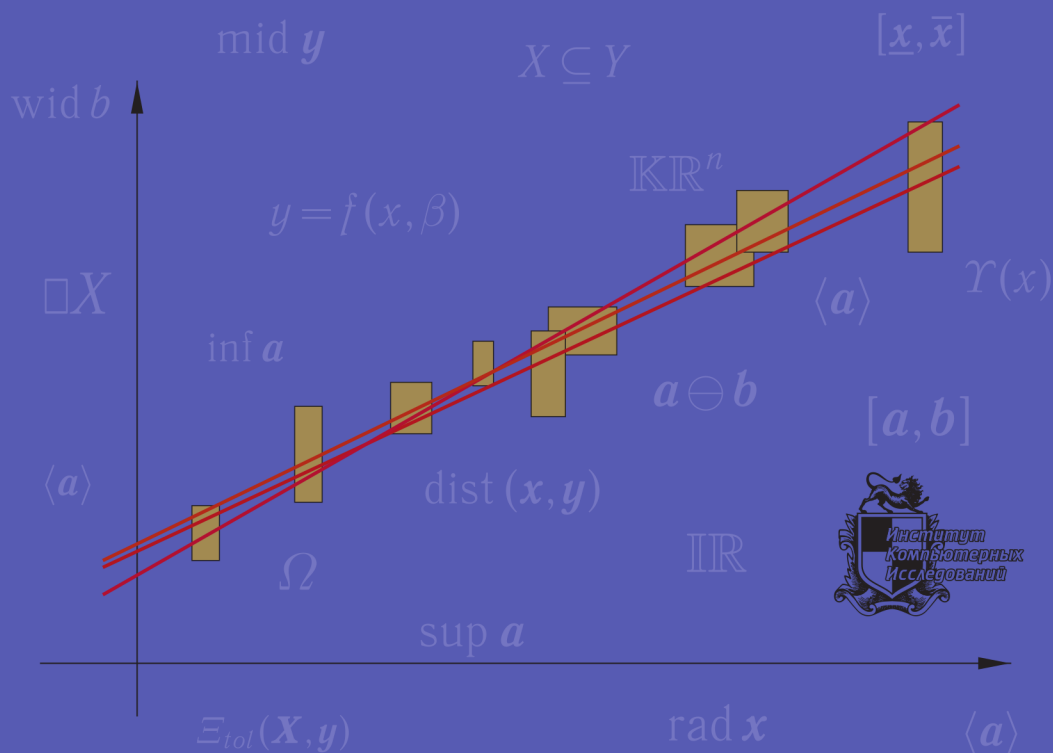


А. Н. Баженов, С. И. Жилин,
С. И. Кумков, С. П. Шарый

ОБРАБОТКА И АНАЛИЗ ИНТЕРВАЛЬНЫХ ДАННЫХ



А. Н. Баженов, С. И. Жилин,
С. И. Кумков, С. П. Шарый

ОБРАБОТКА И АНАЛИЗ ИНТЕРВАЛЬНЫХ ДАННЫХ



Москва ♦ Ижевск

2024

УДК 518+519.4+519.6, 658.562.012+519.254

ББК 22.161

О230

Баженов А. Н., Жилин С. И., Кумков С. И., Шарый С. П.

О230 **Обработка и анализ интервальных данных.** — М.; Ижевск :
Институт компьютерных исследований, 2024. — 356 с.

ISBN 978-5-4344-1018-2

В книге рассматриваются методы обработки результатов измерений и наблюдений, которые неточны и имеют интервальную неопределённость, тогда как условия применения теории вероятностей не выполнены и какие-либо вероятностные характеристики неопределённости отсутствуют. В подобных ситуациях стандартные статистические методы, основанные на теоретико-вероятностной модели погрешностей, могут применяться лишь формально. Развиваемые в книге подходы к анализу интервальных данных основаны на методах интервального анализа и смежных дисциплин, опираются на использование специализированных программных средств для ЭВМ.

В книге развивается взаимосогласованная система понятий и терминов новой дисциплины — интервального анализа данных, и на конкретных примерах иллюстрируется её использование для решения задач обработки информации.

