,\mathbf{X}_{i-1},\mathbf{X}_{i+1},\dots$ 

Восстановленные значения  $\widehat{Y}_i$  в точках  $\mathbf{X}_i$  определяются равенством

$$\widehat{Y}_{i} = \frac{(V_{Y}^{-1}(\mathbf{X}_{i}/\mathbf{X}_{1}, \dots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_{n})\mathbf{1}, \mathbf{Y}_{i})}{(V_{Y}^{-1}(\mathbf{X}_{i}/\mathbf{X}_{1}, \dots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_{n})\mathbf{1}, \mathbf{1})},$$
(14.43)

где  $V_Y^{-1}(\mathbf{X}_i/\mathbf{X}_1,\dots,\mathbf{X}_{i-1},\mathbf{X}_{i+1},\dots,\mathbf{X}_n)$  — обратная  $(n-1)\times (n-1)$ -матрица неопределенности  $V_Y(\mathbf{X}_i/\mathbf{X}_1,\dots,\mathbf{X}_{i-1},\mathbf{X}_{i+1},\dots,\mathbf{X}_n)$  восстановленного значения функции Y в точке  $X_i$  относительно совокупности точек  $\mathbf{X}_1,\dots,\mathbf{X}_{i-1},\mathbf{X}_{i+1},\dots,\mathbf{X}_n,1=(1,\dots,1)^T,\ \mathbf{Y}_i=(Y_1,\dots,Y_{i-1},Y_{i+1},\dots,Y_n)^T-(n-1)$ -мерные векторы.

Затем подсчитывается среднеквадратичное отклонение восстановленных значений от реализованных по всей совокупности точек  $\mathbf{X}_1, \dots, \mathbf{X}_n$  согласно формуле

$$\Delta Y(\sigma^2) = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2.$$
 (14.44)

Подходящее значение  $\sigma^2$  находится как решение экстремальной задачи

 $\sigma^2 = \arg\min \triangle Y(\sigma^2). \tag{14.45}$ 

Замечание 14.7. До сих пор предполагалось, что в восстановлении с помощью равенства (14.40) участвуют п точек  $\mathbf{X}_1,\ldots,\mathbf{X}_n$ , в которых известны (с погрешностями) значения исследуемой функции. Однако, особенно при решении практических задач, целесообразно проводить локализацию совокупности точек, по которым восстанавливается значение функции в рассматриваемой точке  $\mathbf{X}^*$ . При реализации схемы локализации подбирается подходящее число R>0, и в формуле (14.40) участвуют лишь те точки  $\mathbf{X}_1,\ldots,\mathbf{X}_n$ , которые попали внутрь шара  $\|\mathbf{X}-\mathbf{X}^*\|_{\overline{w}} < R$ . В частности, если не найдется ни одной точки из совокупности реализованных  $\mathbf{X}_1,\ldots,\mathbf{X}_n$ , попавших внутрь вышеуказанного шара, то это означает, что неопределенность восстановления в точке  $\mathbf{X}^*$  по совокупности точек  $\mathbf{X}_1,\ldots,\mathbf{X}_n$  имеет слишком большое значение и погрешность восстановления неприемлемо велика.

Выше была приведена схема подбора весов  $w_1, \dots, w_n$  метрики. Отметим, что локализацию можно использовать и при подборе весов метрики, причем в этом случае локализация будет особенно эффективна, если исследуемая функция имеет разные степени чувствительности от аргументов в разных частях исследуемой области изменения аргументов. В этом случае подходящая локализация при определении весов метрики обеспечивает большую точность интерполирования или восстановления функции.

После определения метрических весов  $\mathbf{w}$  определим матрицу функционально-метрической неопределенности точки  $\mathbf{X}^*$  относительно известных значений функции в совокупности точек  $\mathbf{X}_1, \dots, \mathbf{X}_n$  равенством

$$W_Y(X^*/\mathbf{X}_1,\dots,\mathbf{X}_n)=W, \qquad (14.46)$$

где W дается равенством (13.3).

Обозначим через  $W_Y$  псевдообратную матрицу матрицы (14.46). Величину  $I_Y(\mathbf{X}/\mathbf{X}_1,\dots,\mathbf{X}_n)=(W_Y\mathbf{1},\mathbf{1})=\sum_{i=1}^n\sum_{j=1}^nW_Y>0$  будем называть метрической информацией о значении Y в точке  $\mathbf{X}^*$  относительно совокупности точек  $\mathbf{X}_1,\dots,\mathbf{X}_n$ .

Метрическая информация о значении функции Y в точке  $X^*$  относительно совокупности точек  $X_1, \ldots, X_n$  содержит в себе информацию не только о геометрическом расположении совокупности точек  $X_1, \ldots, X_n$  относительно точки  $X^*$ , но и о степени зависимости исследуемой функции от аргументов с включением этой степени зависимости в метрическое расположение точек  $X_1, \ldots, X_n$  относительно точки  $X^*$  путем изменения метрики.

Если  $I_Y(\mathbf{X}/\mathbf{X}_1,\dots,\mathbf{X}_n) > I(\mathbf{X}/\mathbf{X}_1,\dots,\mathbf{X}_n)$ , где  $I(\mathbf{X}/\mathbf{X}_1,\dots,\mathbf{X}_n)$  — метрическая информация в безвесовой метрике, то это означает, что степень влияния аргументов на изменение функции различно, причем расположение точек таково, что более геометрически близкие к  $\mathbf{X}^*$  точки совокупности  $\mathbf{X}_1,\dots,\mathbf{X}_n$  более близки относительно координат соответствующих аргументов с наибольшим влиянием на изменение исследуемой функции.

Наоборот, если  $I_Y(\mathbf{X}/\mathbf{X}_1,\dots,\mathbf{X}_n) < I(\mathbf{X}/\mathbf{X}_1,\dots,\mathbf{X}_n)$ , то это означает увеличение в новой метрике расстояния точек  $\mathbf{X}_1,\dots,\mathbf{X}_n$  от точки  $\mathbf{X}^*$  для наиболее близких к  $\mathbf{X}^*$  точек в геометрической безвесовой метрике.

Таким образом, вышеуказанная схема выбора весовой метрики дает возможность выявить степень влияния каждого из аргументов и учесть эту различную степень влияния путем перехода к новой метрике с соответствующими неравными весами.

Рассмотрим вырожденный случай, когда в исследуемой функции введен аргумент (фактор), от которого функция не зависит. Тогда при реализации схемы нахождения весов новой метрики будет получен однозначный результат: метрический вес, соответствующий вышеуказанному фактору, будет равен нулю, и этот фактор будет автоматически исключен из дальнейшего рассмотрения. Вышеприведенная схема перехода к метрике с весами дает возможность учесть, подобно тому как это делается в факторном анализе, влияние факторов на изменение функции и исключить несущественные из них, понижая размерность факторного пространства.

Пример 14.2. Рассматривается зависимость  $Y = f(\mathbf{x}) + \varepsilon = (V\mathbf{x}.\mathbf{x}) + (\mathbf{c}.\mathbf{X}^*) + \varepsilon$ , где  $\mathbf{x} = (x_1, \dots, x_n)^T$ ,  $0 \le x_i \le 1$ ,  $i = 1, \dots, 5$ , V — симметричная матрица,  $\mathbf{c}$  — заданный вектор.  $\varepsilon$  — хаотическая компонента. В таблице 1 представлены резуль-

таты восстановления значений Y при следующих условиях:  $n=2,\, \varepsilon$  — белый шум с  $\varepsilon\sim N[0;0.1]$ , точки  $\mathbf{X}^*$  брались в разных частях 5-мерного куба  $0\leqslant x_i\leqslant 1,\, i=1,\ldots,5$ .

Таблица 1

						таолица
ET/AITE	THE CHORE	Вектор Х	MISTORY MISTORY	TANY S	Y	Результат восстанов- ления
0,321004	0,660444	0,133273	1,221744	0,725291	20,9909	20,6793
0,553684	0,972302	0,854523	0,175905	0,511616	37,0047	36,3486
0,791003	0,928066	0,146925	0,118069	0,007696	11,4883	13.0046
0,844583	0,971661	0,167349	0,562888	0,992063	52,7896	52,0912
0,511118	0,152113	0,088374	0,886334	0,968449	33,3384	35,6585
0,506079	0,48184	0,710547	0,405388	0,851228	39,9653	39,0228
0,146236	0,497067	0,539012	0,882125	0,954199	43,7182	43,7485
0,264033	0,087588	0,690458	0,125688	0,639964	17,2629	15,5257
0,309556	0,170704	0,759311	0,939503	0,726604	39,9813	40,5395
0,740922	0,029829	0,409211	0,016721	0,880375	20,6636	21,5067
0,237767	0,860013	0,878114	0,266829	0,051345	19,5284	19,3066
0,707865	0,047810	0,639003	0,376524	0,747099	30,04	30,0895
0,885714	0,705842	0,3656	0,115566	0,778709	33,728	34,1422
0,548071	0,592355	0,534014	0,972202	0,231008	34,4034	34,6933
0,797738	0,279911	0,018476	0,950551	0,434507	27,4361	28,6641
0,828089	0,597407	0,957641	0,079108	0,4444	32,2177	32,2573
0.717324	0,604474	0,74687	0,714585	0,54289	46,5764	44,8653
0,593894	0,035818	0,160847	0,675622	0,525737	20,266	19,7185
0,741381	0,674717	0,866045	0,032282	0,25448	23,7496	24,0375

### Глава 15

# АЛГОРИТМЫ И ВЫЧИСЛИТЕЛЬНЫЕ СХЕМЫ ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ МЕТРИЧЕСКОГО АНАЛИЗА

Рассмотрим функциональную зависимость y=f(x), заданную на равномерной сетке  $[x_1,\ldots,x_N]$ ,  $x_i=x_1+(i-1)h$ ,  $y_i=f(x_i)$ ,  $i=1,\ldots,N$ . Пусть требуется найти значение функции  $y_{N+1}=f(x_{N+1})$  в точке  $x_{N+1}=x_N+h$ . Задачу нахождения экстраполированного значения  $y_{N+1}$  сведем к задаче многомерной интерполяции с помощью нелинейной авторегрессии:

где F — неизвестная m-мерная функция, m — параметр авторегрессионной модели.

В таком виде одномерная функция y = f(x) представлена в виде m-мерной функции, заданной в N-m точках:

$$\mathbf{X}_{1} = (y_{1}, \dots, y_{m}),$$
 $\mathbf{X}_{2} = (y_{2}, \dots, y_{m+1}),$ 
 $\dots$ 

$$\mathbf{X}_{N-m} = (y_{N-m}, \dots, y_{N-1}).$$
(15.2)

Экстраполированное значение  $\tilde{y}_{N+1}$  находится как интерполяционное значение m-мерной функции F в точке  $\mathbf{X}^* = (y_{N-m+1}, \dots, y_N)$ :

 $\widetilde{y}_{N+1} = F(\mathbf{X}^*). \tag{15.3}$ 

Используя различные методы многомерной интерполяции, можно получить различные экстраполированные значения  $\tilde{y}_{N+1}$ .

