

Министерство образования науки Российской
Федерации

САНКТ-ПЕТЕРБУРГСКИЙ
ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА
ВЕЛИКОГО

А. Н. Баженов

Обобщение мер совместности
для анализа данных с интервальной
неопределённостью.

Учебное пособие

Санкт-Петербург

2022

УДК 519.9 Р32

А в т о р :

А.Н.Баженов. Обобщение мер совместности для анализа данных с интервальной неопределённостью. – СПб., 2022. – 80 с.

Учебное пособие соответствует образовательному стандарту высшего образования Федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого» по направлению подготовки бакалавров 01.03.02 «Прикладная математика и информатика», по дисциплинам «Математическая статистика» и «Интервальный анализ» и магистров 01.04.02 «Прикладная математика и информатика» по дисциплине «Анализ данных с интервальной неопределённостью».

Настоящее пособие представляет оригинальный материал, относящийся к классу задач классификации. В последние годы в этом классическом разделе анализа данных появилось много работ с учётом интервальной неопределённости данных. Вместе с тем, в этих публикациях рассматриваются только выборки совместных данных. Мы расширяем методическую основу рассмотрения вопроса, привлекая полную интервальную арифметику Каухера. Это позволяет обобщить применение классических методов на анализ несовместных выборок данных.

Предложенные меры и различные математические конструкции могут быть использованы для характеристики совместности отдельных интервалов, составных интервальных объектов и интервальных множеств, а также применены в схемах вычислительных методов.

© Санкт-Петербургский политехнический университет Петра Великого, 2022

Ключевые слова. Интервальный анализ, анализ данных с интервальной неопределённостью, мера совместности, коэффициент Жаккара, множества решений, регуляризация, арифметика Каухера, классическая интервальная арифметика

Оглавление

Обозначения	8
Введение	10
1 Краткие сведения об интервальном анализе	12
1.1 Интервальный анализ и интервальные арифметики.	13
2 Мера совместности (покрытия) интервальных величин	16
2.1 Анализ данных с интервальной неопределённостью.	16
2.2 Приём варьирования неопределённости	18
2.3 Бинарная мера совместности множеств.	19
2.4 Бинарная мера совместности (покрытия) интервалов.	20
2.5 Мера совместности (покрытия) для интервальной выборки.	24
2.6 Случай многомерных интервалов (брусков) и выборок многомерных интервалов.	26
2.7 Случай неограниченных интервалов.	27
3 Обсуждение меры покрытия в общем контексте анализа интервальных данных.	30
3.1 Применение JK к твинной арифметике и представление данных измерений с несколькими видами погрешности.	30
3.2 Место JK в ряду характеристик интервальных выборок.	33
4 Применения меры покрытия для характеристики вычислительных процессов.	37
4.1 Одномерный случай.	37
4.2 Многомерный случай.	39

5	Характеризация множеств решений систем интервальных уравнений	42
5.1	Множества решений систем интервальных уравнений . . .	42
5.2	Аналитическая характеристика множеств АЕ-решений ИСЛАУ.	43
6	Применения меры покрытия в интервальном анализе данных.	48
6.1	Характеризация области значения переменной	48
6.2	Оценка степени совместности данных при восстановлении зависимостей	52
7	Построение гистограмм интервальных данных	55
8	Регуляризация данных на основе меры совместности	58
8.1	Нахождение оптимального параметра по мере покрытия	59
8.1.1	Применение коэффициента Жаккара для регуляризации. Случай совместной выборки.	59
8.1.2	Применение коэффициента Жаккара для регуляризации. Случай несовместной выборки.	62
8.2	Характеристика вариабельности меры сходства	71
	Литература	75
	Предметный указатель	79

Список примеров

1	Вычисление несимметричных мер сходства	24
2	Мера совместности для интервальной неперекрывающейся вы- борки	25
3	Мера совместности для интервальной перекрывающейся выборки	25
4	Вычисление меры совместности для неограниченных ин- тервалов	28
5	Температура как твин	31
6	Характеризация твинов	32
7	Пример вычисления меры совместности для интервальной выборки	33
8	Мера совместности для интервальной выборки и мера совместности	35
9	Метод Ньютона — одномерный случай [31].	37
10	Метод Ньютона — многомерный случай.	39
11	Допусковое множество ИСЛАУ и график функционала Tol_{JK}	45
12	Объединённое множество ИСЛАУ и график функционала Uni_{JK}	46
13	Характеризация области значения переменной	49
14	Характеризация области значения переменной — контроль- ные измерения.	51
15	Оценка степени совместности данных при восстановлении зависимостей	52
16	Гистограмма интервальной выборки	56
17	Применение коэффициента Жаккара для регуляризации. Сравнение рабочей выборки с эталонной	59
18	Применение коэффициента Жаккара для регуляризации. Сравнение двух выборок	62

Список иллюстраций

2.1	Различные варианты положения интервалов и значения JK.	23
2.2	Диаграмма рассеяния интервальной выборки (3.11).	26
3.1	Температура как твин.	31
3.2	Данные по физике нейтрино в форме твинов типа « \leq ».	32
3.3	Диаграмма рассеяния интервальной выборки (3.11) и элементы выборки \mathbf{z}	34
3.4	Значения частот μ_i , интервальная мода $\text{mode } \mathbf{X}$ выборки (3.11) и элементы выборки $\mathbf{z}_k : k \in K$	35
3.5	Варианты отношений положения интервалов для накрывающих выборок, интервальная мода и значения JK.	35
4.1	Брусы оператора Кравчика и графики компонент $\delta \mathcal{H}_i$	41
5.1	Допусковое множество и график функционала J_{K_A}	46
5.2	Объединённое множество и график функционала JK.	47
6.1	Измерения температуры двумя датчиками.	49
6.2	График значений функционала JK для данных Рис. 6.1.	50
6.3	Графики температуры для области стабилизации.	50
6.4	Графики JK для области стабилизации.	51
6.5	Графики температуры для контрольного измерения.	51
6.6	Графики температуры и JK для контрольного измерения.	52
6.7	Графики данных датчика и вектор JK.	54
7.1	Диаграмма рассеяния и гистограмма интервальной выборки (7.2).	56

8.1	Диаграммы рассеяния интервальных измерений величины X_1 и X_2 [6].	60
8.2	Индекс Жаккара, показано информационное множество значений R_{21} по формуле (8.9).	61
8.3	Гистограмма значений объединённой выборки.	62
8.4	Диаграмма рассеяния интервальных измерений величины [39] для пика полного поглощения, Табл. 8.1.	65
8.5	Индекс Жаккара, показан оптимум по формуле (8.17).	68
8.6	Диаграмма рассеяния интервальных измерений величины для исходной выборки (столбцы 4-5) и скорректированных данных (столбцы 2-3).	69
8.7	Мода интервальной выборки величины δ для всех данных [39].	70
8.8	Графическое представление интервальных данных и результаты обработки по методике [41].	70
8.9	Элементы характеристика variability множества совместности	72

Обозначения

\Rightarrow	логическая импликация
\Leftrightarrow	логическая равносильность
$\&$	логическая конъюнкция, связка «и»
\rightarrow	отображение множеств; предельный переход
\emptyset	пустое множество
$x \in X$	элемент x принадлежит множеству X
$x \notin X$	элемент x не принадлежит множеству X
$X \cup Y$	объединение множеств X и Y
$X \cap Y$	пересечение множеств X и Y
$JK(x, y)$	мера совместности интервалов x и y
$JK_p(x, y)$	мера совместности интервалов x и y положительно определённая
$s_x(x, y)$	мера совместности интервалов x и y несимметричная
$X \setminus Y$	разность множеств X и Y
$X \subseteq Y$	множество X есть подмножество множества Y
$X \subset Y$	X есть собственное подмножество множества Y
$X \times Y$	прямое декартово произведение множеств X и Y
\mathbb{R}	множество вещественных (действительных) чисел
\mathbb{IR}	классическая интервальная арифметика
\mathbb{KR}	полная интервальная арифметика Каухера
\mathbb{R}^n	множество вещественных n -мерных векторов

$\mathbb{IR}^n, \mathbb{KR}^n$	множества n -мерных интервальных векторов
$\mathbb{R}^{m \times n}$	множество вещественных $m \times n$ -матриц
$\mathbb{IR}^{m \times n}, \mathbb{KR}^{m \times n}$	множества интервальных $m \times n$ -матриц
$[a, b]$	интервал с нижним концом a и верхним b
$]a, b[$	открытый интервал с концами a и b
$\underline{a}, \inf \mathbf{a}$	левый конец интервала \mathbf{a}
$\bar{a}, \sup \mathbf{a}$	правый конец интервала \mathbf{a}
$\text{mid } \mathbf{a}$	середина интервала \mathbf{a}
$\text{wid } \mathbf{a}$	ширина интервала \mathbf{a}
$\text{rad } \mathbf{a}$	радиус интервала \mathbf{a}
$\text{dual } \mathbf{a}$	дуальный (двойственный) к \mathbf{a} интервал
$\text{pro } \mathbf{a}$	правильная проекция интервала \mathbf{a}
$[[\underline{a}, \bar{a}], [\underline{b}, \bar{b}]]$	интервал с интервальными концами, твин
$[[\underline{X}^{in}, \bar{X}^{in}], [\underline{X}^{out}, \bar{X}^{out}]]$	твин в форме Нестерова
$ \mathbf{a} $	абсолютное значение (модуль) интервала \mathbf{a}
$\square X$	интервальная оболочка множества $X \subseteq \mathbb{R}^n$
\wedge	операция минимума по включению
\vee	операция максимума по включению
dist	расстояние (метрика) на множестве интервалов
Dist	векторнозначное расстояние (мультиметрика)
$\text{ran}(f, X)$	область значений функции f на множестве X
\min, \max	операции взятия минимума и максимума
\sum	символ суммы нескольких слагаемых
$\ \cdot\ $	векторная или матричная норма
x^*	истинное значение измеряемой величины
\hat{x}	базовое измеренное значение величины
Ω	информационное множество задачи
Ξ	множество решений интервальной системы

Введение

На протяжении последних десятилетий возникло новое направление в анализе данных, которое учитывает их неопределенность. Манифестом этого подхода, обобщающим различные исследования, является книга коллектива авторов А.Н. Баженов, С.И. Жилин, С.И. Кумков, С.П. Шарый «Обработка и анализ данных с интервальной неопределённостью» [1]. В этой книге представлена согласованная система понятий и терминов, относящихся к обработке данных, имеющих интервальную неопределённость, и дан краткий обзор основных и наиболее значимых результатов научного направления, который авторы книги называют «статистикой интервальных данных» или же «анализом интервальных данных». Для более краткого введения в предмет служит учебное пособие автора «Введение в анализ данных с интервальной неопределённостью» [5], содержащее краткий обзор понятий и идей этой книги и примеры их использования.