ББК 22.161

УДК 518+519.4+519.6,

658.562.012+519.254

ISBN 978-5-4344-1018-2

© А. Н. Баженов, С. И. Жилин,
С. И. Кумков, С. П. Шарый, 2024

© АНО «Ижевский институт компьютерных исследований», 2024

Оглавление

Введение	6
ГЛАВА 1. Обзор методов статистики и обработки данных	10
1.1. Статистика и теория вероятностей	10
1.2. Данные, погрешности и их обработка	15
1.3. Критика вероятностной статистики	19
1.4. Альтернативы вероятностной статистике	31
1.4.1. Статистика интервальных данных	31
1.4.2. Статистика нечётких данных	32
1.4.3. Почему интервалы?	35
1.5. Место и особенности интервального подхода	41
ГЛАВА 2. Базовые понятия и математический аппарат	45
2.1. Интервалы	45
2.2. Классическая интервальная арифметика	53
2.3. Преобразования характеристик интервалов	55
2.4. Независимые и связанные интервальные величины	57
2.5. Преобразования интервалов	61
2.6. Примеры простейших интервальных расчётов	65
2.7. Интервальная арифметика Кэхэна	70
2.8. Полная интервальная арифметика Каухера	72
2.9. Аффинная и функциональная арифметики	76
2.10. Твины — интервалы интервалов	77
2.11. Мультиинтервалы	83
2.12. Интервальные векторы и матрицы	85
2.13. Двойственный характер интервальной неопределённости	90
2.14. Сравнение интервалов	92
2.15. Истинное значение измеряемой величины	98
2.16. Измерения и их результаты	101
2.17. Погрешность измерений	111
2.18. Накрывающие и ненакрывающие измерения	113

ОГЛАВЛЕНИЕ

2.19. Принцип соответствия	116
2.20. Выбросы и промахи	118
2.21. Выборки интервальных данных	120
2.22. Информационное множество	123
2.23. Дальнейшая классификация интервальных данных	127
2.24. Оценки точечные и интервальные	129
2.25. Доказательные интервальные вычисления	131
ГЛАВА 3. Измерение постоянной величины	135
3.1. Выборка измерений и интервалы их неопределённости	135
3.2. Обработка накрывающих выборок	139
3.3. Совместность выборок	143
3.3.1. Определение и общие соображения	143
3.3.2. Граф совместности выборки	148
3.3.3. Примеры использования графа совместности	150
3.3.4. Индекс Жаккара	153
3.4. Мода интервальной выборки	156
3.5. Выборки унимодальные и мультимодальные	161
3.6. Медиана интервальной выборки	162
3.7. Обработка ненакрывающих выборок	163
3.8. Вариабельность оценки	168
3.9. Критические соображения П. Е. Эльясберга и их разбор	171
3.10. Приём варьирования ширины интервальных данных	174
ГЛАВА 4. Задача восстановления функциональных зависимостей	181
4.1. Постановка задачи и необходимые понятия	181
4.2. Накрывающие и ненакрывающие измерения и выборки	185
4.3. Информационное множество задачи	187
4.4. Прогнозный коридор и коридор совместных зависимостей	193
4.5. Случай точных измерений входных переменных	198
4.6. Конкретный пример и сравнение с оценками МНК	203
4.7. Интервальные уравнения и системы уравнений	208
4.8. Решения интервальных уравнений и их систем	210
4.9. Общий случай задачи восстановления зависимостей	217
4.10. Приближение и оценивание информационного множества	223
4.11. Состоятельность и несмещённость	230
4.12. Численные методы для интервальных систем уравнений	234
4.13. Парадоксы интервального анализа данных	240