Рассмотрим многомерную интерполяцию методом метрического анализа с использованием матрицы метрической неопределенности (13.3). Из представления (13.3), с учетом (15.2) получим следующее представление для матрицы метрической неопределенности через значения функции  $y_1, \ldots, y_N$ :

$$W = Y \operatorname{diag}(\omega_1, \dots, \omega_m) Y^T, \tag{15.4}$$

где

$$Y = \begin{pmatrix} y_1 - y_{N-m+1} & y_2 - y_{N-m+2} & \dots & y_m - y_N \\ y_2 - y_{N-m+1} & y_3 - y_{N-m+2} & \dots & y_{m+1} - y_N \\ \dots & \dots & \dots & \dots \\ y_{N-m} - y_{N-m+1} & y_{N-m+1} - y_{N-m+2} & \dots & y_{N-1} - y_N \end{pmatrix}.$$
(15.5)

При  $N-m\leqslant m\Rightarrow m\geqslant N/2$  матрица W невырождена и интерполяционные веса  $\mathbf{z}=(z_1,\ldots,z_{N-m})$  можно найти по формуле (13.10) или по формуле (14.39). После нахождения интерполяционных весов экстраполированное значение  $\widetilde{y}_{N+1}$  найдем по формуле

$$\tilde{y}_{N+1} = z_1 y_{m+1} + z_2 y_{m+2} + \dots + z_{N-m} y_N.$$
 (15.6)

Параметр m играет важную роль в этом методе экстраполяции, так как от его выбора зависит, каким будет результат экстраполяции, а априори он неизвестен. Для выбора параметра m поступим следующим образом. Найдем экстраполированные значения  $\widetilde{y}_N^{(m)}$  как интерполяционные значения m-мерной функции F при всевозможных  $m=2,\ldots,N-2$  в точках  $\mathbf{X}^*=(y_{N-m},\ldots,y_{N-1})$ , имея N-m-1 значений  $y_{k+1},\ldots,y_{N-1}$  функции F в N-m-1 точках

соответственно.

Тогда

$$\widetilde{y}_N^{(m)} = F^*(\mathbf{X}_1, \dots, \mathbf{X}_{N-m-1}, \widetilde{\mathbf{Y}}; \widetilde{\mathbf{X}}^*), \tag{15.8}$$

где  $\widetilde{\mathbf{Y}}=(y_{m+1},\dots,y_{N-1})^T,\; \mathbf{Y}_i=y_{m+i}+F(\mathbf{X}_i).$  Так как значение  $y_N$  известно, то определим m как

$$m = \arg\min(\widehat{y}_N^{(m)} - y_N). \tag{15.9}$$

После определения параметра m с помощью формулы (15.6) находим экстраполированное значение  $y_{N+1}$ . Далее для ряда  $(y_2,\ldots,y_N,\tilde{y}_{N+1})$  длины N с тем же параметром m нелинейной авторегрессионной модели строится матрица метрической неопределенности (15.4), для нее находится вектор интерполяци-

онных весов  $\mathbf{z}=(z_1,\dots,z_{N-m})$  и вычисляется экстраполированное значение  $\widetilde{y}_{N+2}$ :

$$\tilde{y}_{N+1} = z_1 y_{m+2} + \dots + z_{N-m-1} y_N + z_{N-m} y_{N+1}$$

Продолжая этот процесс, можно найти экстраполированные значения на любое количество шагов вперед.

Экстраполяция методом метрического анализа является достаточно устойчивой и точной, а также однозначно восстанавливает ряды для класса функций, управляемых линейными рекуррентными формулами (ЛРФ).

Для случая, когда значения функции  $y_1,\ldots,y_N$  зашумлены, можно предложить два подхода для их прогнозирования. В первом подходе одномерный ряд  $y_1,\ldots,y_N$  сглаживается методом метрического анализа и далее применяется описанный выше метод экстраполяции с помощью метрического анализа. Во втором подходе экстраполированное значение  $\tilde{y}_{N+1}$  находится не как интерполяционное значение, а как сглаженное значение m-мерной функции F в точке  $\mathbf{X}^* = y_{N-m-1},\ldots,y_N$ , найденное методом метрического анализа.

Определение. Говорят, что ряд  $F_N = (y_1, \dots, y_N)$  имеет конечно-разностную размерность, не большую чем d, и обозначают  $f\dim(F_N) \leqslant d$ , если  $1 \leqslant d < N-1$  и существуют  $a_1, \dots, a_d$ , такие что

$$y_{d+i} = a_1 y_i + a_2 y_{i+1} + \dots + a_d, \quad i = 1, \dots, N - d, \quad a_1 \neq 0;$$
(15.10)

 $d = \min\{k : f\dim(F_N) \leq k\}$  — называется размерностью ряда и обозначается  $f\dim(F_N) = d$ .

Формула (15.10) называется ЛРФ. ЛРФ с  $d = f \dim(F_N)$  называется минимальной ЛРФ. Если для ряда  $F_N$  выполняется

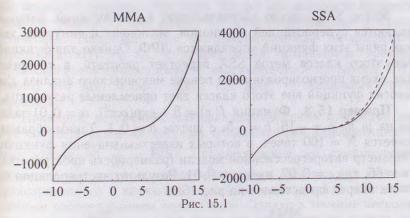
(15.10), то говорят, что ряд  $F_N$  управляется ЛРФ.

Еще одним методом, позволяющим однозначно восстанавливать ряды, управляемые ЛРФ, является метод SSA, который определяет коэффициенты ЛРФ, и по этим коэффициентам продолжает ряд с помощью представления (15.10). Однако этот метод является менее устойчивым и точным, а для функций, выходящих из класса управляемых ЛРФ, теряет точность экстраполяции намного раньше, чем метод экстраполяции на основе метрического анализа.

Продемонстрируем результаты прогнозирования на конкретных примерах и сравним результаты экстраполяции методом метрического анализа и методом SSA. Для сравнения в качестве

экстраполированного значения методом SSA будем брать наилучший прогноз из рекурсивного SSA- и векторного SSA-прогноза.

**Пример 15.1.** Рассмотрим функцию  $y(x) = (x-2)(x-1) \times$  $\times (x+1)$ , заданную на интервале [a,b], a=-10, b=-5, с шагом h = 0, 1. Таким образом, имеется N = 50 точек, в которых известны значения функции. Результат экстраполяции на  $N_{ext} = 200$  точек вперед представлен на рис. 15.1. Слева представлен результат экстраполяции методом метрического анализа, справа — методом SSA. Сплошной линией представлены точные значения функции, штриховой — экстраполированные. Точные значения функции в точках  $x_{N+1},\ldots,x_{N+N_{ext}}$  представлены для наглядности сравнения с экстраполированными значениями.



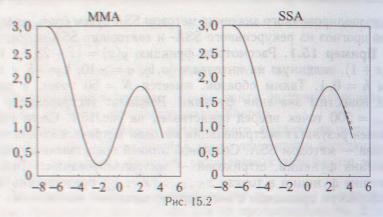
Параметр авторегрессионной модели (размерность пространства) m=35. Найдем точку  $x_i, \ i=N+1,\dots,N+N_{ext}$ , для которой величина отклонения  $|y_{ext}(x_i)-y(i)|$  максимальна, и подсчитаем в этой точке относительные отклонения для метода SSA и метода метрического анализа:

$$\varepsilon = \left| \frac{y_{ext}(x_i) - y(i)}{y(i)} \right|, \tag{15.11}$$

 $\varepsilon_{SSA} = 0,34, \, \varepsilon_{MMA} = 0,01.$ 

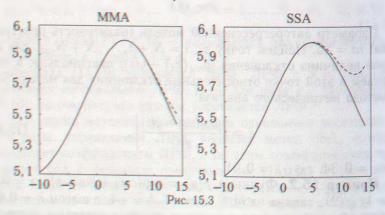
Пример 15.2. Функция  $f(x) = \exp(\alpha x) + \sin(\omega x)$ ,  $\alpha = 0, 1$ ,  $\omega = 1/\sqrt{(2)}$ , задана на [a,b], a=-7, b=-4, с шагом h=0,05. Таким образом, имеется c точек, в которых известны значения функции.

Параметр авторегрессионной модели (размерность пространства) m=43,  $arepsilon_{SSA}=0.2$ ,  $arepsilon_{MMA}=9,5\cdot 10^{-4}$ . Результат экстраполяции на  $N_{ext}=165$  точек вперед представлен на рис. 15.2.



Метод SSA хорошо экстраполирует класс функций, которые являются суперпозицией полиномов, экспонент и синусоид, так как ряды этих функций управляются ЛРФ. Однако для функций вне этого класса метод SSA перестает работать, в то время как схема прогнозирования на основе метрического анализа для многих функций вне этого класса дает приемлемые результаты.

Пример 15.3. Функция  $f(x)=5+\exp(\alpha x^2)$ ,  $\alpha=0,01$  задана на [a,b], a=-15, b=-5, с шагом h=0,1. Таким образом, имеется N=100 точек, в которых известны значения функции. Параметр авторегрессионной модели (размерность пространства) m=66.  $\varepsilon_{SSA}=0.07$ ,  $\varepsilon_{MMA}=0.01$ . Результат экстраполяции на точек вперед представлен на рис. 15.3.



### Глава 16

## ДИНАМИЧЕСКИЙ ДЕТЕРМИНИРОВАННЫЙ ХАОС

Одной из возможных причин появления неопределенности при прогнозировании динамических процессов является так называемый динамический детерминированный хаос. До недавнего времени считалось, что все динамические процессы делятся только на два класса — детерминированные и стохастические. Например, для детерминированных динамических процессов  $\mathbf{x}(t), t \geqslant t_0$ , с непрерывным изменением времени предполагается существование эволюционного оператора U(t), такого что  $\mathbf{x}(t) = U(t) \cdot \mathbf{x}(t_0)$ , где  $\mathbf{x}(t_0)$  — начальная точка процесса.

Примером детерминированных динамических процессов могут служить решения системы обыкновенных дифференциальных уравнений  $\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x})$ . Стохастические же динамические процессы  $\mathbf{y}(t)$  с непрерывным изменением времени для каждого фиксированного t являются случайной величиной с некоторым данным распределением вероятности. Однако в течение нескольких последних десятилетий было осознано, что из класса детерминированных процессов целесообразно выделить подкласс процессов, которые по своим проявлениям напоминают стохастические и при рассмотрении их реализаций не могут быть отличимы по многим их свойствам от стохастических процессов. Такие детерминированные динамические процессы принято называть процессами детерминированного хаоса [55, 61, 62, 94, 105].

### 1. Примеры процессов детерминированного хаоса

Рассмотрим систему обыкновенных дифференциальных уравнений

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}), t \geqslant t_0, \ \mathbf{x} = (x_1, \dots, x_M)^T.$$
 (16.1)

При определенных достаточно общих условиях каждое решение  $\mathbf{x}(t)$  системы (16.1) однозначно определяется на всем проме-

жутке определения решения  $t \in [t_0, t_{0+T}]$  с начальным значением  $\mathbf{x}_0 = \mathbf{x}(t_0)$ . Это означает, что, зная точное значение  $\mathbf{x}_0$ , мы можем однозначно определить (предсказать) значение  $\mathbf{x}(t, \mathbf{x}_0)$  для  $t \ge t_0$ .

Однако на практике точное значение  $\mathbf{x}_0$  никогда не известно. В случае же неточного задания  $\mathbf{x}_0$  возможны варианты такого поведения  $\mathbf{x}(t,\mathbf{x}_0)$ , при котором близкие в начальный момент  $t_0$  траектории  $\mathbf{x}(t,\mathbf{x}_0)$ ,  $\mathbf{x}(t,\mathbf{x}_0+\boldsymbol{\delta})$ , где  $\|\boldsymbol{\delta}\|$  мала, с течением времени «разбегаются» и прогноз даже для сравнительно небольших времен  $t>t_0$  имеет неприемлемо большую погрешность, как и в случае стохастических временных процессов.

Исторически первой прикладной задачей, в которой имел место детерминированный хаотический процесс, явилась знаменитая задача трех тел в небесной механике. При попытке решения этой задачи с помощью разложения в ряды теории возмущений А. Пуанкаре обнаружил, что эти ряды расходятся. В дальнейшем стало ясно, что расходимость этих рядов является проявлением детерминированного хаоса, присутствующего в задаче трех тел.