Для чтения настоящего пособия необходимо знакомство с интервальным анализом, системой его понятий и инструментов. Наиболее основательное изложение идей и методов интервального анализа дано в книге С.П. Шарого «Конечномерный интервальный анализ» [4]. Изучение материала этой книги требует более основательной математической подготовки и рекомендуется для углублённого изучения вопроса.

Основы интервального анализа представлены в учебном пособии автора «Интервальный анализ. Основы теории и учебные примеры» [2]. В учебном пособии автора «Естественнонаучные и технические применения интервального анализа» [3] дана картина применения интервальных данных и интервального анализа в более широком контексте.

В практике обработки экспериментальных данных сейчас повсеместно используются статистические методы, основанные на идеях и результатах теории вероятностей. Эти подходы опираются на исполь-

зование 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, а погрешности измерений не могут быть адекватно описаны с помощью инструментов теории вероятностей или же информация о вероятностных характеристиках погрешностей отсутствует. Для решения таких задач можно применить методы «интервальной статистики», основанные на идеях и результатах интервального анализа, использующие его подходы, алгоритмы и соответствующее программное обеспечение. Интервальные методы поддерживаются практически на всех популярных платформах программирования. В некоторых интегрированных средах, как например, *Mathematica*, *Octave* поддержка базовых интервальных конструкций встроена. Для наиболее популярного в настоящее время языка программирования *Python* также есть реализации основных конструкций и методов интервального анализа.

Терминология интервальной статистики естественным образом наследует многое из традиционной статистики, где сложился развитый понятийный аппарат. Различным аспектам анализа интервальных данных посвящены, в частности, публикации [8] и книга [9].

Следует заметить, что современная теория и практика обработки данных выходит далеко за рамки теоретико-вероятностной математической статистики. Дж. Тьюки в конце 50-х годов прошлого века предложил оформить новую научную дисциплину — «анализ данных», в котором охватывались те математические методы обработки данных, которые не попадали под «математическую статистику» в узком смысле этого слова [10].

Настоящее пособие представляет оригинальный материал, относящийся к классу задач классификации. В последние годы в этом классическом разделе анализа данных появилось много работ с учётом интервальной неопределённости данных. Вместе с тем, в них рассматриваются только выборки совместных данных. Мы расширяем методическую основу рассмотренного вопроса, привлекая полную интервальную арифметику Каухера. Это позволяет обобщить применение классических методов на анализ несовместных выборок данных. Первыми примерами таких исследований являются работы [6] и [7].

Глава 1

Краткие сведения об интервальном анализе

Данные, погрешности и их обработка На практике данные не бывают точными. Реально нам известно приближённое значение измеряемой величины, а также некоторая информация (качественная и количественная) о погрешности этого значения. На результаты измерений могут оказывать влияние изменчивость самих измеряемых величин, их непостоянство во времени или пространстве, на измерения могут влиять внешние неконтролируемые факторы, так называемые «шумы». У применяемой аппаратуры имеются собственные погрешности. В процессе математической обработки данных на результат влияют неизбежные неточности расчётов (ошибки представления и округления и т. п.).

В наиболее популярном подходе к обработки данных, теоретико-вероятностной математической статистике, оперируют с вещественными величинами, их множествами и функциями. Для описания погрешностей данных вводятся дополнительные конструкции. В то же время, для описания величин с погрешностями, естественно и удобно использовать понятия и методы специального раздела математики, который называется интервальный анализ. В интервальном анализе, как следует из названия, интервал является первичным объектом, на основе которого строится специальная арифметика и развиваются методы решения различных задач.

1.1 Интервальный анализ и интервальные арифметики.

В настоящей публикации меры сходства будут использоваться для математических объектов с интервальной неопределённостью. Приведём далее необходимые сведения об интервальном анализе, современная система обозначений интервальных объектов приводится в публикации [14].

Классическая интервальная арифметика \mathbb{IR} — алгебраическая система, образованная интервалами $\mathbf{x} = [\underline{x}, \bar{x}] \subset \mathbb{R}$ так, что для любой арифметической операции « \star » из множества $\{+, -, \cdot, /\}$ результат операции между интервалами определяется как

$$\mathbf{x} \star \mathbf{y} = \{ x \star y \mid x \in \mathbf{x}, y \in \mathbf{y} \}. \quad (1.1)$$

Развёрнутые формулы для арифметических операций в \mathbb{IR} выглядят следующим образом:

$$\mathbf{x} + \mathbf{y} = [\underline{x} + \underline{y}, \bar{x} + \bar{y}], \quad (1.2a)$$

$$\mathbf{x} - \mathbf{y} = [\underline{x} - \bar{y}, \bar{x} - \underline{y}], \quad (1.2b)$$

$$\mathbf{x} \cdot \mathbf{y} = [\min\{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\}, \max\{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\}], \quad (1.2c)$$

$$\mathbf{x}/\mathbf{y} = \mathbf{x} \cdot [1/\bar{y}, 1/\underline{y}] \quad \text{для } \mathbf{y} \not\ni 0. \quad (1.2d)$$

Действующий IEEE1788 стандарт для интервальных вычислений принят в 2015 году [15].

Далее в публикации будут использоваться обозначения:

$$\text{mid } \mathbf{a} = \frac{1}{2}(\underline{a} + \bar{a}) \quad \text{— середина интервала,} \quad (1.3)$$

$$\text{wid } \mathbf{a} = \bar{a} - \underline{a} \quad \text{— ширина интервала.} \quad (1.4)$$

На основе одномерных интервалов строятся более сложные объекты, интервальные векторы и матрицы. Арифметические операции с ними определяются «по представителям».

Теоретико-множественные операции между интервалами. Если интервалы \mathbf{a} и \mathbf{b} имеют непустое пересечение, т. е. $\mathbf{a} \cap \mathbf{b} \neq \emptyset$

\emptyset , то можно дать простые выражения для результатов теоретико-множественных операций пересечения и объединения через концы этих интервалов

$$\mathbf{a} \cap \mathbf{b} = [\max\{\underline{\mathbf{a}}, \underline{\mathbf{b}}\}, \min\{\bar{\mathbf{a}}, \bar{\mathbf{b}}\}],$$

$$\mathbf{a} \cup \mathbf{b} = [\min\{\underline{\mathbf{a}}, \underline{\mathbf{b}}\}, \max\{\bar{\mathbf{a}}, \bar{\mathbf{b}}\}].$$

Если же $\mathbf{a} \cap \mathbf{b} = \emptyset$, т. е. интервалы \mathbf{a} и \mathbf{b} не имеют общих точек, то эти равенства уже неверны.

Обобщением операций пересечения и объединения являются операции взятия минимума и максимума относительно включения « \subseteq ». Они обозначаются специальными символами \wedge и \vee :

$$\mathbf{a} \wedge \mathbf{b} = [\max\{\underline{\mathbf{a}}, \underline{\mathbf{b}}\}, \min\{\bar{\mathbf{a}}, \bar{\mathbf{b}}\}], \quad (1.5)$$

$$\mathbf{a} \vee \mathbf{b} = [\min\{\underline{\mathbf{a}}, \underline{\mathbf{b}}\}, \max\{\bar{\mathbf{a}}, \bar{\mathbf{b}}\}]. \quad (1.6)$$

Интервальная арифметика Каухера. В более общей постановке задачи интервального анализа можно решать в так называемой полной интервальной арифметике или арифметике Каухера, обозначаемой \mathbb{KR} . Эта арифметика получается присоединением неправильных интервалов $[\underline{x}, \bar{x}]$, $\underline{x} > \bar{x}$ к множеству \mathbb{IR} .

Введение интервальной арифметики Каухера мотивировано рядом ограничений классической интервальной арифметики. Среди них отсутствие обратных элементов относительно операций сложения и умножения, невозможность в общем случае взятия минимума по включению, сложности с постановкой минимаксных задач.

Интервальная арифметика Каухера обладает многими замечательными алгебраическими свойствами. В ней существуют обратные элементы по отношению к операциям сложения и умножения, для любых интервалов определена операция взятия минимума по включению. Применение неправильных интервалов позволяет строить оригинальные методы решения задач с интервальными объектами. Краткое изложение арифметики Каухера дано в публикации С.П.Шарого [?].