ОГЛАВЛЕНИЕ

4.13.1. Парадокс Е. З. Демиденко	240
4.13.2. Парадокс А. И. Хлебникова	242
4.13.3. Анализ парадоксов и их объяснение	243
4.14. Обзор методов восстановления зависимостей по интер- вальным данным	246
4.14.1. Простейшие методы	247
4.14.2. Прямое построение информационного множества	248
4.14.3. Метод парциальных информационных множеств	252
4.14.4. Метод центра неопределённости	256
4.14.5. Метод максимума совместности	261
4.14.6. Примеры применения метода максимума совмест- ности	269
4.14.7. Исчерпывающее оценивание информационного мно- жества	278
4.14.8. Пример Миланезе – Вичино	287
4.14.9. Метод нарезки информационного множества	290
4.15. Вариабельность оценок параметров	296
4.15.1. Общие соображения	296
4.15.2. Ограниченность информационного множества	303
4.15.3. Мера вариабельности для оценки максимума сов- местности	308
4.16. Восстановление зависимостей по ненакрывающим выбор- кам	318
Обозначения	322
Литература	326
Предметный указатель	350

Введение

Главная цель нашей книги — представить внутренне согласованную систему понятий, терминов и фактов, относящихся к обработке данных, которые неточны и представлены интервалами своих возможных значений. Часто говорят, что такие данные имеют *интервальную неопределённость* (или, в более общем случае, ограниченную по величине неопределённость). Мы будем называть их также *интервальными данными*. В книге даётся также краткий обзор основных и наиболее значимых результатов этого научного направления, которое можно назвать «статистикой интервальных данных» или же «анализом интервальных данных». В то же время материал книги относится и к метрологии, поскольку введение нового типа данных — интервалов — для результатов измерений или наблюдений, обоснование этого шага, разбор вытекающих из этого следствий являются метрологическими вопросами.

Возникнув в последние десятилетия XX века как альтернатива традиционной вероятностной статистике, основанной на методах теории вероятностей, анализ интервальных данных вскоре сделался важным и практически востребованным. Его интенсивное развитие вызвало к жизни необходимость введения различных понятий, обозначений, конструкций и т. п., которые не всегда согласовывались друг с другом. К настоящему моменту возникла настоятельная необходимость унификации понятий, терминов и обозначений в этой научной области, чтобы специалисты могли лучше понимать друг друга и были понятными для тех, кто применяет методы обработки интервальных данных.

Помимо собственно анализа данных, такую же работу необходимо провести в отношении тех вопросов, которые обычно относятся к области метрологии, — переосмыслить понятия измерения, результата измерений и погрешности, ревизовать базовые математические методы их обработки. Это совершенно неизбежно при введении нового интервального типа данных для результатов измерений.

В практике обработки экспериментальных данных сейчас повсеместно используются статистические методы, основанные на идеях, результатах и общей методологии теории вероятностей (см., например,

[40, 41, 42, 139, 169, 240]). Эти подходы опираются на использование ряда допущений о вероятностных свойствах погрешностей измерений, а также наличие выборок представительной длины (как минимум в несколько десятков измерений). Однако практики часто сталкиваются с ситуациями, когда выборки измерений коротки (например, не более 20–30 или даже не более десятка), а погрешности измерений не могут быть адекватно описаны с помощью инструментов теории вероятностей или же информация о вероятностных характеристиках погрешностей отсутствует. В этих условиях погрешности и неопределённости измерений нужно описывать и обрабатывать уже по-другому. В частности, для анализа данных можно применить методы интервальной статистики, основанные на идеях и результатах интервального анализа, использующие его подходы, алгоритмы и соответствующее программное обеспечение.

Напомним, что *интервалом* вещественной оси называется множество всех чисел, расположенных между заданными числами, включая их самих. Интервалы, таким образом, дают одну из форм представления диапазонов значений интересующих нас величин, и в этом качестве интервалы главным образом и используются в самой математике и её практических приложениях. Интервальный анализ — это, в узком смысле, исчисление интервалов, т.е. наука о том, как оперировать с ними как самостоятельными объектами, наподобие того, как мы делаем это с обычными числами. Развитая на этой основе техника используется для решения различных задач, где интервалы встречаются в виде данных, входных или промежуточных.