Другим широко известным примером проявления детерминированного хаоса является турбулентное течение жидкости. Правда, при исследовании турбулентности необходимо рассматривать более сложную по сравнению с ОДУ систему нелинейных уравнений в частных производных. Отметим, что наличие детерминированного хаоса для систем линейных дифференциальных уравнений невозможно. Более того, детерминированные хаотические процессы не могут наблюдаться для системы обыкновенных дифференциальных уравнений вида (16.1) с порядками M=1,2. Таким образом, минимальный порядок системы ОДУ (16.1), в которой может наблюдаться детерминированный хаос, M=3.

Детерминированным хаотическим динамическим процессам как решениям системы ОДУ (16.1) соответствуют так называемые странные аттракторы. Аттрактором принято называть притягивающий установившийся режим. Примерами аттракторов для ОДУ вида (16.1) являются устойчивые по Ляпунову стационарные решения (решение системы алгебраических уравнений  $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ ).

Если M = 1, то единственным типом аттрактора может быть асимптотическое устойчивое стационарное решение.

Если M=2, то возможно наличие еще одного типа аттрактора — предельного цикла, которому соответствует орбитально устойчивое периодическое решение ОДУ.

Если M=2, то возможно наличие аттракторов только двух типов — асимптотически устойчивых по Ляпунову стационарных

решений и предельных циклов. Эти два типа аттракторов не порождают динамического хаоса.

Если  $M\geqslant 3$ , то возможно наличие странного аттрактора, представляющего собой притягивающее множество, каждая траектория которого представляет собой геометрический образ динамического хаотического временного процесса.

Впервые пример странного аттрактора был представлен в связи с системой трех дифференциальных уравнений Лоренца:

$$\begin{cases}
\frac{dx}{dt} = \sigma \cdot (y - x); \\
\frac{dy}{dt} = r \cdot x - y - x \cdot z; \\
\frac{dz}{dt} = -b \cdot z + x \cdot y,
\end{cases} (17.2)$$

где  $\sigma, r, b$  — параметры системы.

Система уравнений (16.2) описывает несколько физических процессов, связанных с течением жидкости и работой лазера.

Лоренц показал, что при некоторых значениях параметров системы (16.2) появляется странный аттрактор и тем самым детерминированный хаос.

Например, при значениях  $\sigma=10, r=28, b=8/3$ , взятых Лоренцем, система (16.2) имеет странный аттрактор и динамические процессы x(t), y(t), z(t) изменяются нерегулярным образом, проявляя хаотическую структуру.

К настоящему времени сконструированы десятки систем обыкновенных дифференциальных уравнений третьего порядка, имеющих странные аттракторы и тем самым динамические хаотические процессы.

Наряду с детерминированным эволюционными системами с непрерывным изменением времени t рассмотрим детерминированные динамические процессы с дискретным изменением времени. Такие процессы могут задаваться с помощью соотношения

$$\mathbf{x}_{n+1} = \mathbf{F}(x_n), \tag{16.3}$$

где  $n=0,1,2,\ldots$  — нумерация дискретной фиксации времени через определенные временные шаги,  $\mathbf{x}_n=(x_{n1},\ldots,x_{nM})^T$ .

Оказывается, дискретные модели (16.3) могут порождать детерминированный хаос даже для случая размерности M=1. Впервые факт наличия хаоса в одномерных дискретных моделях  $x_{n+1}=F(x_n)$  был обнаружен А.Н. Шарковским и подтвержден затем в работе [122].

**Пример 16.1.** Логистический дискретный процесс задается равенством

$$x_{n+1} = \alpha \cdot (x_n - x_n^2), \tag{17.4}$$

где  $\alpha$  — параметр логистической модели.

В работе [122] показано, что при  $\alpha=4$  путем выбора начального значения  $x_0$  можно получить любую наперед заданную смену знаков для смещенной последовательности  $x_n-c$ , где c- константа, не зависящая от n.

### Пример 17.2. Дискретный процесс Лемера

$$x_{n+1} = D(c \cdot x), \tag{16.5}$$

где c>0 — фиксированная константа, D(x) — дробная часть числа.

Напомним, что дискретный процесс Лемера используется в качестве датчика псевдослучайных величин при достаточно большом c>0.

Использование детерминированных хаотических процессов в качестве датчиков случайных величин в настоящее время нашло широкое применение во многих областях прикладного анализа и, в частности, при обработке неопределенных данных.

# 2. Свойства детерминированных хаотических процессов

Рассмотрим решения  $\mathbf{x}(t,\mathbf{x}_0,t_0),\mathbf{x}(t_0,\mathbf{x}_0,t_0)=\mathbf{x}_0$  системы дифференциальных уравнений (16.1). Пусть траектория решения  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  отлична от станционарного решения  $\mathbf{x}(t,\mathbf{x}_0,t_0)\equiv\mathbf{x}_0$ . Рассмотрим  $\varepsilon$ -окрестность  $||\mathbf{x}-\mathbf{x}_0||\leqslant \varepsilon$  точки  $\mathbf{x}_0$ . Пусть траектория решения  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  возвращается через время T в  $\varepsilon$ -окрестность точки  $\mathbf{x}_0$ . Если решение  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  таково, что  $\mathbf{x}_0$  из окрестности, принадлежащей аттрактору — предельному циклу, то T может зависеть от  $\varepsilon$  и  $T(\varepsilon) \to T^*$  при  $\varepsilon \to 0$ , где  $T^*$  — период предельного цикла.

Обозначим через  $t_1(\varepsilon), t_2(\varepsilon), \dots$  времена последовательного возврата Пуанкаре траектории  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  в  $\varepsilon$ -окрестность точки  $\mathbf{x}_0$ . Если  $t_1(\varepsilon)=t_2(\varepsilon)=\dots=T(\varepsilon)$ , но  $T(\varepsilon)\to+\infty$  при  $\varepsilon\to0$ , то решение  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  — квазипериодическая функция. Примером квазипериодической функции может служить линейная суперпозиция нескольких периодических функций, отношения периодов которых — иррациональные числа.

Пусть теперь решение  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  таково, что времена возвратов  $t_1(\varepsilon),t_2(\varepsilon),\ldots$  не равны и их числовая последовательность имеет характер выборки некоторой случайной величины. В этом случае решению  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  соответствует детерминированная динамическая хаотическая траектория, принадлежащая странному аттрактору.

Выясним теперь связь между странными аттракторами и устойчивостью по Ляпунову решений системы (16.1). Для ответа на вопрос об устойчивости по Ляпунову решения  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  системы (16.1) необходимо перейти к линейной системе относительно возмущений  $\delta\mathbf{x}=\mathbf{x}(t,\mathbf{x}_0+\delta,t_0)-\mathbf{x}(t,\mathbf{x}_0,t_0)$ , где  $\delta>0$  — малое число,

$$\frac{d\delta \mathbf{x}}{dt} = A(t)\delta \mathbf{x},\tag{16.6}$$

 $A(t)=rac{\partial \mathbf{F}}{\partial \mathbf{x}}|_{\mathbf{x}=\mathbf{x}(t,\mathbf{x}_0,t_0)}$  — матрица Якоби для исследуемого решения  $\mathbf{x}(t,\mathbf{x}_0,t_0).$ 

Для произвольного фиксированного решения  $\delta {f x}$  линейной системы (16.6) определим ляпуновский характеристический по-казатель  $\Lambda(\delta {f x})$  равенством

$$\Lambda(\delta \mathbf{x}) = \overline{\lim_{t \to +\infty}} \frac{1}{t} \ln ||\delta \mathbf{x}||, \tag{16.7}$$

где lim — верхний предел.

При достаточно общих условиях для каждого решения  $\delta x$  системы (16.6) существует ограниченный по модулю ляпуновский характеристический показатель. Существует набор из M характеристических показателей, такой что для каждого решения его характеристический показатель совпадает с одним из показателей набора.

Пусть  $\mathbf{x}_1(t),\ldots,\mathbf{x}_M(t)$  — такая фундаментальная система решений уравнения (17.6), для которой получаем весь набор из M характеристических показателей,  $\Lambda_j = \Lambda(\mathbf{x}_j(t)), j=1,\ldots,M$  — ляпуновские характеристические показатели решений, образующих фундаментальную систему. Не умаляя общности, предположим, что показатели  $\Lambda_j, j=1,\ldots,M$ , ранжированы, т. е.

$$\Lambda_1 \geqslant \Lambda_2 \geqslant \dots \geqslant \Lambda_M. \tag{16.8}$$

Набор ляпуновских характеристических показателей (16.7) называют спектром ляпуновских показателей системы (16.6) или решения  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  системы (16.1). Показатель  $\Lambda_1$  называют

старшим характеристическим показателем системы (16.7) или решения  $\mathbf{x}(t, \mathbf{x}_0, t_0)$  системы (16.1).

Для вычисления всех характеристических показателей спектра можно использовать эволюционную матрицу системы (16.6).

Решение задачи Коши для системы (16.6) дается равенством

$$\delta \mathbf{x}(t) = V(t) \cdot \delta \mathbf{x}_0, \tag{17.9}$$

где V(t) — эволюционная матрица.

Подставляя (16.9) в (16.7), имеем

$$\Lambda(\delta \mathbf{x}) = \overline{\lim}_{t \to +\infty} \frac{1}{2t} \ln(V^T(t)V(t)\delta \mathbf{x}_0, \delta \mathbf{x}_0). \tag{16.10}$$

Замечание 16.1. В (16.7) можно брать любую норму M-мерного фазового пространства. В дальнейшем предполагается, что взята евклидова норма  $\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$ .

Следовательно, характеристические показатели спектра определяются через положительные собственные значения положительно определенной симметричной матрицы  $W(t) = V^T(t)V(t)$ :

$$\Lambda_j = \overline{\lim}_{t \to +\infty} \frac{1}{t} \cdot \ln \mu_j(t), j = 1, \dots, M, \tag{16.11}$$

rme 
$$W(t)\cdot \Psi_j(t)=\mu_j(t)\cdot \Psi_j(t), j=1,\ldots,M.$$

Замечание 16.2. Числовые алгоритмы расчета эволюционной матрицы V(t) должны предусматривать периодическую нормировку решений из-за возможности быстрого их роста в силу наличия положительного характеристического показателя.

Напомним, что согласно теореме А.М. Ляпунова об устойчивости по первому линейному приближению исследуемое решение  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  будет асимптотически устойчиво, если старший характеристический показатель  $\Lambda_1$  отрицателен. Наоборот, решение  $\mathbf{x}(t,\mathbf{x}_0,t_0)$  будет неустойчиво, если старший показатель  $\Lambda_1$  положителен.

Как переносится свойство устойчивости по Ляпунову решений системы (16.1) на свойства аттракторов? Если аттрактор является стационарной точкой или предельным циклом, то он образован одним решением, ассимптотическая устойчивость этого решения и обеспечивает существование аттрактора.

Если аттрактор образован совокупностью разных решений, то согласно теореме В.И.Оселедеца [70] почти все решения  $\mathbf{x}(t, \mathbf{x}_0, t_0)$ , образующие аттрактор, имеют один и тот же спектр характеристических показателей, который тем самым характери-

зует сам аттрактор в целом.

Отметим, что сумма всех характеристических показателей Ляпунова  $\sum\limits_{i=1}^{M} \Lambda_i < 0$  определяет степень сжатия M-мерного фазового пространства аттрактора.

Спектр характеристических показателей аттрактора должен удовлетворять некоторым общим условиям.

- 1. Если аттрактор не является стационарной точкой, то нуль принадлежит спектру, поскольку для возмущений вдоль траектории системы (16.1), образующей аттрактор, характеристический показатель равен нулю.
- 2. Поскольку аттрактор ограниченное множество в фазовом пространстве траекторий, то сумма всех характеристиче-  $\frac{M}{}$

ских показателей должна быть отрицательной: 
$$\sum\limits_{i=1}^{M}\Lambda_{i}<0.$$

Дополнительно для странного аттрактора, порождающего детерминированный хаос, по крайней мере один характеристический показатель должен быть положительным.

Последнее требование для странного аттрактора означает, что для автономных детерминированных дифференциальных уравнений первого и второго порядков странные аттракторы не могут существовать.