В связи с наличием в \mathbb{KR} неправильных интервалов, используется операции *дуализации* и взятия *правильной проекции*.

Правильные и неправильные интервалы, переходят друг в друга в результате отображения *дуализации* $\text{dual} : \mathbb{KR} \rightarrow \mathbb{KR}$, меняющего ме-

стами (переворачивающего) концы интервала, т. е. такого что

$$\text{dual } \mathbf{a} := [\bar{\mathbf{a}}, \underline{\mathbf{a}}].$$

Правильной проекцией интервала \mathbf{a} из \mathbb{KR} называется интервал, обозначаемый $\text{pro } \mathbf{a}$ и такой, что

$$\text{pro } \mathbf{a} = \begin{cases} \mathbf{a}, & \text{если } \mathbf{a} \text{ — правильный,} \\ \text{dual } \mathbf{a}, & \text{если } \mathbf{a} \text{ — неправильный.} \end{cases}$$

С помощью правильной проекции из произвольного интервала получается его «правильный образ», с которым можно обращаться как с обычным числовым интервалом в \mathbb{R} , т. е. множеством всех точек вещественной оси между двумя заданными концами. Арифметические операции между интервалами в \mathbb{KR} продолжают операции в \mathbb{IR} .

Для работы с интервальными объектами чрезвычайно важно следующее

Определение 1.1.1 *Для интервалов $\mathbf{a}, \mathbf{b} \in \mathbb{KR}$ выполняется включение $\mathbf{a} \subseteq \mathbf{b}$, если*

$$\underline{\mathbf{a}} \geq \underline{\mathbf{b}} \quad \text{и} \quad \bar{\mathbf{a}} \leq \bar{\mathbf{b}}.$$

Относительно введённого таким образом отношения включения в \mathbb{KR} для любых двух интервалов существует минимальный и максимальный по включению, т. е. результаты операций $\mathbf{a} \wedge \mathbf{b}$ и $\mathbf{a} \vee \mathbf{b}$ всегда определены.

Другие определения и конструкции интервального анализа будут приводиться в тексте по необходимости.

Глава 2

Мера совместности (покрытия) интервальных величин

2.1 Анализ данных с интервальной неопределённостью.

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. Различным аспектам анализа интервальных данных посвящены, в частности, публикации [8], [16], [17], [43], [20], [21], [22], [23], [24] и книга [9].

Анализу данных с интервальной неопределённостью (иначе, статистика интервальных данных, анализ интервальных данных) посвящена книга [1]. В ней представлена система понятий и терминов, относящихся к обработке данных, которые имеют интервальную неопределённость. В настоящем разделе мы следуем определениям, введённый в книге.

Понятие накрывающего измерения (замера) и накрывающей выборки. Одним из важнейших понятий интервальной статистики

является понятие накрывающего измерения (замера). Согласно [1]

Определение 2.1.1 *Накрывающее измерение (накрывающий замер) — это интервальный результат измерения, который гарантированно содержит истинное значение измеряемой величины.*

Измерение, для которого нельзя утверждать, что оно содержит истинное значение измеряемой величины, будем называть ненакрывающим.

Накрывающее измерение является гарантированной двусторонней «вилкой» значений измеряемой величины, тогда как для ненакрывающего измерения ничего подобного утверждать нельзя. Это придаёт накрывающим измерениям принципиально более высокий статус и позволяет строить на их основе более содержательные конструкции.

При перенесении свойства «накрытия истинного значения» на выборки простейший путь — объявить «накрывающей выборкой» совокупность накрывающих измерений, а выборки, в которых присутствует хотя бы одно ненакрывающее измерение станут «ненакрывающими».

Далее мы будем называть *накрывающей выборкой* совокупность измерений, в которой доминирующая часть (большинство и т. п.) измерений (наблюдений) являются накрывающими. Напротив, выборка называется *ненакрывающей*, если преобладающая часть входящих в неё измерений — ненакрывающие. Возможные альтернативные термины — «включающее измерение», «охватывающее измерение» (их отрицание — «невключающее», «неохватывающее»). Предлагаемые английские эквиваленты — enclosing measurement, covering measurement.

В настоящей публикации предлагается обобщение классической конструкции меры совместности объектов, рассматриваются её свойства и применение в анализе данных с интервальной неопределённостью.

Принцип соответствия. При формулировании новых определений будем стремиться следовать принципу соответствия. *Принцип соответствия* в методологии науки — это утверждение, что любая новая научная теория должна включать старую теорию и её результаты как частный предельный случай. Применительно к мере совместности, более общие конструкции, которые будут определены далее, содержат как частные случаи ранее определённые конструкции. Символически это можно выразить как

$$\mathbb{KR} \supseteq \mathbb{IR} \supseteq \mathbb{R}.$$

Также важно, чтобы меры, введенные на более общих конструкциях, именно, на выборках интервальных величин и многомерных интервальных объектах, естественным образом переходили в конструкции для более простых объектов.

2.2 Приём варьирования неопределённости

Фундаментальный факт, касающийся интервальных измерений состоит в том, что если некоторое Δ является величиной неопределённости накрывающего интервального измерения или выборки, то и любое Δ' , удовлетворяющее $\Delta' \geq \Delta$, также может служить величиной неопределённости.

Сказанное выше приводит к мысли о том, что при обработке интервальных данных величиной неопределённости можно управлять, виртуально варьируя её, с целью исследования интервальных измерений, их выборок и построения оценок с нужными свойствами. В этом и состоит суть «приёма варьирования неопределённости», вынесенного в заголовок этого параграфа.

При обработке интервальных данных величиной неопределённости можно управлять, виртуально варьируя её, с целью исследования интервальных измерений, их выборок и построения оценок с нужными свойствами. Если выборка интервальных измерений несовместна, то, увеличивая одновременно величину неопределённости всех измерений, мы всегда сможем добиться того, чтобы выборка сделалась совместной, т. е. чтобы пересечение интервалов стало непустым, а интервал минимума по включению (1.5) — правильным. Кроме того, точка (или точки), которая первой появляется в непустом пересечении интервалов при расширении интервальных измерений, и тем самым требует наименьшего увеличения неопределённости измерений для достижения совместности выборки, является «наименее несовместной». Её разумно брать в качестве оценки величины (или оценки параметров зависимости).

Представленный подход будет применён для оценки совместности выборки данных в примере, рассмотренном в разделе §8.1. В конкретной ситуации данных Табл. 8.1, измерения выборки являются существенно неравноширинными. Одновременное изменение величины неопределённости для всех измерений на одно и то же значение может оказаться неразумным. Пусть задан некоторый положительный весовой вектор

$w = (w_1, w_2, \dots, w_n)$, $w_k > 0$, размерность которого равна длине исследуемой выборки, причём изменение величины неопределённости k -го измерения — $\text{rad } \mathbf{x}_k$, должно быть пропорциональным w_k , т. е. для любых k и l справедливо

$$\frac{\text{изменение } \text{rad } \mathbf{x}_k}{\text{изменение } \text{rad } \mathbf{x}_l} = \frac{w_k}{w_l}. \quad (2.1)$$

2.3 Бинарная мера совместности множеств.

Мера сходства двуместная $S(A, B)$ — это вещественная функция между объектами A и B . Формально принадлежность к мерам сходства определяется системой аксиом:

- ограниченность $0 \leq S(A, B) \leq 1$;
- симметрия $S(A, B) = S(B, A)$;
- рефлексивность $S(A, B) = 1 \Leftrightarrow A = B$;
- транзитивность $A \subseteq B \subseteq C \rightarrow S(A, B) \geq S(A, C)$.

Эти свойства также называют t -норма. Существуют и иные системы аксиом сходства. Перечисленные свойства можно реализовать разными способами. Меры включения и сходства широко применяются в практике, начиная со второй половины прошлого века. В публикации Б.И. Сёмкина подробно рассмотрены различные меры сходства и их взаимные связи [12].

Первая подобная мера была предложена Жаккардом в 1901 году применительно к биологии. Коэффициент (индекс) Жаккара для множеств имеет вид

$$S(A, B) = \frac{n(A \cap B)}{n(A \cup B)}. \quad (2.2)$$

где n — мощность (число представителей) множеств.

В компьютерных приложениях (обработка изображений, машинное обучение) меру сходства множеств часто обозначают IoU — Intersection over Union.

Популярное введение в теорию бинарных отношений даётся в книге Ю.А.Шрейдера [13].

2.4 Бинарная мера совместности (покрытия) интервалов.

По мере развития интервального анализа, были введены различные определения и конструкции оценки меры совместности интервальных объектов. Весьма подробно вопрос сравнения интервалов рассмотрен в публикации В.И. Левина [25]. Однако в этой работе рассмотрены только абсолютные меры сравнения.

Вместе с тем, в практике обработки данных часто необходимо оперировать с относительными величинами. В частности, это нужно из-за необходимости сопоставления допусков и размеров деталей, погрешности измерителей и значений измеряемых величин, и т.п.

В публикации S.Kabir et all, [26], вводятся различные меры сходства для объектов нечётких множеств и интервалов, используются обозначения $S_J(A, B)$.