Интервальный анализ возник в середине XX века в связи с бурным развитием компьютерных вычислений, в которых требовалось строго и аккуратно оценивать эффект от замены точных значений на приближённые. Сам термин «интервальный анализ» был введён американцем Р. Э. Муром в середине 60-х годов прошлого века (см. [293]), но он лишь удачно завершил оформление нового научного направления, которое реально зародилось несколькими десятилетиями ранее. В его становление внесли свой вклад многие исследователи из СССР, США, Великобритании, Германии, Польши, Японии, начиная с первых десятилетий XX века. В частности, у нас в стране идеи интервального анализа развивались В. М. Брадисом, математиком-педагогом и вычислителем, автором широко известных математических таблиц для средней школы, а затем Л. В. Канторовичем [74], выдающимся математиком и экономистом. Со второй половины 60-х годов XX века интервальный анализ и его приложения интенсивно развивались научной школой Н. Н. Яненко в Сибири.

В настоящее время интервальный анализ является развитой математической дисциплиной со своим кругом задач и своими специфическими методами, которые проникают во многие прикладные отрасли знаний. Для знакомства с различными аспектами современного интервального анализа можно порекомендовать книги [3, 49, 200, 267, 288, 295, 300]. Но терминология интервальной статистики — интервальных методов обработки данных — по необходимости наследует многое из традиционной статистики, где сложился развитый понятийный аппарат, который можно и нужно использовать в новых условиях как для удобства, так и для преемственности новой науки.

Наша книга предназначена для широкого круга читателей, и прежде всего для прикладников — метрологов, инженеров и техников, сталкивающихся с необходимостью обработки данных, результатов измерений и наблюдений, к которым традиционные теоретико-вероятностные методы плохо применимы. Немало полезного найдётся в книге также для математиков и специалистов по математическому моделированию, развивающих и использующих интервальный анализ и смежные области знаний.

О содержании книги можно составить представление по её оглавлению. Отдельным крупным темам, в том числе математическим, посвящены отдельные главы книги, и они получились неравноценными по объёму. Около половины всей книги занимает глава 4, рассматривающая задачу восстановления функциональных зависимостей. Это вызвано, конечно, большой практической важностью задачи, а также количеством и качеством полученных по этой теме результатов. Несмотря на довольно общее название книги, в ней не получили отражение те вопросы обработки интервальных данных, которые относятся, например, к кластерному анализу, факторному анализу, планированию эксперимента и пр. Причина состоит главным образом в малой разработанности теории соответствующих задач для случая интервальных данных. Эти постановки ещё ждут своих исследователей. С другой стороны, общность названия книги оправдывается тем, что в ней впервые столь тщательно прорабатываются самые общие вопросы, которые возникают во всех частных дисциплинах, имеющих дело с обработкой и анализом интервальных данных.

Вклад авторов в текст книги распределился следующим образом. А. Н. Баженов написал § 2.6, § 2.10 и § 2.11. С. И. Жилину принадлежат § 3.3.2 и § 3.3.3. Перу С. И. Кумкова принадлежит § 4.14.3, а также общая идея издания этой книги. С. П. Шарый написал главу 1, §§ 2.1–2.5,

ВВЕДЕНИЕ

§§ 2.7–2.9, §§ 2.12–2.14, §§ 2.16–2.22, §§ 2.24–2.25, § 3.9, §§ 4.1–4.13, § 4.14.1, § 4.14.2, §§ 4.14.4–4.14.9, а также §§ 4.11–4.16. Остальные разделы книги написаны авторами совместно.

Материал книги использовался в учебных курсах по интервальному анализу данных, которые первый и последний авторы читают в Санкт-Петербургском политехническом университете и Новосибирском государственном университете.

Авторы искренне благодарны Ларисе Аркадьевне Игнатенковой, руководителю Центра «СЕРТИМЕТ» УрО РАН, за ценную информацию, консультации и обсуждения, Всероссийскому интервальному веб-семинару, на котором обсуждалась значительная часть материала книги, а также А. В. Пролубникову, Е. В. Чаусовой, А. А. Карповой, Д. Ю. Надёжину, К. К. Семёнову и С. С. Шарой за полезные замечания.