Рассмотрим теперь дискретный детерминированный процесс  $\{x_n\}$ , задаваемый многошаговым рекуррентным равенством

$$x_{n+1} = F(x_{n-M+1}, x_{n-M+2}, \dots, x_n).$$
 (16.12)

Число M называется размерностью (порядком) модели (16.12).

Для того чтобы получить конкретный временной ряд, порождаемый моделью (16.12), необходимо задать начальный вектор  $\mathbf{x}_0 = (x_{-M+1}, x_{-M+2}, \dots, x_0)^T$  размерности M.

Для исследования характеристических показателей Ляпунова решения  $(x_1,x_2,\ldots)$  уравнения (16.12) с начальным положением  $\mathbf{x}_0$  рассмотрим другое близкое по начальному положению решение  $(y_1,y_2,\ldots)$  с начальным вектором  $\mathbf{y}_0=(y_{-M+1},y_{-M+2},\ldots)$ 

 $(x_0, y_0)^T, ||\mathbf{x}_0 - \mathbf{y}_0|| \le \delta.$ Введем векторы  $\mathbf{x}_n = (x_{n-M+1}, \dots, x_n)^T, \mathbf{y}_n = (y_{n-M+1}, \dots, y_n)^T$ 

 $(x_n,y_n)^T$ ,  $\delta x_n = y_n - x_n$ . Линеаризованная модель относительно возмущения  $\delta x_n$  имеет вид

$$\delta x_{n+1} = \frac{\partial F(x_{n-M+1}, \dots, x_n)}{\partial x_{n-M+1}} \cdot \delta x_{n-M+1} + \dots$$

$$\dots + \frac{\partial F(x_{n-M+1}, \dots, x_n)}{\partial x_n} \cdot \delta x_n. \tag{16.13}$$

Определим характеристический показатель Ляпунова решения уравнения (16.13) равенством

$$\Lambda = \overline{\lim}_{n \to \infty} \frac{1}{n} \ln |\delta \mathbf{x}_n|. \tag{16.14}$$

Совокупность всех характеристических показателей Ляпунова временных рядов, порожденных моделью (16.9), образует спектр решения х модели (16.12) с начальным положением хомли, что то же самое, спектр линейной модели (16.13).

Существует фундаментальная система решений  $\{\delta \mathbf{x}^{(j)}(t)\}$ ,  $j=1,\ldots,M$  дискретного линейного уравнения (16.13), такая что совокупность характеристических показателей Ляпунова решений, образующих фундаментальную систему, совпадает со спектром линейной модели (16.13).

Не умаляя общности, будем считать, что характеристические показатели ранжированы в порядке убывания:

$$\Lambda_1 \geqslant \Lambda_2 \geqslant \ldots \geqslant \Lambda_M.$$
 (16.15)

Показатель  $\Lambda_1$  называется старшим характеристическим показателем.

Если рассматривается аттрактор дискретной модели (16.12). то аналогично непрерывной модели (16.1) с вероятностью единиа любой временной ряд, порождающий рассматриваемый аттрактор, имеет один и тот же спектр характеристических показателей Ляпунова. Единый спектр временных рядов, порождающих аттрактор, называется спектром характеристических показателей Ляпунова этого аттрактора.

Проще всего найти старший характеристический показатель  $\Lambda_1$ , поскольку с вероятностью единица в каждом решении будет присутствовать аддитивная компонента, динамика которой определяется показателем  $\Lambda_1$ , и поэтому использование формулы (16.14), примененной к произвольному решению уравнения (16.13), будет определять на практике старший показатель  $\Lambda_1$ .

Для нахождения всего спектра характеристических показателей можно аналогично непрерывной модели (16.6) использовать эволюционную матрицу.

Из (16.13) следует, что

$$\delta \mathbf{x}_{n+1} = A(\mathbf{x}_n) \cdot \delta \mathbf{x}_n, \tag{16.16}$$

где эволюционная (M imes M)-матрица определяется равенством

$$A(\mathbf{x}_n) = \begin{pmatrix} \frac{\partial F(x_{n-2M+2}, \dots, x_{n-M+1})}{\partial x_{n-2M+2}} & \dots & \frac{\partial F(x_{n-2M+2}, \dots, x_{n-M+1})}{\partial x_{n-M+1}} \\ \dots & \dots & \dots \\ \frac{\partial F(x_{n-M+1}, \dots, x_n)}{\partial x_{n-M+1}} & \dots & \frac{\partial F(x_{n-2M+2}, \dots, x_{n-M+1})}{\partial x_n} \end{pmatrix}.$$

Очевидно равенство (16.16) эквивалентно равенству  $\delta \mathbf{x}_{n+1} = A(\mathbf{x}_n) \cdot A(\mathbf{x}_{n-1}) \cdot \ldots \cdot A(\mathbf{x}_0) \cdot \delta \mathbf{x}_0$ .

Рассмотрим выражение

$$\overline{\lim_{n \to +\infty}} \frac{1}{n} \ln ||\delta \mathbf{x}_n||, \tag{16.17}$$

для которого справедливо равенство

$$\overline{\lim}_{n\to+\infty} \frac{1}{n} \ln ||\delta \mathbf{x}_n|| = \overline{\lim}_{n\to\infty} \frac{1}{n} \ln |\delta x_n| = \Lambda.$$

**Замечание 16.3.** В выражении (16.17) можно брать любую норму M-мерного пространства. В дальнейшем предполагается, что  $||\mathbf{x}|| = \sqrt{(\mathbf{x}, \mathbf{x})}$  — эвклидова норма.

Из вышесказанного следует, что характеристические показатели Ляпунова решения  $(x_1,x_2,\ldots)$  с начальным положением  $\mathbf{x}_0$  можно находить из равенства

$$\Lambda = \overline{\lim}_{n \to +\infty} \frac{1}{2n} \ln(\delta \mathbf{x}_n, \delta \mathbf{x}_n),$$

или с учетом определения матриц  $A(\mathbf{x}_n)$  из равенства

$$\Lambda = \overline{\lim}_{n \to \infty} \frac{1}{2n} \ln(V_n \delta \mathbf{x}_0, \delta \mathbf{x}_0),$$

где симметричная положительно определенная (M imes M)-матрица  $V_n$  определяется равенством

$$V_n = A^T(\mathbf{x}_0) \cdot A^T(\mathbf{x}_1) \cdot \ldots \cdot A^T(\mathbf{x}_{n-1}) \cdot A(\mathbf{x}_{n-1}) \cdot \ldots \cdot A(\mathbf{x}_1) \cdot A(\mathbf{x}_0).$$

Весь спектр характеристических показателей Ляпунова определяется M положительными собственными значениями матрицы  $V_n$ :

$$V_n \overrightarrow{\Psi}_{nj} = \mu_{nj} \overrightarrow{\Psi}_{nj}, j = 1, \dots, M,$$
 (16.18)

где  $\{\overrightarrow{\Psi}_{nj}\},\ j=1,\ldots,M$  — ортонормированная система собственных векторов матрицы  $V_n$ .

Следовательно, спектр характеристических показателей Ля-пунова можно найти, используя равенство

$$\Lambda_{j} = \overline{\lim}_{n \to \infty} \frac{1}{n} \ln \mu_{nj}, j = 1, \dots, M.$$
 (16.19)

Замечание 16.4. Для расчета характеристических показателей Ляпунова можно использовать схему итерационной ортогонализации Бенеттина [55, 108].

Ниже даны примеры расчета показателей Ляпунова для различных аттракторов непрерывных моделей вида (16.1) и дискретных моделей вида (16.3).

**Пример 16.3.** Рассмотрим систему уравнений Лоренца (16.2). Система уравнений (16.6) относительно возмущений имеет вид

$$\frac{d\delta x}{dt} = \sigma \cdot (\delta y - \delta x),$$

$$\frac{d\delta y}{dt} = r \cdot \delta x - \delta y - z \cdot \delta x - x \cdot \delta z,$$

$$\frac{d\delta z}{dt} = -b \cdot \delta z + y \cdot \delta x + x \cdot \delta y,$$
(16.20)

где  $\mathbf{x} = (x, y, z)^T$  — решение, порождающее странный аттрактор. Решаем задачу Коши для системы (16.2) со значениями параметров  $\sigma = 10, r = 28, b = 8/3$ , соответствующими странному аттрактору.

Затем подставляем полученное решение в (16.20) и для нахождения эволюционной матрицы решаем последовательно три задачи Коши с начальными условиями  $\delta \mathbf{x}^{(1)}(0) = (1,0,0)^T$ .  $\delta \mathbf{x}^{(2)}(0) = (0,1,0)^T$ ,  $\delta \mathbf{x}^{(3)}(0) = (0,0,1)^T$  и используем алгоритм Бенеттина [55, 108].

Замечание 16.5. На самом деле для данного примера достаточно подсчитать только один старший положительный ляпуновский показатель  $\Lambda_1$ , поскольку еще один показатель равен нулю из-за автономности системы Лоренца, а сумма всех показателей Ляпунова равна  $-(\sigma+1+b)$  — сумме коэффициентов при диагональных элементах линейной части системы Лоренца.

Старший показатель Ляпунова  $\Lambda_1 = 0,897$ . Поэтому осталь-

ные показатели равны:  $\Lambda_2 = 0$ ,  $\Lambda_3 = -17,564$ .

Поскольку старший показатель Ляпунова положителен, а сумма показателей отрицательна, исследуемое решение имеет каотический характер и принадлежит странному аттрактору.

**Пример 16.4.** Рассмотрим непрерывную автономную модель Ресслера

$$\frac{dx}{dt} = -y - z, 
\frac{dy}{dt} = x + ay + v, 
\frac{dz}{dt} = b + xz, 
\frac{dv}{dt} = cz - dv.$$
(16.21)

Пусть параметры модели a,b,c,d принимают значения: a=0,25,b=3,c=0,5,d=0,05.

В этом случае система (16.21) имеет странный аттрактор. Система уравнений (16.6) относительно возмущений имеет вид

$$\frac{d\delta x}{dt} = -\delta y - \delta z, 
\frac{d\delta y}{dt} = \delta x + a\delta y + \delta v, 
\frac{d\delta z}{dt} = z\delta x + x\delta z, 
\frac{d\delta v}{dt} = c\delta z - d\delta v.$$
(16.22)

Решаем задачу Коши для системы (16.21). Затем подставляем решение в (16.22) и решаем последовательно 4 задачи Коши для линейной системы (16.22) с начальными условиями  $\delta \mathbf{x}^{(1)} = (1,0,0,0)^T, \delta \mathbf{x}^{(2)} = (0,1,0,0)^T, \delta \mathbf{x}^{(3)} = (0,0,1,0)^T, \delta \mathbf{x}^{(4)} = (0,0,0,1)^T$ , формируем эволюционную матрицу, находим 4 ее сингулярных числа и используем формулу (16.11).

Четыре характеристических показателя Ляпунова, соответствующие странному аттрактору, определяются равенствами

$$\Lambda_1 = 0,313; \Lambda_2 = 0,052; \Lambda_3 = 0; \Lambda_4 = -23,35; \sum_{j=1}^4 \Lambda_j < 0.$$

Таким образом, система Ресслера (16.21) имеет странный аттрактор с двумя положительными показателями и определяет так называемый гиперхаос.

Пример 16.5. Рассмотрим одномерный дискретный динамический процесс, порожденный логистической моделью

$$x_{n+1} = 1 - 2 \cdot x_n^2. \tag{16.23}$$

Уравнение относительно возмущений имеет вид

$$\delta x_{n+1} = -4x_n \cdot \delta x_n. \tag{16.24}$$

Можно показать, что

$$\delta x_n = 2^n \cdot \sqrt{\frac{1 - x_n^2}{1 - x_0^2}} \cdot \delta x_0.$$

Следовательно, характеристический показатель Ляпунова находится из равенств

$$\Lambda = \overline{\lim_{n \to +\infty}} \frac{1}{n} \ln |\delta x_n| = \ln 2 + \overline{\lim_{n \to +\infty}} \frac{|\delta x_0|}{n} + \overline{\lim_{n \to +\infty}} \frac{1}{2n} \ln \left| \frac{1 - x_n^2}{1 - x_0^2} \right| = \ln 2 > 0.$$

В последнем равенстве учтен тот факт, что временной ряд  $\{x_n\}$  образует аттрактор и тем самым ограничен.