Кроме этого, авторы вводят несимметричную меру *overlapping ratio* для интервала I_i относительно пары интервалов $\{I_i, I_j\}$

$$OR(I_i, I_j) = \frac{|I_i \cap I_j|}{|I_i|}. \quad (2.3)$$

Подходы, развитые в работе [26], получили применение в ряде публикаций, например, [27], [28]. В частности, в работе [27] на основе конструкции (2.3) вводится понятие *сходства*, *similarity* интервалов

$$s = \min(OR(I_i, I_j), OR(I_j, I_i)). \quad (2.4)$$

Мера *несходства*, *dissimilarity* соответственно имеет вид $d = 1 - s$.

Все меры, представленные в перечисленных работах, развивают идею Жаккара, и оперируют только с пересекающимися интервалами. В публикации [29] обсуждаются также и непересекающиеся интервалы, но вычисления меры совместности не производятся.

Обобщение бинарной меры совместности (покрытия) интервалов. Приступим к обобщению меры совместности интервалов, имея целью описывать единообразно как накрывающие, так и не накрывающие выборки. Для анализа данных необходимо проводить сравнение интервальных объектов универсальным образом, независимо от свойства накрытия.

В первую очередь, введём базовую конструкцию совместности для двух интервалов. Для иллюстрации идеи, рассмотрим следующую числовую характеристику степени совместности двух интервалов \mathbf{x}, \mathbf{y}

$$\frac{\text{wid}(\mathbf{x} \wedge \mathbf{y})}{\text{wid}(\mathbf{x} \vee \mathbf{y})}. \quad (2.5)$$

В выражении (2.6) в качестве абсолютной меры величины сходства используется ширина интервала (1.4), а вместо операций пересечения и объединения множеств — операции взятия минимума (\wedge) (1.5) и максимума (\vee) (1.6) по включению двух величин в полной интервальной арифметике (Каухера).

В общем случае, минимум по включению в выражении (2.6) может быть неправильным интервалом. При этом его ширина не определена и нужно использовать либо его правильную проекцию, либо задать нужную конструкцию в явном виде. В связи с этим, выпишем относительную меру покрытия как

$$\text{JK}(\mathbf{x}, \mathbf{y}) = \frac{\min\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\} - \max\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\}}{\max\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\} - \min\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\}}. \quad (2.6)$$

В записи формулы (2.6), вместо ширин интервалов используются явные выражения взятия минимума и максимума по включению, обеспечивающие универсальный характер конструкции, независимо от того, является ли результат операции взятия минимума по включению (\wedge) правильным или неправильным интервалом.

Альтернативным способом записи (2.6) может быть использование в числителе ширины $\text{wid}(\mathbf{x} \wedge \mathbf{y})$, с учётом правильности или неправильности полученного интервала знаком.

$$\text{JK}(\mathbf{x}, \mathbf{y}) = \frac{\text{wid}(\mathbf{x} \wedge \mathbf{y})}{\text{wid}(\mathbf{x} \vee \mathbf{y})}. \quad (2.7)$$

Автору представляется использование явных выражений (2.6) более предпочтительным ввиду дальнейшего обобщения на выборки интервальных величин и общего минимаксного духа арифметики Каухера. В именовании $\text{JK}(\mathbf{x}, \mathbf{y})$ буква J отвечает фамилии Jaccard, а K указывает на арифметику Каухера.

Свойства конструкции JK. Рассмотрим свойства предложенной конструкции (2.6).

В первую очередь, проверим, что для в случае пересекающихся интервалов \mathbf{x} и \mathbf{y} , выражение (2.6) даёт разумные результаты.

Если $\mathbf{x} \cap \mathbf{y} \neq \emptyset$, $\mathbf{x} \wedge \mathbf{y}$ — правильный интервал, в выражении (2.6) можно использовать операции пересечения и объединения множеств

$$\text{JK}(\mathbf{x}, \mathbf{y}) = \frac{\text{wid}(\mathbf{x} \cap \mathbf{y})}{\text{wid}(\mathbf{x} \cup \mathbf{y})}. \quad (2.8)$$

В этом случае, (2.6) переходит в индекс Жаккара (2.2).

В частности, при $\mathbf{x} = \mathbf{y}$, имеем

$$\text{JK}(\mathbf{x}, \mathbf{y}) = 1. \quad (2.9)$$

Свойство (2.9) отвечает максимальной совместности данных \mathbf{x}, \mathbf{y} .

При пересечении интервалов \mathbf{x}, \mathbf{y} в точке, имеем

$$\text{JK}(\mathbf{x}, \mathbf{y}) = 0. \quad (2.10)$$

В случаях (2.8), (2.9) и (2.10) прослеживается удовлетворение мерой (2.6) принципа соответствия.

Перейдём к случаям, не подпадающим под область применения индекса Жаккара в традиционной форме (2.2). В случае $\mathbf{x} \cap \mathbf{y} = \emptyset$, $\mathbf{x} \wedge \mathbf{y}$ — неправильный интервал, числитель (2.6) имеет отрицательное значение.

$$\text{JK}(\mathbf{x}, \mathbf{y}) < 0, \text{ если } \mathbf{x} \cap \mathbf{y} = \emptyset. \quad (2.11)$$

В предельном случае вещественных значений $x \neq y$, имеем

$$\text{JK}(x, y) = -1. \quad (2.12)$$

Суммируя факты (2.8) — (2.12), получаем, что в целом имеет место

$$-1 \leq \text{JK}(\mathbf{x}, \mathbf{y}) \leq 1. \quad (2.13)$$

Таким образом, величина JK непрерывно описывает ситуации от полной несовместности вещественных значений $x \neq y$ до полного перекрытия интервалов $\mathbf{x} = \mathbf{y}$.

Альтернативное определение меры совместности JK. Можно переопределить JK следующим образом

$$\text{JK}_p(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(1 + \text{JK}(\mathbf{x}, \mathbf{y})). \quad (2.14)$$

Индекс p (positive) указывает на положительность конструкции (2.14).

В таком случае, величина JK_p неотрицательна и

$$0 \leq \text{JK}_p(\mathbf{x}, \mathbf{y}) \leq 1. \quad (2.15)$$

При этом $\text{JK}_p = 0$ соответствует случаю вещественных значений $x \neq y$, $\text{JK}_p = \frac{1}{2}$ — пересечению \mathbf{x}, \mathbf{y} в точке, $\text{JK}_p = 1$ — случаю $\mathbf{x} = \mathbf{y}$.

Графическое представление меры JK для двух интервалов.
Приведём иллюстрации различных случаев меры бинарной совместности.

Рассмотрим случай двух интервалов \mathbf{x}, \mathbf{y} , таких что

$$\min\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\} = 1, \quad \max\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\} = 6.$$

На Рис. 2.1 представлены различные примеры относительного положения интервалов \mathbf{x}, \mathbf{y} и соответствующие им значения JK по формуле (2.6).

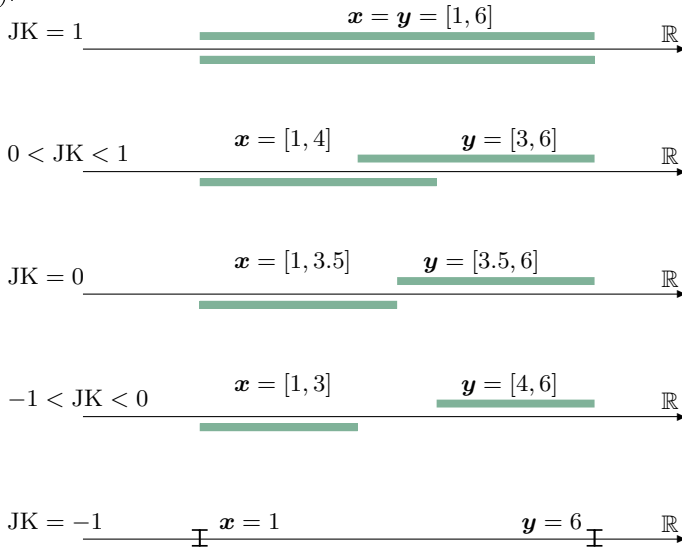


Рис. 2.1. Различные варианты положения интервалов и значения JK.

Несимметричные варианты меры JK. Мера (2.6) симметрична относительно своих аргументов. Это может оказаться неудобным в случае большой разницы ширин аргументов. В таком случае могут оказаться полезными несимметричные варианты.

$$s_x(\mathbf{x}, \mathbf{y}) = \frac{\min\{\overline{\mathbf{x}}, \overline{\mathbf{y}}\} - \max\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\}}{\text{wid } \mathbf{x}}, \quad s_y(\mathbf{x}, \mathbf{y}) = \frac{\min\{\overline{\mathbf{x}}, \overline{\mathbf{y}}\} - \max\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\}}{\text{wid } \mathbf{y}}. \quad (2.16)$$

Подобные конструкции именуют мера сходства («similarity», «overlapping ratio» [26], [27]). Выражение (2.16) отличается от них тем, что допускает неправильность интервала при взятии минимума по включению.

Пример 1 (Вычисление несимметричных мер сходства) Пусть

$$\mathbf{x} = [1, 2], \mathbf{y} = [3, 7].$$

Вычислим меры

$$s_x(\mathbf{x}, \mathbf{y}) = \frac{2 - 3}{1} = -1, \quad (2.17)$$

$$s_y(\mathbf{x}, \mathbf{y}) = \frac{2 - 3}{4} = -\frac{1}{4}. \quad (2.18)$$

Абсолютная величина (2.17) отражает равенство «несходства» операндов относительно ширины первого из них. При этом величина (2.18) существенно меньше ширины 2-го операнда. Этот факт соответствует разнице ширин операндов и позволяет строить на этой основе содержательные конструкции, например, при процедурах регуляризации. ■

2.5 Мера совместности (покрытия) для интервальной выборки.

Для анализа данных полезной характеристикой является мера, описывающая количественно степень совместности накрывающей или ненакрывающей выборки, определение которой введено в §2.1. Мера совместности, введенная в разделе §2.4 для двух интервалов в форме (2.6), допускает естественное обобщение на случай интервальной выборки.

Пусть имеется интервальная выборка $\mathbf{X} = \{\mathbf{x}_i\}$, $i = 1, 2, \dots, n$, где \mathbf{x}_i — интервалы в \mathbb{IR} .

Определим меру $\text{JK}(\mathbf{X})$ на множестве \mathbf{X} как

$$\text{JK}(\mathbf{X}) = \frac{\min_i \bar{\mathbf{x}}_i - \max_i \underline{\mathbf{x}}_i}{\max_i \bar{\mathbf{x}}_i - \min_i \underline{\mathbf{x}}_i}. \quad (2.19)$$

Заметим, что выражение (2.19) переходит в случае интервальной выборки из 2 элементов в выражение (2.6). Таким образом, принцип соответствия выполнен.

Случай ненакрывающей выборки. Приведем численный пример.

Пример 2 (Мера совместности для интервальной ненакрывающей выборки) Рассмотрим пример выборки интервальных данных 3.4.1. из книги [1].

$$\mathbf{X} = \{ [1, 4], [5, 9], [1.5, 4.5], [6, 9] \}. \quad (2.20)$$

Для наглядного представления выборки часто рисуют образующие её интервалы в виде графика, изображённого на Рис. 8.4, который по статистической традиции мы будем называть *диаграммой рассеяния*.

Диаграмма рассеяния выборки \mathbf{X} приведена на рисунке 3.3.

Для выборки \mathbf{X} (3.11) имеем согласно (2.19)

$$\text{JK}(\mathbf{X}) = \frac{4 - 6}{9 - 1} = -0.25.$$

Отрицательность JK говорит о несовместности выборки \mathbf{X} , а значение — о степени несовместности её элементов. ■

Случай накрывающей выборки. Теперь рассмотрим пример накрывающей выборки.

Пример 3 (Мера совместности для интервальной накрывающей выборки) Применим процедуру регуляризации выборки (3.11) методом варьирования величины неопределённости интервальных измерений по Оскорбину. Вычисления проведём с использованием кода `estimate_uncertainty_center` [30] С.И. Жилина.

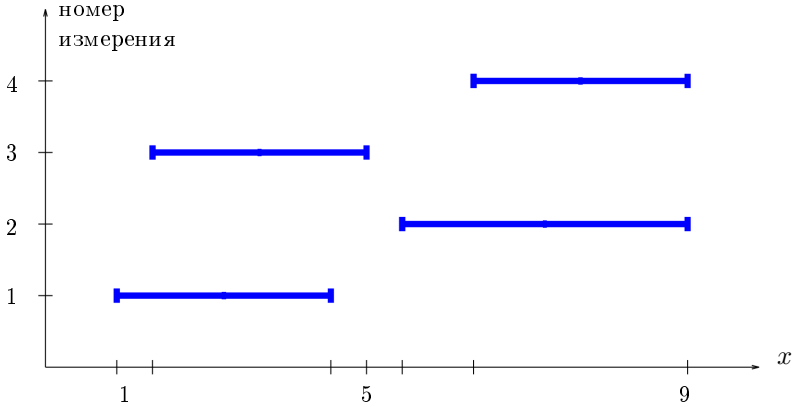


Рис. 2.2. Диаграмма рассеяния интервальной выборки (3.11).

Получаем накрывающую выборку \mathbf{X}^{Cover}

$$\mathbf{X}^{Cover} = \{ [0, 5], [3.6666, 10.334], [0.5, 5.5], [5, 10] \}. \quad (2.21)$$

Мера (2.19) от регуляризованной накрывающей выборки даёт

$$JK(\mathbf{X}^{Cover}) = 0.$$

Нулевое значение $JK(\mathbf{X}^{Cover})$ соответствует пересечению элементов выборки \mathbf{X}^{Cover} в одной точке, конкретно $x = 5$. ■

2.6 Случай многомерных интервалов (брусов) и выборок многомерных интервалов.

Мера совместности, введенная в разделе §2.4 для двух одномерных интервалов в форме (2.6), допускает естественное обобщение на многомерный случай.

В случае многомерности операндов $\mathbf{x} = \{\mathbf{x}_k\}$, $\mathbf{y} = \{\mathbf{y}_k\}$, $k = 1, 2, \dots, m$, в качестве меры совместности имеем вектор JK_k

$$JK_k(\mathbf{x}, \mathbf{y}) = \frac{\min\{\bar{\mathbf{x}}_k, \bar{\mathbf{y}}_k\} - \max\{\underline{\mathbf{x}}_k, \underline{\mathbf{y}}_k\}}{\max\{\bar{\mathbf{x}}_k, \bar{\mathbf{y}}_k\} - \min\{\underline{\mathbf{x}}_k, \underline{\mathbf{y}}_k\}}, \quad k = 1, 2, \dots, m. \quad (2.22)$$

В качестве примера пары $\mathbf{x} = \{\mathbf{x}_k\}$, $\mathbf{y} = \{\mathbf{y}_k\}$ можно привести два бруса последовательных итераций вычислительного процесса решения системы интервальных уравнений.

Если операнды — множества многомерных интервальных векторов $\mathbf{X} = \{\mathbf{x}_k^i\}$, $\mathbf{Y} = \{\mathbf{y}_k^i\}$, $k = 1, 2, \dots, m$, $i = 1, 2, \dots, n$, то конструкция (2.22) переходит в вектор

$$\text{JK}_k(\mathbf{X}, \mathbf{Y}) = \frac{\min_i \bar{\mathbf{x}}_k^i - \max_i \underline{\mathbf{x}}_k^i}{\max_i \bar{\mathbf{x}}_k^i - \min_i \underline{\mathbf{x}}_k^i}, \quad k = 1, 2, \dots, m. \quad (2.23)$$

Операции взятия минимума и максимума в выражении (2.23) проводятся для множеств $\mathbf{X} = \{\mathbf{x}_k^i\}$, $\mathbf{Y} = \{\mathbf{y}_k^i\}$ покомпонентно.

В качестве примера множеств $\mathbf{X} = \{\mathbf{x}_k^i\}$, $\mathbf{Y} = \{\mathbf{y}_k^i\}$ можно привести множества брусов последовательных итераций вычислительного процесса решения системы интервальных уравнений различными методами или с различными начальными приближениями.

2.7 Случай неограниченных интервалов.

Все виды мер совместности обобщаются на случай неограниченных интервалов. Такие обобщения могут быть полезны в тех случаях, когда в постановке задач возникают неравенства.

Интервальная арифметика Кэхэна В первую, очередь, укажем, что в интервальном анализе есть аппарат для работы с неограниченными интервалами. Так называемая интервальная арифметика Кэхэна имеет своими элементами бесконечные и полубесконечные интервалы вида

$$[-\infty, p], \quad [q, +\infty], \quad [-\infty, p] \cup [q, +\infty].$$

В контексте обработки данных арифметика Кэхэна нужна для того, чтобы описывать односторонние ограничения на рассматриваемые величины, к примеру, $x \geq 0$ или $y \leq 1$ и им подобные. Вместо выписанных неравенств можно записать на интервальном языке

$$\mathbf{x} = [0, +\infty] \quad \text{или} \quad \mathbf{y} = [-\infty, 1].$$

Для любых обычных интервалов $\mathbf{a}, \mathbf{b} \in \mathbb{IR}$ результаты сложения, вычитания, умножения и деления \mathbf{a}/\mathbf{b} при $0 \notin \mathbf{b}$ в классической интервальной арифметике и арифметике Кэхэна полностью совпадают. Но в

арифметике Кэхэна дополнительно определено деление интервалов \mathbf{a} и \mathbf{b} с $0 \in \mathbf{b}$, которое и приводит к бесконечным интервалам. Подробнее об интервальной арифметике Кэхэна можно прочитать, к примеру, в книге [4].

Мера совместности на неограниченных интервалах. Перейдем к мере совместности. Базовая конструкция мер совместности (2.5)

$$\frac{\text{wid}(\mathbf{x} \wedge \mathbf{y})}{\text{wid}(\mathbf{x} \vee \mathbf{y})},$$

непосредственно неприменима, ввиду неограниченности знаменателя.

В связи с этим, необходимо использовать несимметричные меры совместности §2.4 (2.16)

$$s_x(\mathbf{x}, \mathbf{y}) = \frac{\min\{\bar{x}, \bar{y}\} - \max\{\underline{x}, \underline{y}\}}{\text{wid } \mathbf{x}}, \quad s_y(\mathbf{x}, \mathbf{y}) = \frac{\min\{\bar{x}, \bar{y}\} - \max\{\underline{x}, \underline{y}\}}{\text{wid } \mathbf{y}},$$

в зависимости от того, какой из аргументов \mathbf{x}, \mathbf{y} неограничен.

Пример 4 (Вычисление меры совместности для неограниченных интервалов) Рассмотрим численные примеры, применяя конструкцию (2.16).