Странные аттракторы обладают еще одним замечательным

свойством - фрактальной структурой.

Пусть задана ограниченная область G в M-мерном евклидовом пространстве. Покроем область G сеткой M-мерных кубиков со стороной длины  $\varepsilon>0$  и подсчитаем минимальное количество кубиков  $N(\varepsilon)$ , полностью покрывающее область G.

По определению фрактальной размерностью (емкостью) d об-

ласти С является величина

$$d = -\lim_{\varepsilon \to +0} \frac{\log N(\varepsilon)}{\log \varepsilon}.$$
 (16.25)

Например. если рассматривается одномерный случай, т.е. G — кривая, то для гладкой кривой d=1. В общем случае для гладкой M-мерной области d=M, т.е. фрактальная размерность области совпадает с геометрической размерностью. Однако в M-мерных пространствах существуют области, у которых фрактальная размерность больше геометрической размерности M и является дробным числом. Например, если рассмотреть линию  $\mathbf{r}(t)$ , являющуюся графиком белого шума, то ее фрактальная размерность равна 1.5, т.е. превышает ее геометрическую размерность в 1.5 раза. Такое превышение фрактальной размерностью линии ее геометрической размерности есть следствие изрезанной структуры вплоть до нулевого масштаба. Оказывается, что странные аттракторы имеют дробную фрактальную размерность и по этой характеристике подобны случайным процессам.

Для аттракторов, порожденных двумерными дискретными

моделями, справедливо соотношение

$$d = m + \sum_{i=1}^{m} \Lambda_i / |\Lambda_{m+1}|, \qquad (16.26)$$

где m < M такое, что  $\Lambda_m > 0$ , а  $\Lambda_{m+1} < 0$ .

Соотношение (16.26) называют формулой Каплана-Йорка.

В общем случае для странных аттракторов, порожденных моделями (16.1) и дискретными моделями (16.3) и (16.12) формула (16.26) неверна, хотя и дает, как правило, хорошее приближение фрактальной размерности через спектр показателей Ляпунова аттрактора.

В настоящее время фрактальную размерность d странного аттрактора, определяемую равенством (16.26), принято называть размерностью Ляпунова, а сумму  $\sum\limits_{i=1}^m \Lambda_i$  всех положительных характеристических показателей аттрактора, энтропией Колмогорова—Синая.

Кроме вышеуказанной фрактальной размерности, используют и другие характеристики фрактальности, в частности, учитывающие частоту «посещения» траекторией разных подобластей аттрактора.

Для того чтобы построить такие характеристики, снова рассмотрим сетку из M-мерных кубиков M-мерного фазового пространства  $\mathbf{x}$ , покрывающую исследуемую ограниченную область странного аттрактора. Обозначим через  $p_i$  вероятность того, что траектория принадлежит кубику с номером i.

Введем характеристики  $d_q$ , называемые обобщенными раз-

мерностями, с помощью равенств

$$d_q = -\lim_{\varepsilon \to +0} \frac{\log \sum_i p_i^q}{\ln \varepsilon},\tag{16.27}$$

где  $q = 0, 1, 2, \dots$ 

Очевидно, что  $d_0$  — фрактальная размерность (емкость), определяемая равенством (16.25). Величина  $d_1$  называется информационной размерностью, а  $d_2$  — корелляционным показателем Грассбергера—Прокачиа.

Рассмотрим теперь обратную задачу: пусть дан ограниченный временной ряд  $\{x_n\}, n=0,1,\dots$  Можно ли ответить на вопрос, — является ли временной ряд значениями через равные интервалы решения некоторого автономного дифференциального уравнения?

Более точно, будем говорить, что временной ряд  $\{x_n\}$  может быть воспроизведен гладкой детерминированной моделью, если существует автономное дифференциальное уравнение (16.1) с гладкой функцией  $F(\mathbf{x})$  и гладкая функция  $g(\mathbf{x})$ , такие что  $\mathbf{x}_n = g(\mathbf{x}(n \cdot \Delta t))$ , где  $\mathbf{x}(t)$  — решение задачи Коши системы (16.1).

В работе Ф. Такенса [122] была сформулирована и доказана следующая

**Теорема 16.1.** Временной ряд  $\{x_n\}$  может быть воспроизведен гладкой детерминированной моделью, если  $\ln c_{n,\varepsilon}/(n-\ln\varepsilon)$  равномерно ограничена при  $n-\ln\varepsilon\to\infty$ . В противном случае временной ряд  $\{x_n\}$  не может быть воспроизведен гладкой детерминированной моделью. Здесь для  $\varepsilon>0$  и натурального п  $c_{n,\varepsilon}$  определено как число элементов множества  $R_{n,\varepsilon}$ , а множества натуральных чисел  $R_{n,\varepsilon}$  определены правилом:  $i\in R_{n,\varepsilon}$  только в том случае, если для всех  $0< j< i, j\in R_{n,\varepsilon}$ . выполняется неравенство

$$max\{|x_i-x_j|, |x_{i+1}-x_{j+1}|, \ldots, |x_{i+n}-x_{j+n}|\} \geqslant \varepsilon.$$

Пусть исследуемый хаотический временной ряд  $\{x_n\}$  порожден детерминированной моделью.

Поставим теперь вопрос о прогнозировании значений временного ряда. Если нам известна детерминированная модель, то, принимая последнее реализованное значение  $x_n$  за начальное, можно вычислить  $x_{n+1}$ , которое и принимается в качестве прогнозируемого. Этот процесс можно продолжить и на последующие шаги. Однако в силу неустойчивости по Ляпунову сколь угодно малое отклонение от точного значения через несколько шагов приведет к неуправляемой ошибке в прогнозировании. Тем не менее возможность прогнозирования с приемлемой точностью на один и несколько (небольшое число) шагов достаточно привлекательна. Если временной ряд  $\{x_n\}$  порожден нелинейной дискретной моделью (16.12), но модель неизвестна, то одной из схем прогнозирования может быть следующая.

Пусть  $x_n$  — последнее реализованное значение дискретного временного процесса. На основе реализованной части временного ряда  $x_i$  : = 1.... n-M. с помощью какой-либо интерполяционной схемы оценим функциональную зависимость  $x_{n+1} = F(x_{n-M+1}, \dots, x_n)$  а затем по последним реализованным значениям  $x_{n-M+1}, \dots, x_n$  по формуле (16.12) вычислим  $\overline{x}_{n+1}$ , которое и принимается за прогнозируемое значение. Оценку

размерности M системы можно осуществить на реализованной части временного ряда, оценив функцию F на первой части ряда, прогнозируя значения временного ряда на второй части ряда и сравнивая прогнозируемые значения с реализованными. Подходящее значение M можно выбирать, используя принцип минимума суммы квадратов невязок между прогнозируемыми и реализованными значениями временного ряда.

В качестве интерполяционный схемы можно брать схемы, использующие нейронные сети или схемы, основанные на метри-

ческом анализе (см. гл. 13).

#### Глава 17

# 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 взвещенная МНК-оценка (7.22), где  $(n\times m)$ -матрица A определяется равенством

$$A^T = (\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n)),$$

 $\varphi(x_i) = (\varphi_1(x_i), \dots, \varphi_m(x_i))^T$  — 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  $x_i$ ,  $i=1,\dots,n$ , фиксации независимой переменной x, то  $K_{\widehat{u}}$  также зависит от  $x_i$ ,  $i=1,\dots,n$ . Таким образом. В зависимости от того, при каких значениях  $x_i$ ,  $i=1,\dots,n$ , проведены измерения, будем получать различные по точности МНК-оценки. Из вышесказанного естественным образом возникает следующая задача. Пусть исследователь имеет возможность разместить  $x_i$ ,  $i=1,\dots,n$ , 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.

## 1. Постановка задач планирования оптимальных измерений

В настоящей главе 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  $K_{\varepsilon}=\mathrm{diag}\,(\sigma_1^2,\dots,\sigma_n^2)$ . Обозначим  $W_i=1/\sigma_i^2,\,i=1,\dots$ , n. Тогда  $K_{\varepsilon}^{-1}=\mathrm{diag}\,(W_1,\dots,W_n)$ . Числа  $W_i$  называют весами. Для рассматриваемого случая МНК-оценку (7.22) можно представить в виде

 $\widehat{u}_{\mathsf{MHK}} = M^{-1}Y,$ 

гле

$$M = A^T K_{\varepsilon}^{-1} A = \sum_{i=1}^{n} W_i \varphi(x_i) \varphi^T(x_i) = \sum_{i=1}^{n} M_i,$$
$$Y = A^T K_{\varepsilon}^{-1} f = \sum_{i=1}^{n} W_i y_i \varphi(x_i)$$

— 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  $\widehat{y}_{\mathrm{MHK}}=\varphi^T(x)\widehat{u}_{\mathrm{MHK}}$  с наименьшей дисперсией  $\sigma^2(\widehat{y}_{\mathrm{MHK}})=\varphi^T(x)M^{-1}\varphi(x)\equiv d(x).$  Функцию  $\sqrt{d(x)}$  называют коридором ошибок.

Часто в одной точке  $x=x_i$  проводится несколько независимых измерений  $y_{i1},\dots,y_{i\tau_i}$  функции y, каждое с дисперсией  $\sigma_i^2$ .

Тогда всю совокупность измерений  $y_{ik},\ k=1,\dots,r_i,$  можно заменить одним

 $y_i = \frac{1}{r_i} \sum_{k=1}^{r_i} y_{ik}$ 

c becom  $W_i = r_i/\sigma_i^2$ .

В большинстве прикладных задач дисперсия  $\sigma_i^2$  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  $\sigma_0^2$ , т. е.  $\sigma^2(x)=\sigma_0^2\tilde{\sigma}^2(x)$ , то функцию  $\tilde{\lambda}(x)=1/\tilde{\sigma}^2(x)$  назовем функцией относительной эффективности измерений.

Очевидно, что  $\widehat{u}_{\rm MHK} = M^{-1}Y$ , 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

$$s_0^2 = \frac{1}{n-m} \sum_{i=1}^n \widetilde{\lambda}(x_i) (y_i - \varphi^T(x_i) \widehat{u}_{\text{MHK}})^2$$

является несмещенной и состоятельной оценкой неизвестной  $\sigma_0^2$ . Пусть измерения проводятся при значениях  $x_1,\dots,x_n$ , причем теперь в точке  $x_i$  проводится  $r_i$  измерений. Тогда  $\sum_{i=1}^n r_i = N$  — 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  $p_i=r_i/N$ ,  $\sum_{i=1}^n p_i=1$ , называют нормированным планом (эксперимента) и обозначают символом  $\varepsilon(N)$ .

Совокупность точек  $x_1, \dots, x_n$  называется спектром плана эксперимента.

Поскольку информационная матрица M зависит от плана эксперимента  $\mathcal E$  (или  $\varepsilon$ ), в дальнейшем часто будем использовать явную запись этой зависимости  $M=M(\mathcal E)$  (или  $M(\varepsilon)$ ).

Имеем:  $M(\varepsilon)=N\sum_{i=1}^n p_i\widetilde{\lambda}(x_i)\varphi(x_i)\varphi^T(x_i)$ . В дальнейшем мат-

рицу  $M^{-1}$ , обратную информационной, будем обозначать  $\mathcal{D}(\mathcal{E})$  (или  $\mathcal{D}(\varepsilon)$ ).