Пусть один из интервалов неограничен и минимум по включению — правильный интервал:

$$\mathbf{x} = [-\infty, 2], \quad \mathbf{y} = [1, 3].$$

$$s_x(\mathbf{x}, \mathbf{y}) = \frac{3 - 1}{\infty} = 0,$$

$$s_y(\mathbf{x}, \mathbf{y}) = \frac{3 - 1}{2} = 1.$$

Пусть оба интервала неограничены, минимум по включению — правильный интервал:

$$\mathbf{x} = [-\infty, 2], \quad \mathbf{y} = [1, +\infty].$$

$$s_x(\mathbf{x}, \mathbf{y}) = s_y(\mathbf{x}, \mathbf{y}) = \frac{2 - 1}{\infty} = 0.$$

Пусть один из интервалов неограничен и минимум по включению — неправильный интервал:

$$\mathbf{x} = [-\infty, -1], \mathbf{y} = [1, 3].$$

$$s_x(\mathbf{x}, \mathbf{y}) = 0,$$

$$s_y(\mathbf{x}, \mathbf{y}) = \frac{-2}{2} = -1.$$

Наконец, оба интервала неограничены, минимум по включению — неправильный интервал:

$$\mathbf{x} = [-\infty, -1], \mathbf{y} = [1, +\infty].$$

$$s_x(\mathbf{x}, \mathbf{y}) = s_y(\mathbf{x}, \mathbf{y}) = \frac{-2}{+\infty} = 0.$$

■

Конструкции, основанные на формулах (2.16), могут быть использованы, например, для характеристики множеств решений систем интервальных уравнений с неравенствами — §5.

Глава 3

Обсуждение меры покрытия в общем контексте анализа интервальных данных.

3.1 Применение JK к твинной арифметике и представление данных измерений с несколькими видами погрешности.

На практике концы интервалов, представляющие результаты измерений, сами могут быть известны неточно, так что возникает необходимость работы с интервалами, имеющими интервальные концы. Такие объекты известны в интервальном анализе и называются *твинами* (по английски *twin*, как сокращение фразы *twice interval*, «двойной интервал»).

Твин, как «интервал интервалов» или интервал с интервальными концами, можно представить как

$$X = [a, b] = [\underline{a}, \bar{a}], [\underline{b}, \bar{b}]. \quad (3.1)$$

Альтернативное рассмотрение твинов как объектов, дающих внут-

реннюю и внешнюю оценку интервальных вычислений, предложено В.М. Нестеровым [35].

$$\mathbf{X} = [\mathbf{X}^{in}, \mathbf{X}^{out}] = [[\underline{\mathbf{X}}^{in}, \overline{\mathbf{X}}^{in}], [\underline{\mathbf{X}}^{out}, \overline{\mathbf{X}}^{out}]]. \quad (3.2)$$

В обоих случаях, твин определяется 4-мя значениями.

Функционал $Rwid$ также зависит от 4-х значений, концов двух интервалов \mathbf{x}, \mathbf{y} . Если взять в качестве аргументов $Rwid$ внутреннюю и внешнюю компоненты твина, получим относительную величину, характеризующую соотношение внутренней и внешней оценки интервального объекта

$$\delta \mathbf{X} = JK(\mathbf{X}^{in}, \mathbf{X}^{out}). \quad (3.3)$$

Пример 5 (Температура как твин) Пример 2.14.2. книги [1].

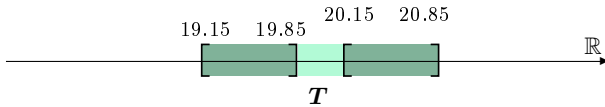


Рис. 3.1. Температура как твин.

Дан твин температуры в форме (3.1)

$$\mathbf{T} = [[19.15, 19.85], [20.15, 20.85]] \text{ } ^\circ C. \quad (3.4)$$

Запишем его в форме Нестерова

$$\mathbf{T} = [[19.85, 20.15], [19.15, 20.85]] \text{ } ^\circ C. \quad (3.5)$$

В соответствии с конструкцией (3.3) имеем «относительную ширину твина» как

$$\delta \mathbf{T} = JK(\mathbf{T}^{in}, \mathbf{T}^{out}) = 0.18. \quad (3.6)$$

Положительность величины $\delta \mathbf{T}$ отражает факт согласованности внутренней и внешней оценок интервального объекта, а абсолютная величина — «запас надёжности» оценки. ■

Вычисления с твинами в форме Нестерова часто приводят к точности внутренней оценки, что снижает информативность таких результатов. В случае допущения $\mathbf{a} \cap \mathbf{b} \neq \emptyset$ в представлении твина в форме

(3.1), можно оперировать с более широким классом вариантов твинов. Такие ситуации весьма часто возникают при измерениях, имеющих как статистическую, так и систематическую погрешность. При этом форма ЖК не изменяется.

Пример 6 (Характеризация твинов) Пример 2.14.1. книги [1].

На рисунке 3.2 графически представлены твины \mathbf{R}_1 и \mathbf{R}_2 . Численные значения концов правого интервала смещены вверх. На сей раз твин \mathbf{R}_1 «уже», чем твин \mathbf{R}_2 , ввиду более широких систематических погрешностей для величины R_2 . Следует заметить также, интервалы $\underline{\mathbf{R}}_2, \overline{\mathbf{R}}_2$ в форме (3.1) имеют ненулевое пересечение. Это пересечение дано более тёмной заливкой, чем концы твина.

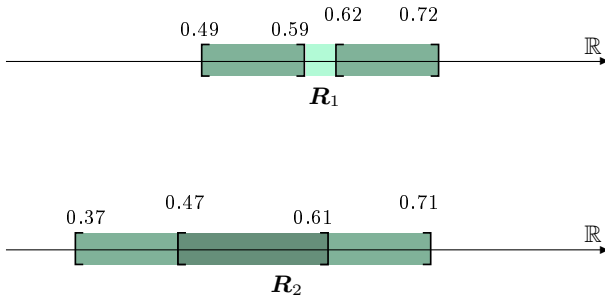


Рис. 3.2. Данные по физике нейтрино в форме твинов типа « \leq ».

В примере для осцилляций нейтрино, мы имели в качестве представления результатов разных экспериментов твины \mathbf{R}_1 и \mathbf{R}_2 :

$$\mathbf{R}_1 = [[0.49, 0.59], [0.62, 0.72]], \quad (3.7)$$

$$\mathbf{R}_2 = [[0.37, 0.61], [0.47, 0.71]]. \quad (3.8)$$

Поступая аналогично вычислениям (3.6), получаем

$$\delta \mathbf{R}_1 = 0.13, \quad (3.9)$$

$$\delta \mathbf{R}_2 = -0.41. \quad (3.10)$$

Отрицательность величины $\delta \mathbf{R}_2$ соответствует непустому пересечению интервалов нижних и верхних оценок величины \mathbf{R}_2 в формуле (3.8). ■

3.2 Место JK в ряду характеристик интервальных выборок.

В традиционной статистике важной характеристикой выборки является её *мода* — значение из выборки, которое встречается наиболее часто. Если же рассматривается случайная величина с непрерывным вероятностным распределением, то её мода — это точка (или точки), в которых плотность вероятности имеет локальный максимум. Выборки и распределения с одной модой называются *унимодальными*, а с двумя и более модами — *мультимодальными*.

Мода лучше и полнее, чем среднее значение (математическое ожидание) характеризует выборки с большим разбросом значений. Кроме того, мода, как характеристика «средней величины», может применяться при обработке данных, имеющих нечисловую природу. Следуя работе [34], введём определение [1]

Модой интервальной выборки назовём совокупность интервалов пересечения наибольших совместных подвыборок рассматриваемой выборки. Наибольшая длина совместных подвыборок данной выборки называется *частотой моды*.

В книге [1] приводится алгоритм для нахождения моды интервальной выборки. Процедура реализована на языке Octave с помощью пакетов [30], [40].

Для облегчения дальнейшего изложения приведём пример.

Пример 7 (Пример вычисления моды интервальной выборки)

Пусть имеется интервальная выборка из 4 элементов

$$\mathbf{X} = \{ [1, 4], [5, 9], [1.5, 4.5], [6, 9] \}. \quad (3.11)$$

Диаграмма рассеяния выборки \mathbf{X} приведена на рисунке 3.3

В соответствии с алгоритмом из книги [1], проверим совместность \mathbf{X} . Пересечение элементов выборки пусто

$$\mathbf{I} = \bigcap_{i=1}^n \mathbf{x}_i = \emptyset.$$

Таким образом, необходимо выполнить основные шаги алгоритма.

Сформируем массив интервалов \mathbf{z} из концов интервалов \mathbf{X}

$$\mathbf{z} = \{ [1, 1.5], [1.5, 4], [4, 4.5], [4.5, 5], [5, 6], [6, 9], [9, 9] \}. \quad (3.12)$$

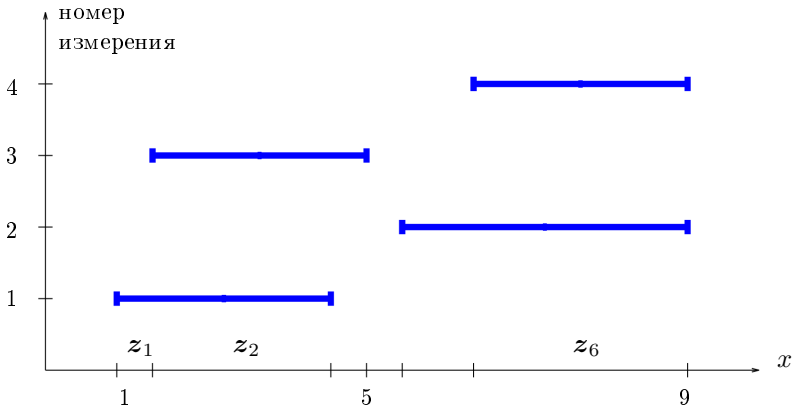


Рис. 3.3. Диаграмма рассеяния интервальной выборки (3.11) и элементы выборки z .

Для каждого интервала z_i подсчитываем число μ_i интервалов из выборки \mathbf{X} , включающих z_i , получаем массив μ_i в виде

$$\{1, 2, 1, 0, 1, 2, 2\}. \quad (3.13)$$

Максимальные μ_i , равные 2, достигаются для индексного множества

$$K = \{2, 6, 7\},$$

так что частота моды равна $\mu = 2$. Как итог, мода является мультиинтервалом

$$\text{mode } \mathbf{X} = \bigcup_{k \in K} z_k = [1.5, 4] \cup [6, 9]. \quad (3.14)$$

На Рис. 3.4 значения частот μ_i (3.13) показаны синим цветом, а интервальная мода $\text{mode } \mathbf{X}$ (3.14) — красным цветом. ■

Важным аспектом введения интервальной моды является соблюдение принципа соответствия §2.1. При уменьшении интервалов неопределённости до нуля, интервальная мода должна перейти в обычную. В частности, в алгоритме вычисления учитывается, что если два интервала имеют пересечение в одной точке, то при вычислении частоты этой точки такой факт учитывается как добавление 2.

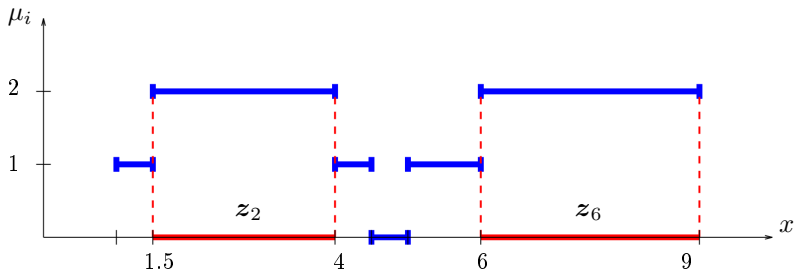


Рис. 3.4. Значения частот μ_i , интервальная мода \mathbf{X} выборки (3.11) и элементы выборки $z_k : k \in K$.

Рассмотрим на примере, как могут соотноситься мода интервальной выборки и мера совместности.

Пример 8 (Мода интервальной выборки и мера совместности)

На Рис. 3.5 представлены различные выборки из двух интервалов $\mathbf{X} = \{x, y\}$.

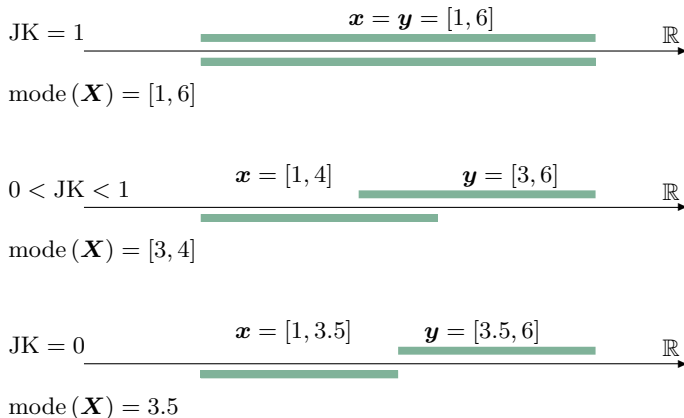


Рис. 3.5. Варианты отношений положения интервалов для накрывающих выборок, интервальная мода и значения JK.



Пример показывает родственность понятий моды интервальной выборки и меры совместности. При этом интервальная мода является сложным объектом, дающим богатую информацию о выборке. Интервальная мода является, в общем случае, мультиинтервалом.

Природа JK иная и носит численный характер. В числителе выражения (2.6) находится минимум по включению, и его положение в случае мультимодовых распределений может не принадлежать ни одному интервалу выборки. В этом случае нужны дополнительные конструкции для численного описания интервальной моды. Предварительно можно заметить, что величина $JK(\text{mode}(\mathbf{X}))$ в случае мультиинтервальности $\text{mode}(\mathbf{X})$ требует трактовки моды как множества интервалов. В таком случае значение $JK(\text{mode}(\mathbf{X}))$ неположительно. Абсолютная величина отвечает степени «изолированности» локальных экстремумов в распределении значений частот μ_i .

Несимметричные варианты JK имеют некоторую аналогию с *условной вероятностью*. С точки зрения обработки данных, их достоинство может заключаться в более обширной информации о взаимном размере двух интервалов, например, в итерационных методах.

Глава 4

Применения меры покрытия для характеризации вычислительных процессов.

4.1 Одномерный случай.

Рассмотрим использование предложенной меры для контроля хода итерационных процессов.

Пример 9 (Метод Ньютона — одномерный случай [31]) Используя интервальный метод Ньютона, найдем корни интервального нелинейного уравнения

$$[2, 2.2]\sqrt{x} + [0.3, 0.4]x - [3, 3.1] = 0. \quad (4.1)$$

Воспользуемся для решения (4.1) интервальным методом Ньютона. Отображение

$$\mathcal{N} : \mathbb{IR} \times \mathbb{R} \rightarrow \mathbb{IR},$$

действующее по правилу

$$\mathcal{N}(\mathbf{X}, \tilde{x}) := \tilde{x} - \frac{\mathbf{f}(\mathbf{a}, \tilde{x})}{\mathbf{f}'(\mathbf{a}, \mathbf{X})}, \quad (4.2)$$

называется (одномерным) *интервальным оператором Ньютона* для f с интервальным параметром a . После выбора $\mathbf{X}^{(0)} := \mathbf{X}$ итерационный процесс уточнения одного корня уравнения $f(\mathbf{a}, x) = 0$ с помощью *интервального метода Ньютона* можно организовать следующим образом:

$$\mathbf{X}^{(k+1)} \leftarrow \mathbf{X}^{(k)} \cap \mathcal{N}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}), \quad \tilde{x}^{(k)} \in \mathbf{X}^{(k)}, \quad k = 0, 1, 2, \dots$$

Зададимся $\mathbf{X}^{(0)} = [1, 2]$ и проведём несколько итераций. Помимо вычисления значений выражений, необходимых для метода Ньютона, на каждом шаге будем вычислять значение «относительную ширину» оператора Ньютона как

$$\delta \mathcal{N}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}) = \text{JK}(\mathcal{N}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}), \mathbf{X}^{(k)}). \quad (4.3)$$

В формуле (4.3) использована симметричная конструкция меры совместности (2.6).

Представим результаты вычисления в таблице 4.1

k	$\mathbf{X}^{(k)}$	$\mathcal{N}(\mathbf{X}^{(k)}, \tilde{x}^{(k)})$	$\delta \mathcal{N}(\mathbf{X}^{(k)}, \tilde{x}^{(k)})$
0	[1.0000, 2.0000]	[1.2076, 1.6991]	0.3487
1	[1.2076, 1.6991]	[1.2345, 1.6903]	0.9640
2	[1.2345, 1.6903]	[1.2329, 1.6894]	0.9999
3	[1.2345, 1.6894]	[1.2330, 1.6894]	1.0000

Таблица 4.1. Итерационное уточнение корня уравнения (4.1) с помощью интервального метода Ньютона, $\tilde{x}^{(k)} = \text{mid } \mathbf{X}^{(k)}$

Из данных таблицы 4.1 видно, что «относительная ширина» оператора Ньютона относительно текущего бруса решения удобно представляет ход итерационного процесса. ■

Также возможно использование конструкции

$$1 - \delta \mathcal{N}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}) = 1 - \text{JK}(\mathcal{N}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}), \mathbf{X}^{(k)}). \quad (4.4)$$

В случае успешных итераций величина $1 - \delta \mathcal{N}(\mathbf{X}^{(k)}, \tilde{x}^{(k)})$ монотонно стремится к 0. При малых значениях (4.4), можно использовать её десятичный логарифм для характеристики скорости сходимости.

4.2 Многомерный случай.

Пример 10 (Метод Ньютона — многомерный случай.) В качестве иллюстрации, рассмотрим пример решения систему интервальных уравнений [2] $f : \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$\begin{pmatrix} x_1^2 + x_2^2 - 1 \\ x_1 - x_2^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (4.5)$$

Решим систему (4.5) методом Кравчика. Отображение $\mathcal{H} : \mathbb{D} \times \mathbb{R} \rightarrow \mathbb{D} \times \mathbb{R}$, задаваемое выражением

$$\mathcal{H}(\mathbf{X}, \tilde{x}) := \tilde{x} - \Lambda \cdot F(\tilde{x}) - (I - \Lambda \cdot L) \cdot (\mathbf{X} - \tilde{x}) \quad (4.6)$$

называется *оператором Кравчика на \mathbb{D} относительно точки \tilde{x}* .

Дальнейшее уточнение решения можно организовать совершенно так же, как и для интервального метода Ньютона, положив $X^{(0)} = X$ и запуская итерации

$$X^{(k+1)} \leftarrow X^{(k)} \cap \mathcal{H}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}),$$

$k = 0, 1, 2, \dots$ для какой-то последовательности точек $\tilde{x}^{(k)} \in \mathbf{X}^{(k)}$.