Два плана

$$\mathcal{E}_1 = \{x_1^{(1)}, \dots, x_{n_1}^{(1)}; \quad r_1^{(1)}, \dots, r_{n_1}^{(1)}\},$$

$$\mathcal{E}_2 = \{x_1^{(2)}, \dots, x_{n_2}^{(2)}; \quad r_1^{(2)}, \dots, r_{n_2}^{(2)}\}$$

называют одинаковыми тогда и только тогда, когда  $n_1=n_2=n$  и

$$x_{(i)}^{(1)} = x_{(i)}^{(2)}, \quad r_{(i)}^{(1)} = r_{(i)}^{(2)}, \quad i = 1, \dots, n,$$

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 различ-

ные планы, для которых выполнены равенства

$$n_1 = n_2 = n$$
,  $\sum_{i=1}^{n} r_i^{(1)} = \sum_{i=1}^{n} r_i^{(2)} = N$ .

Аналогично вводится понятие различия двух нормированных планов.

Различные планы можно сравнивать друг с другом с точки зрения качества MHK-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 сравне-

ния 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  $\sigma^2(\widehat{y}_{\mathrm{MHK}}(x))$ , 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 18.1.** Пусть восстанавливается функциональная зависимость  $y = u_1 + u_2 x + u_3 x^2$  и имеются два плана эксперимента  $\mathcal{E}_1 = \{-1.0; 4, 4, 4\}, \ \mathcal{E}_2 = \{-1,0; 3,6,3\}.$  Имеем

$$\mathcal{D}(\mathcal{E}_1) = \frac{1}{12} \begin{pmatrix} 3 & 0 & -3 \\ 0 & 3/2 & 0 \\ -3 & 0 & 9/2 \end{pmatrix}, \quad \mathcal{D}(\mathcal{E}_2) = \frac{1}{12} \begin{pmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{pmatrix},$$

 $|\mathcal{D}(\mathcal{E}_1)| = 6.75 \cdot 12^{-3}, \quad |\mathcal{D}(\mathcal{E}_2)| = 8 \cdot 12^{-3}, \quad \operatorname{Sp} \mathcal{D}(\mathcal{E}_1) = 3/4, \\ \operatorname{Sp} \mathcal{D}(\mathcal{E}_2) = 2/3.$ 

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  $au(\mathcal{E})$  берут сумму  $k_B au_B(\mathcal{E})+k_M au_M(\mathcal{E})$ , 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:

$$L[\mathcal{D}(\mathcal{E})] = |\mathcal{D}(\mathcal{E})|, \qquad L[\mathcal{D}(\mathcal{E})] = \operatorname{Sp} \mathcal{D}(\mathcal{E}),$$

$$L[\mathcal{D}(\mathcal{E})] = \max_{j=1,\dots,m} \mathcal{D}_{jj}(\mathcal{E}),$$

$$L[\mathcal{D}(\mathcal{E})] = \max_{x \in X} d(x; \mathcal{E}) = \max_{x \in X} \varphi^{T}(x) \mathcal{D}\varphi(x),$$

$$L[\mathcal{D}(\mathcal{E})] = \int_{\mathcal{X}} d(x; \mathcal{E}) dx.$$

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

$$\mathcal{E}^* = \arg\min_{\tau(\mathcal{E}) \leqslant T} L[\mathcal{D}(\mathcal{E})], \tag{17.1}$$

rде 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

$$\mathcal{E}^* = \arg\min_{L[\mathcal{D}(\mathcal{E})] \leqslant L_0} \tau(\mathcal{E}), \tag{17.2}$$

где  $L_0$  задано.

Функционал затрат часто представим в виде  $\tau(\mathcal{E}) = \sum_{i=1}^{n} c_i r_i$ , где  $c_i > 0$  — стоимость одного измерения в точке  $x_i$ . В частности, если стоимость измерений не зависит от  $x_i$ , то  $c_i = c$ ,  $i = 1, \dots, n$ , и  $\tau(\mathcal{E}) = cN$ .

Экстремальные задачи с ограничением (17.1), (17.2) эквивалентны экстремальной задаче без ограничения

$$\mathcal{E}^* = \arg\min R(\mathcal{E}), \tag{17.3}$$

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 (17.1) эквивалентна задаче без ограничения

$$\mathcal{E}^* = \arg\min L[\mathcal{D}(\mathcal{E})]. \tag{17.4}$$

План  $\mathcal{E}^*$  из (17.4) называется  $\mathcal{D}$ -оптимальным, если  $L[\mathcal{D}(\mathcal{E})] = |\mathcal{D}(\mathcal{E})|$ ; А-оптимальным, если  $L[\mathcal{D}(\mathcal{E})] = \operatorname{Sp} \mathcal{D}(\mathcal{E})$ ; минимаксным в пространстве параметров, если  $L[\mathcal{D}(\mathcal{E})] = \max_{j=1,\dots,m} \mathcal{D}_j(\mathcal{E})$ ; 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 18.2. Пусть  $y=u_1+u_2x,\ x\in[-1;1],\ N=3,\ \lambda(x)=$  = 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  $-1\leqslant x_i\leqslant 1,\,i=1,2,3,\,\mathcal{E}_1^*=\{-1,1;2,1\}$ 

 $\mathsf{H} \ \mathcal{E}_{2}^{*} = \{-1,1;1,2\}.$ 

Если априори задана приемлемая погрешность оценки  $L_0$ , а N не фиксируется, то решение задачи (17.4) 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  $\varepsilon$  называется совокупность величин  $\{x_1,\ldots,x_n;p_1,\ldots,p_n\}$ , где  $0\leqslant p_i\leqslant 1,\;\sum_{i=1}^n p_i=1.$ 

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:

$$1^{\circ} |\mathcal{D}(\mathcal{E})| \approx \frac{1}{N_{\bullet}^{m}} |\mathcal{D}(\varepsilon)|;$$

$$2^{\circ} \operatorname{Sp} \mathcal{D}(\mathcal{E}) \approx \frac{1}{N} \operatorname{Sp} \mathcal{D}(\varepsilon);$$

$$3^{\circ} \max_{j=1,...,m} \mathcal{D}_{jj}(\mathcal{E}) \approx \frac{1}{N} \max_{j=1,...,m} \mathcal{D}_{jj}(\varepsilon);$$

$$4^{\circ} \max_{x \in X} d(x; \mathcal{E}) \approx \frac{1}{N} \max_{x \in X} d(x; \varepsilon);$$

$$5^{\circ} \int_{X} d(x; \mathcal{E}) dx \approx \frac{1}{N} \int_{X}^{\infty} d(x; \varepsilon) dx.$$

Таким образом, для случая  $au(\mathcal{E})=cN,\ N\gg n$  функционал потерь  $R(\mathcal{E})$  имеет вид

$$R(\mathcal{E}) \approx Nc + \frac{k_L}{N^{\beta}} L[\mathcal{D}(\varepsilon)],$$
 (17.5)

где eta=m для критерия  $L[\mathcal{D}(arepsilon)]=|\mathcal{D}(arepsilon)|$  и eta=1 для остальных

четырех критериев.

Поскольку  $L[\mathcal{D}(\varepsilon)]$  не зависит от N, решение экстремальной задачи (17.3), когда  $R(\mathcal{E})$  дается равенством (17.5), можно разбить на два этапа: на первом этапе находится оптимальный непрерывный нормированный план  $\varepsilon^* = \arg\min_i L[\mathcal{D}(\varepsilon)] = \{x_1^*, \ldots, x_n^*; p_1^*, \ldots, p_n^*\};$  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  $\mathcal{E}^* = \{x_1, \ldots, x_n^*\}$ , 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.

## 2. Методы решения задач планирования оптимальных измерений

В дальнейшем рассматриваются только непрерывные нормированные планы  $\varepsilon$ . Заметим, что информационная матрица  $M(\varepsilon)$  вырождена, если спектр плана  $\varepsilon$  содержит менее m точек. Поэтому количество точек n спектра любого разумного плана  $\varepsilon$  должно состоять не менее чем из m точек.

Естественно конструировать планы с наименьшим числом точек спектра.

**Теорема 18.1** (Де Ла Гарза). Для любого плана  $\varepsilon$  с числом точек спектра  $n \geqslant m$  всегда найдется такой план  $\widetilde{\varepsilon}$  с числом точек спектра m+1, что  $M(\overline{\varepsilon})=M(\varepsilon)$ .

Из теоремы 18.1 следует. в частности, что оптимальные планы  $\varepsilon^*$  можно искать в классе планов  $\varepsilon$  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  $\varepsilon$  называется невырожденным, если матрица плана  $M(\varepsilon)$  невырожденная.

Теорема 18.2. Пусть  $\varepsilon = \{x_1, \dots, x_n; p_1, \dots, p_n\}$  — невырожденный план. Тогда  $1^\circ \sum_{i=1}^n p_i \widetilde{\lambda}(x_i) d(x_i; \varepsilon) = m;$   $2^\circ \max_{x \in X} \overline{\lambda}(x) d(x; \varepsilon) \geqslant m$ . 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 18.3** (Вольфовиц-Кифер). Следующие утверждения эквивалентны:  $\varepsilon^*$  —  $\mathcal{D}$ -оптимальный план;  $\varepsilon^*$  минимизирует  $\max_{x\in X}\widetilde{\lambda}(x)d(x;\varepsilon);\ max_{x\in X}\widetilde{\lambda}(x)d(x;\varepsilon)=m.$ 

Теорему 18.3 принято называть теоремой эквивалентности. Из теоремы эквивалентности, в частности, следует, что при  $\widetilde{\lambda}(x)=\mathrm{const}~\mathcal{D}$ -оптимальный непрерывный план одновременно является минимаксным непрерывным планом, и наоборот.

**Теорема 18.4.** 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 18.4 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 18.3. Пусть  $y=u_1+u_2e^{-\alpha x},\ x\geqslant 0,\ \widehat{\lambda}(x)=e^{-\alpha x},\ \alpha>0.$  Найдем  $\mathcal{D}$ -оптимальный план  $\varepsilon_z^*$ . Сделаем замену переменной  $z=e^{-\alpha x}$ . Имеем:  $y=u_1+u_2x,\ \widehat{\lambda}(z)=z,\ 0\leqslant z\leqslant 1.$  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  $\varepsilon_*^*=\{1,1/3;1/2,1/2\}$ . Имеем:  $\max\lambda(z)d(z;\varepsilon_*^*)=2$ . Из теоремы эквивалентности следует, что  $\varepsilon_*^*-\mathcal{D}$ -оптимальный непрерывный план для  $y=u_1+u_2z,\ \lambda(z)=z,\ 0< z\leqslant 1$ . Тогда из теоремы 18.4 следует, что  $\varepsilon_*^*=\{0,-\ln 3;1/2,1/2\}-\mathcal{D}$ -оптимальный непрерывный план для восстановления функциональной зависимости  $y=u_1+u_1e^{-\alpha z},\ \lambda(x)=e^{-\alpha x},\ x\geqslant 0$ . Таким образом, если половину измерений делать при x=0, а вторую половину — при  $x=-\ln 3$ , то план  $\mathcal E$  будет  $\mathcal D$ -оптимальным.

Понск Д-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  $x_1,\dots,x_n,p_1,\dots,p_n,\sum_{i=1}^n p_i=1,\ 1\geqslant p_i\geqslant 0.$  Можно показать, что при n=m справедливо равенство

 $|M(arepsilon)| = \prod_i p_i \widetilde{\lambda}(x_i) |A|^2$ , right  $A = (\varphi(x_1), \dots, \varphi(x_m))$  — (m imes m)х т - матрица. Из последнего равенства следует, что максимальное значение  $M(\varepsilon)$  по  $p_1,\ldots,p_m$  достигается при  $p_i=1/m$ ,  $i = 1, \dots, m$ , независимо от  $x_1, \dots, x_m$ . 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 Д-оптимального плана сводится к максимизации по  $x_1,\ldots,x_m$  функционала  $|A|^2\prod\widetilde{\lambda}(x_i)$ .

Пусть  $\varphi_j(x) = x^{j-1}, \ j = 1, \dots, m$ , т. е. рассматривается задача восстановления полиномиальной функциональной зависимости  $y = \sum u_j x^{j-1}.$ 

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  $\lambda(x)$ .

Система непрерывных функций  $f_i(x)$ ,  $i=1,\ldots,m$ , определенная на промежутке X, называется чебышевской, если любая линейная комбинация этих функций  $\sum_{i=1}^{m} c_i f_i(x)$  ( $c_i$  действитель-

ные.  $\sum c_i > 0$ ) имеет не более m-1 корней на промежутке X.

Очевидно, что система  $f_i(x) = x^{i-1}$ ,  $i = 1, \dots, m$ , — чебышев-

**Теорема 18.5.** Пусть  $1, \lambda(x), \ldots, \lambda(x) x^{2(m-1)}$  — чебышевская система на промежутке Х. Тогда при восстановлении полиномиальной зависимости  $y = \sum_{j=1}^m u_j x^{j-1}$  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 полиномнальная зависимость остается полиномиальной (с сохранением степени полинома).

Теорема 18.6. Пусть рассматривается задача восстановления полиномиальной зависимости  $y=\sum_{j=1}^{\infty}u_{j}x^{j-1}$  для следующих четырех случаев:

 $1^{\circ} \lambda(x) = \text{const}, x \in [-1, 1];$ 

$$2^{\alpha} \tilde{\lambda}(x) = (1-x)^{\alpha} (1+x)^{\beta}, \ \alpha, \beta > -1, \ x \in [-1, 1];$$

$$3^{\alpha} \tilde{\lambda}(x) = x^{\alpha+1} e^{-x}, \ \alpha > -1, \ x \geqslant 0;$$

$$3^{\circ} \lambda(x) = x^{\alpha+1}e^{-x}, \ \alpha > -1, \ x \geqslant 0$$

$$4^{\circ} \lambda(x) = e^{-x^2}, x \in (-\infty, \infty).$$

Тогда для каждого из этих четырех случаев существует  $\mathcal{D}$ -оптимальный непрерывный план  $\varepsilon^* = \{x_1, \dots, x_m; 1/m, \dots$ ..., 1/m}, единственный в классе планов с n=m. Точки спектра Д-оптимальных планов являются корнями полиномов для случаев: 1.  $(1-x^2)\frac{dP_{m-1}}{dx}$ , где  $P_{m-1}$  — полином Лежандра степени m-1; 2.  $P_m^{(\alpha,\beta)}(x)$  — полином Якоби степени m с параметрами  $\alpha,\beta;$  3.  $L_m^{(\alpha)}(x)$  — полином Чебышева—Лагерра с параметром  $\alpha$ ; 4.  $H_m(x)$  — полином Чебышева-Эрмита.

Доказательство всех четырех случаев теоремы 18.6 аналогично. Рассмотрим, например, первый случай.

Имеем: 
$$|M(\varepsilon)| = \frac{1}{m} \, C |A|^2$$
, где  $|A|^2 = \prod_{k \neq l, \, k, l = 1}^m (x_k - x_l)^2$ ,

$$\ln |A|^2 = \sum_{k 
eq l,\, k,l=1}^m \ln (x_k - x_l)^2$$
. Поэтому задача максимизации

$$|M(arepsilon)|$$
 эквивалентна задаче максимизации  $\sum_{k 
eq l, \, k, l=1}^m \ln{(x_k-x_l)^2}.$ 

Для нахождения точек спектра имеем систему уравнений  $rac{\partial \ln |A|^2}{\partial x_k}, \ k=1,\ldots,m$ , 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

$$\frac{df(x)}{dx} = (x - x_2)(x - x_3) \cdot \dots \cdot (x - x_m) + \dots \\ \dots + (x - x_1)(x - x_2) \cdot \dots \cdot (x - x_{m-1}),$$

$$\frac{df(x_k)}{dx} = (x_k - x_1) \cdot \dots \cdot (x_k - x_{k-1})(x_k - x_{k+1}) \cdot \dots \cdot (x_k - x_m),$$

$$\frac{d^2 f(x)}{dx^2} = (x - x_3) \cdot \dots \cdot (x - x_m) + \dots + (x - x_1) \cdot \dots \cdot (x - x_{m-2}),$$

$$\frac{d^2 f(x_k)}{dx^2} = (x_k - x_2) \cdot \dots \cdot (x_k - x_{k-1})(x_k - x_{k+1}) \cdot \dots \cdot (x_k - x_m) + \dots + (x_k - x_1)(x_k - x_{k+1}) \cdot \dots \cdot (x_k - x_m) + \dots \cdot (x_k - x_m) + \dots + (x_k - x_1)(x_k - x_2) \cdot \dots \cdot (x_k - x_{m-1}),$$

$$\dots \cdot (x_k - x_{m-1})(x_k - x_{k+1}) \cdot \dots \cdot (x_k - x_{m-1}).$$

Поэтому

$$\frac{d^2 f(x_k)}{dx^2} / \frac{df(x_k)}{dx} = \frac{1}{x_k - x_1} + \frac{1}{x_k - x_2} + \dots + \frac{1}{x_k - x_{k-1}} + \dots + \frac{1}{x_k - x_{m-1}} + \dots + \frac{1}{x_k - x_m} = 0.$$

Из последнего равенства следует, что  $\frac{d^2f(x)}{dx^2}\,(1-x^2)=af(x)$ , где a — постоянная. Приравнивая в обеих частях последнего равенства коэффициенты при  $x^m$ , получаем a=-m(m+1)/2. Следовательно,

$$(1-x^2)\frac{d^2f(x)}{dx^2} + \frac{m(m+1)}{2}f = 0.$$

Но решением последнего уравнения является функция

$$f(x) = (1 - x^2) \frac{dP_{m-1}}{dx}$$
.

Теорема 18.6 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, ..., 2m + 1, неизвестны.

Пусть  $X=[0;2\pi],\ \widetilde{\lambda}(x)=\mathrm{const.}$  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  $\overline{\lambda}(x)=\mathrm{const}~\mathcal{D}$ -оптимальный план совпадает с минимаксным планом. Ниже приведены условия для совпадения планов, оптимальных по разным критериям.

**Теорема 18.7.** План  $\varepsilon^*$  является одновременно Q- и A-оптимальным, если  $\varphi_i(x)$ ,  $i=1,\ldots,m$ , — ортонормированная система на промежутке X, m. е.  $\int\limits_X \varphi(x) \varphi^T(x) \, dx = I_m$ .

Для полиномиальной зависимости с  $\lambda(x)=$  const доказано, что спектры  $\mathcal{D}$ - и Q-оптимальных планов совпадают (n=m), т.е. измерения необходимо проводить в нулях полинома  $(1-x^2)\frac{dP_{m-1}}{dx}$ , но веса для Q-оптимальных планов определяются равенствами

$$p_i = |P_m(x_i)|^{-1} / \sum_{j=1}^m |P_m(x_j)|^{-1}, \quad j = 1, \dots, m,$$

где  $P_m(x)$  — 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  $\varepsilon_0 = \{x_1, \dots, x_n; p_1, \dots, p_n\},$   $n \geqslant m$ . 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, \dots, x_n, x_0; (1 - \alpha_0)p_1, \dots, (1 - \alpha_0)p_n, \alpha_0\},\$$

а «шаг»  $\alpha_0$  выбирается так, чтобы увеличение  $|M(\varepsilon_1)|$  было максимальным.

Для такого  $\alpha_0$  справедливо равенство  $\alpha_0=\delta_0/((\delta_0+m-1)m)$ , где  $\delta_0=\lambda(x_0)d(x_0;\varepsilon_0)-m>0$ . После осуществления третьего этапа  $\alpha_0$  заменяют на  $\alpha_1$  и возвращаются к первому этапу.

Доказано, что последовательность планов  $\varepsilon_k$  сходится к  $\mathcal{D}$ -оптимальному плану  $\varepsilon^*$ .

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:  $1^\circ$  точки спектра с малыми весами, изолированные от других точек спектра, отбрасываются (с равномерным распределением веса отбрасываемых точек по множеству оставшихся, так чтобы выполня-

лось равенство  $\sum_{i=1}^{n} p_i = 1$ , где  $n_k$  — количество остающихся точек спектра);  $2^{\circ}$  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 [33, 91].

### Список литературы

- 1. Айвазян С.А. Основы эконометрики. М.: ЮНИТИ, 2001.
- 2. Айвазян С.А. Прикладная статистика М.: Алавар, 2010.
- 3. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: Юнити, 1998.
- 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. Альберт А. Регрессия, псевдоинверсия и рекурсивное оценивание. М.: Наука, 1977.
- 8. 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.
- 9. *Арсенин В. Я., Крянев А. В., Цупко-Ситников М. В.* Применение робастных методов при решении некорректных задач // ЖВММФ. 1989. Т. 29. № 5. С. 653–661.
- Арсенин В.Я., Крянев А.В. Обобщенный метод максимального правдоподобия решения конечномерных некорректных задач // ЖВММФ. 1991. Т. 31. № 5. С. 643–653.
- 11. 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.
- 12. *Березин И.С., Жидков Н.П.* Методы вычислений. Т. 1, 2. М.: Наука, 1994.
- 13. Бард Я. Нелинейное оценивание параметров. М.: Статистика, 1979.
- 14. 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.
- 15. Богданов Ю.И. Основная задача статистического анализа данных: корневой подход. М.: Изд-во МИЭТ, 2002.
- Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983.
- Вапник В.Н. Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979.
- 18. Вапник В.Н., Стефанюк А.Р. Непараметрические методы восстановления плотности вероятностей // АиТ. 1978. Т. 8. С. 38-52.
- Василенко В.А. Сплайн-функции: теория, алгоритмы, программы. Новосибирск: Наука, 1983.
- 20. Васильев Ф.П. Методы оптимизации. М.: Факториал Пресс, 2002.

- 21 Восильев Ф.П. Методы решения экстремальных задач. М.: Наука, 1981.
- 22 Вероятность и математическая статистика. Энциклопедия. Гл. ред. Ю.В. Прохоров. М.: Изд-во Большая Российская Энциклопедия, 1999.
- Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. М.: Наука. 1984.
- 24. Волков Н.Г., Дорогов В.И., Коньшин В.А., Крянев А.В. Комплексная программа восстановления регрессии и проведение оценки  $\sigma_f$ ,  $\sigma_t$ ,  $\alpha$  и  $\nu_f$  для  $U^{235}$ . Минск, Препринт ИЯЭ АН БССР, 1986.
- 25 Воскобойников Ю. Е., Преображенский Н.Г., Седельников А.И. Математическая обработка эксперимента в молекулярной газодинамике. — Новосибирск: Наука, 1984.
- 26. Гантмахер Ф.Р. Теория матриц. М.: Физматлит, 2004.
- Главные компоненты временных рядов: метод «Гусеница». В сб. под ред. Д. Данилова и А.А. Жиглявского. — СПб.: СПбГУ, 1997.
- Гончаров В.Л. Теория интерполирования и приближения функций. М.: Гостехиздат, 1954.
- Деврой Л., Дьерфи Л. Непараметрическое оценивание плотности.
   L<sub>1</sub>-подход. М.: Мир. 1988.
- 30. Де Гроот М. Оптимальные статистические решения. М.: Мир, 1974.
- Лемиденко Е.З. Линейная и нелинейная регрессии. М.: Финансы и статистика. 1981.
- 32 Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. М.: Мир, 1980, Т. 1; 1981, Т. 2.
- 33. Ермаков С.М., Жиглявский А.А. Математическая теория оптимального эксперимента. М.: Наука, 1987.
- 34 Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применения. М.: 2000.
- 35. Жуковский Е.Л., Мелешко В.И. О помехоустойчивых решениях в линейной условной алгебре // ДАН СССР. 1978. Т. 241. № 5. С. 1009–1012.
- 36 Жуковский Е.Л., Морозов В.А. О последовательной байесовской регуляризации алгебраических систем уравнений // ЖВММФ. 1972. Т. 12. № 2. С. 464–465.
- 37. Завыялов Ю.С., Квасов Б.И., Мирошниченко В.Л. Методы сплайн-функций. М.: Наука, 1980.
- 38. Закс Ш. Теория статистических выводов. М.: Мир, 1975.
- 39. Зрелов П.В., Иванов В.В. Интегральные непараметрические статистики  $\omega_n = n^{\frac{1}{2}} \int_{-\infty}^{\infty} |S_n(x) P(x)|^k dx$  и их основные свойства. Алгебраический вид. функции распределения и критерии согласия. Сообщения ОИЯИ.
- Р10-92-461. Дубна, 1992. 40. Иванов В.К., Васин В.В., Танана В.П. Теория линейных некорректных задач и ее приложения. — М.: Наука, 1978.
- 41. Калиткин Н.Н. Численные методы. М.: Наука, 1978.
- 42. Кендалл М. Временные ряды. М.: Финансы и статистика, 1981.
- Кендалл М. Дж., Стирарт А. Статистические выводы и связи. М.: Наука, 1973.
- 44. Кислицын М.М. Многомерная статистика временных рядов наблюдений в авиационной эргономике // Вопросы кибернетики. 1978. Т. 51.

- 45. Корнейчук Н.П. Сплайны в теории приближений. М.: Наука, 1984.
- 46. Крамер Г. Математические методы статистики. М.: Мир, 1975.
- 47. *Крянев А.В.* Статистическая форма регуляризованного метода наименьших квадратов А.Н. Тихонова. Докл. АН СССР, 1986. Т. 291. № 4. С. 780–785.
- 48. Крянев А.В. Применение современных методов параметрической и непараметрической статистики при обработке данных экспериментов на ЭВМ. М.: МИФИ, 1987.
- 49. Крянев А.В. Применение современных методов математической статистики при восстановлении регрессионных зависимостей на ЭВМ. М.: МИФИ, 1988.
- 50. Крянев А.В., Лукин Г.В. О постановке и решении задач оптимизации инвестиционных портфелей. Препринт 006-2001. М.: МИФИ, 2001.
- 51. Крянев А.В., Лукин Г.В. Метрический анализ для интерполяции и прогнозирования функций многих переменных. М.: Препринт МИФИ, 003-2005.
- 52. Крянев А. В., Лукин Г. В. Математические методы обработки неопределенных данных. М.: Физмалит, 2006.
- 53. Крянев А. В., Лукин Г. В. Метрический анализ и обработка данных М.: Физмалит, 2010.
- 54. *Крянев А.В.*, *Черный А.И.* Робастные линейные сглаживающие сплайны и их применения. Препринт 006-97. М.: МИФИ, 1997.
- 55. *Кузнецов С.П.* Динамический хаос (курс лекций). М.: Физматлит, 2001.
- 56. Лаврентьев М.М., Романов В.Г., Шишатский С.П. Некорректные задачи математической физики и анализа. М.: Наука, 1980.
- 57. 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.
- 58. Леман Э. Теория точечного оценивания. М.: Наука, 1991.
- Лисковец Д.А. Вариационные методы решения неустойчивых задач. Минск: Наука и техника, 1981.
- Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. М.: Финансы и статистика, 2003.
- 61. *Малинецкий Г.Г.*, *Потапов А.Б.* Нелинейная динамика и хаос. Основные понятия. М.: Комкнига, 2009.
- 62. *Малинецкий Г.Г.* Нелинейная динамика: Подходы, результаты, надежды. М.: Комкнига, 2009.
- 63. Марчук Г.И. Методы вычислительной математики. М.: Наука, 1989.
- 64. *Марчук Г.И.*, *Атанбаев С.А.* Некоторые вопросы глобальной регуляризации // ДАН СССР. 1970. Т. 190. № 3.
- 65. *Марчук Г.И.*, *Васильев В.Г.* О приближенном решении операторных уравнений первого рода // ДАН СССР. 1970. Т. 199. № 4.
- 66. Морозов В.А. Регулярные методы решения некорректных задач. М.: Наука, 1987.
- 67. Муравьева М.В. Об оптимальных и предельных свойствах Байесовского решения системы линейных алгебраических уравнений // ЖВММФ. 1973. Т.13. № 7. С. 819–828.
- 68. Мэйндоналд Дж. Вычислительные алгоритмы в прикладной статистике. М.: Финансы и статистика, 1988.

- 69. Надарая Э.А. Непараметрическое оценивание плотности вероятностей и кривой регрессии. Тбилиси: ТГУ, 1983.
- 70 Оселедец В.И. Мультипликативная эргодическая теорема. Характеристические показатели Ляпунова динамических систем. Тр. Моск. матем. общ. 1968. Т. 19. С. 179–210.
- Переамент А.Х. Регуляризованные алгоритмы статистического оценивания функций при наличии корреляционной помехи / Препринт ИПМ им. М.В. Келдыша АН СССР. 1986. № 21.
  - 72. 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, 1979. № 111.
  - Пределы предсказуемости / Под ред. Ю.А. Кравцова. М.: ЦентрКом, 1997.
  - 74. Пригожин И.Р. Конец определенности: Время, хаос и новые законы природы. Ижевск: ИРТ, 1999.
  - Пугачев В.С. Теория вероятностей и математическая статистика. М.: Физматлит, 2002.
  - Пугачев В.С., Синицын И.Н. Теория стохастических систем. М.: Логос. 2000.
  - 77. Пытьев Ю.П. Математические методы интерпретации эксперимента. М.: Высшая школа, 1989.
  - Рао С.Р. Линейные стохастические методы и их применения. М.: Наука. 1968.
  - Рунион Р. Справочник по непараметрической статистике. Современный подход. М.: Финансы и статистика. 1982.
  - Савелова Т.И. Примеры решения некорректно поставленных задач. М.: МИФИ, 1999.
  - 81. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980.
  - 82 Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания. М.: Статистика, 1980.
  - 83. Соболь И.М. Численные методы Монте-Карло. М.: Наука, 1973.
  - Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1979.
  - Тихонов А.Н.. Гончарский А.В., Степанов В.В., Ягола А.Г. Регуляризующие алгоритмы и априорная информация. М.: Наука, 1985.
  - Тихонов А.Н., Леонов А.С., Ягола А.Г. Нелинейные некорректно поставленные задачи. М.: Наука, 1995.
  - 87. 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.
  - Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере / Под ред. Фигурнова В.Э. М.: ИНФРА-М, 1998.
  - 89. Федер E. Фракталы. М.: Мир, 1991.
  - 90 Федоренко Р.П. Введение в вычислительную физику. М.: Изд-во Физтеха, 1994.
  - 91. Федоров В.В. Теория оптимального эксперимента. М.: Наука, 1971.
  - 92 Федотов А.М. Линейные некорректные задачи со случайными ошибками в данных. — Новосибирск: Наука, Сибирское отделение, 1990.

- 93. Фракталы в физике. Труды 6-го международного симпозиума по фракталам в физике / Под редакцией Л. Пьетронеро, Э. Тозетти. М.: Мир, 1988.
- 94. Хакен Г. Синергетика. М.: Мир, 1980.
- 95. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. М.: Мир, 1989.
- Холдендер М., Вулф Д. Непараметрические методы статистики, М.: Финансы и статистика, 1983.
- 97. Хьюбер П. Робастность в статистике. М.: Мир, 1984.
- 98. *Цыпкин Я.З.* Информационная теория идентификации. М.: Наука, 1995.
- Ченцов Н.Н. Почему L<sub>1</sub>-подход и что за горизонтом? Дополнение 1 к книге Деврой Л., Дьерфи Л. Непараметрическое оценивание плотности. — М.: Мир, 1988.
- 100. Ченцов Н.Н. Статистические решающие правила и оптимальные выводы. М.: Наука, 1972.
- 101. Шарковский А.М., Майстренко Ю.А., Романенко Е.Ю. Разностные уравнения и их приложения. Киев: Наукова Думка, 1986, 280 с.
- 102. Шарковский А.М. Сосуществование циклов непрерывного отображения прямой в себя // Укр. матем. журн. 1964. № 1. С. 61–71.
- 103. *Шарп У.Ф.*, *Александер Г.Дж.*, *Бейли Д.В.* Инвестиции. М.: Инфра-М, 2000.
- 104. Шурыгин А.М. Прикладная стохастика: робастность, оценивание, прогноз. — М.: Финансы и статистика, 2000.
- 105. Шустер Г. Детерминированный хаос. М.: Мир, 1988.
- 106. Abarbanel H. D. I. Analysis of Observed Chaotic Data. N. Y.: Springer, 1996.
- 107. 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.
- 108. Benettin G., Galgani L., Giorgilli A., Strelcyn J.M. Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems: A method for computing all of them. P.1: Theory. P.2: Numerical application // Meccanica. 1980. V. 15 P. 9–30.
- 109. 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.
- 110. 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.
- 111 Broomhead D.S., King G.P. Extracting qualitative dynamics from experimental data // Physica D. 1986. V. 20. P. 217-236.
- 112. Daubechies I. Ten lectures on wavelets. SIAM, 1992.
- 113. Draper R. N., Smith H. Applied Regression Analysis. N. Y.: Wiley, 1998.
- 114. Elsner J. B., Tsonis A. A. Singular Spectrum Analysis. New Tool in Time Series Analysis. — N. Y.: Plenum Press, 1996.
- 115. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure. SSA and Related Techniques. — Chapman & Hall / CRS, 2001.
- 116. Hernander E., Weiss G. A First Course of Wavelets. CRS Press, 1996.

- 117. Jollitte I. T. Principal Component Analysis. Springer Series in Statistics. N. Y.: Springer-Verlag, 1986.
- 118. Kantz H., Schreiber T. Nonlinear Time Series Analysis. Cambridge: Cambridge University Press, 1997.
- 119. Keizer G. A Friendly Guide to Wavelets. Birk, 1995.
- 120. Kryanev A. V., Lukin G. V., Udumyan D.K. Metric analysis and applications. Numerical Methods and Programming. 2009. V. 10. P. 408-414.
- 121. Lokenath Debnath. Wavelet Transforms and Their Applications. Hardcover Edition, 2002.
- 122 Ly T.-Y., Yorke J.A. Period Three Implies Chaos. Amer. Math. Monthly. 1975. V. 82, P. 982-985,
- 123. Misiti M., Misiti V., Oppenheim G., Poggy J.-M. Wavelet Toolbox for Use with MATLAB, 1988.
- 124. Mosteller F., Tukey J. W. Data Analysis and Regression: Second Course in Statistics. Reading. MA: Addison-Wesley, 1977.
- 125. Parzen E. On estimation of a probabilities for sums of bounded random variables // Annals of Mathematical Statistics. 1962. V. 33. P. 1065-1076.
- 126 Persival D. B., Walden A. T. Wavelet Methods for Time Series. Hardcover Edition, 2002.
- 127. Rosenblatt M. Remarks on some nonparametric estimates of a density function // Annals of Mathematical Statistics. 1956. V. 27. P. 832-835.
- 128. Takens F. Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence. Lecture Notes in Mathematics. Berlin: Springer-Verlag. 1980. V. 898. P. 366-381.
- 129 Tikhonov A.N., Leonov A.S., Yagola A.G. Nonlinear III-Posed Problems. V. 1, 2. London: Chapman Hall, 1998.
- 130. Vautard R., Yiou P., Ghil M. Singular-spectrum analysis: A toolkit for short noisy chaotic signals // Physica D. 1992. V. 58. P. 95-126.