Система уравнений (4.5) описывает пересечение круга с единичным радиусом и параболы, проходящей через начало координат. Возьмем в качестве первого приближения квадрат, содержащий часть окружности в первом ортанте. $\mathbf{X}^{(0)} = [0, 1] \times [0, 1]$.

Ход вычислительного процесса подробно описан в [2]. Первые шаги вычислений приведены в таблице 4.2.

Численные эксперименты показали, что использование симметричной конструкции меры совместности (2.6) приводит к трудно интерпретируемым результатам. В отличие от одномерного случая, мера ведёт себя немонотонно. Причина этого факта состоит в том, что в процессе итераций изменяется оба аргумента, и оператор Кравчика, и собственно брус приближения, причем по-разному для разных проекций.

В связи с этим, на каждом шаге будем вычислять «относительную ширину» оператора Кравчика в несимметричной форме (2.16)

$$\delta \mathcal{K}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}) = \text{JK}_K(\mathcal{K}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}), \mathbf{X}^{(k)}). \quad (4.7)$$

k	\mathbf{X}_1	\mathbf{X}_2	$\delta \mathcal{K}_1$	$\delta \mathcal{K}_2$
0	[-0.000,1.000]	[0.125,1.000]	0.667	0.583
1	[-0.000,1.000]	[0.274,1.000]	0.790	0.645
2	[0.112,1.000]	[0.406,1.000]	0.865	0.737
3	[0.266,0.974]	[0.534,1.000]	1.000	0.902
4	[0.409,0.827]	[0.643,0.930]	1.000	1.000
5	[0.542,0.694]	[0.735,0.837]	1.000	1.000
6	[0.608,0.628]	[0.780,0.793]	1.000	1.000
7	[0.618,0.618]	[0.786,0.786]	1.000	1.000
8	[0.618,0.618]	[0.786,0.786]	1.000	1.000

Таблица 4.2. Итерационный процесс решения интервальной системы уравнений (4.5).

Данные таблицы 4.2 отражают несколько этапов итерационного процесса, которые в какой-то степени присущи классу задач, в которых изначально не выполнены условия сходимости по спектральному радиусу.

Ход итераций в такой ситуации происходит зачастую следующим образом. Вначале сходимость идёт по компонентам, которым соответствуют собственные числа, меньше 1. В то же время, для других компонент идёт увеличение бруса \mathcal{K} . За счёт пересечения с брусом предыдущего решения, эти компоненты не изменяются. Этот этап можно охарактеризовать как «грубую локализацию» области решения. В конкретном случае, этот этап охватывает итерации 1-3.

После 3-го шага итераций брус оператора Кравчика становится подмножеством бруса начального приближения $\mathcal{K}(\mathbf{X}^{(k)}, \tilde{x}^{(k)}) \subset \mathbf{X}^{(0)}$, и происходит быстрая сходимость к брусу решения. Далее процесс уменьшения брусков на каждой итерации идёт на порядки медленнее, размер брусков практически не изменяется, и в конечном счёте, после колебаний, связанных с точностью вычислений разностей близких величин, $\delta \mathcal{K}_i \rightarrow 1$.

На Рис. 4.1 на верхнем графике красным цветом показаны брусы оператора Кравчика на первых итерациях, синим цветом представлен

Брусы оператора Кравчика

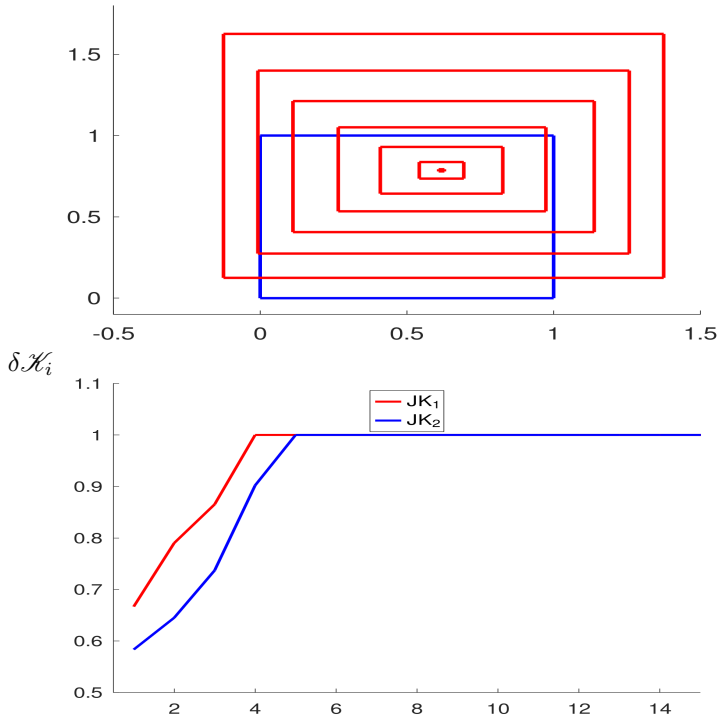


Рис. 4.1. Брусы оператора Кравчика и графики компонент $\delta \mathcal{K}_i$.

брус начального приближения. На нижнем рисунке — графики компонент вектора $\delta \mathcal{K}$. ■

Приведённые примеры показывают, что мера совместности может быть применена для контроля вычислительного процесса при решении интервальных уравнений и систем интервальных уравнений. При этом вид конструкции, наглядно иллюстрирующей вычислительный процесс, не определён однозначно.

Глава 5

Характеризация множеств решений систем интервальных уравнений

5.1 Множества решений систем интервальных уравнений

В интервальном анализе чрезвычайно важную роль имеет понятие множества решений. Приведем необходимые определения и результаты.

Для описания типов множеств решений используется кванторный формализм. При этом различные типы неопределенностей описываются логическими кванторами всеобщности и существования.

Определение 5.1.1 *Выделяющий предикат имеет АЕ-форму, если в нем все вхождения квантора всеобщности « \forall » предшествуют вхождениям квантора существования « \exists ».*

Определение 5.1.2 *Множества решений интервальных уравнений (неравенств и т. п.), для которых выделяющий предикат имеет АЕ-форму, называются множествами АЕ-решений (или АЕ-множествами решений).*

Для интервальных линейных систем вида (5.1) принято выделять три частных случая, представляющие собой крайние точки всего семейства множеств АЕ-решений:

- *объединённое множество решений*

$$\Xi_{uni}(\mathbf{A}, \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\exists A \in \mathbf{A})(\exists b \in \mathbf{b})(Ax = b)\}, \quad (5.4)$$

- *допускосвое множество решений*

$$\Xi_{tol}(\mathbf{A}, \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\forall A \in \mathbf{A})(\exists b \in \mathbf{b})(Ax = b)\}, \quad (5.5)$$

- *управляемое множество решений*

$$\Xi_{ctl}(\mathbf{A}, \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\forall b \in \mathbf{b})(\exists A \in \mathbf{A})(Ax = b)\}. \quad (5.6)$$

Традиционно разрешимость ИСЛАУ, иначе — непустоту множеств решений исследуют с помощью техники *распознающих функционалов*. Распознающие функционалы строят таким образом, что их знак указывает на непустоту множеств решений.

Рассмотрим, как можно поставить в соответствие различным множествам решений ИСЛАУ соответствующие конструкции сходства векторов \mathbf{Ax} и \mathbf{b} . Будем использовать теорему (5.2) и определения множеств решения (5.4) - (5.6).

В случае объединённого множества решений векторы \mathbf{Ax} и \mathbf{b} имеют ненулевое пересечение в обычном смысле классической интервальной арифметики. В связи с этим имеем соответствие или тест принадлежности интервала \mathbf{x} объединённому множествку решений:

$$\mathbf{x} \in \Xi_{uni}(\mathbf{A}, \mathbf{b}) \Leftrightarrow \mathbf{x} : \min_i \text{JK}((\mathbf{Ax}, \mathbf{b})_i) > 0. \quad (5.7)$$

Для допускосого и управляемого множества решений необходимо использовать несимметричные меры совместности (2.16).

В случае допускосого множества решений имеет место включение вида $\mathbf{Ax} \subseteq \mathbf{b}$.

$$\mathbf{x} \in \Xi_{tol}(\mathbf{A}, \mathbf{b}) \Leftrightarrow \mathbf{x} : \min_i \text{JK}_A((\mathbf{Ax}, \mathbf{b})_i) = 1. \quad (5.8)$$

В случае управляемого множества решений имеет место включение вида $\mathbf{Ax} \supseteq \mathbf{b}$.

$$\mathbf{x} \in \Xi_{ctl}(\mathbf{A}, \mathbf{b}) \Leftrightarrow \mathbf{x} : \min_i \text{JK}_b((\mathbf{Ax}, \mathbf{b})_i) = 1. \quad (5.9)$$

Соотношения (5.7) - (5.9) должны быть выполненными для всех уравнений ИСЛАУ (5.2), в связи с чем берется минимум по всем уравнениям. С помощью перечисленных конструкций можно строить распознающие функционалы для различных типов множеств.

Нахождение допускового множества решений ИСЛАУ. Например, для нахождения допускового множества решений ИСЛАУ можно использовать как распознающий функционал конструкцию

$$\text{Tol}_{JK}(\mathbf{x}; \mathbf{A}, \mathbf{b}) = \min_i \{ \text{JK}_A((\mathbf{A}\mathbf{x}, \mathbf{b})_i) - 1 \}. \quad (5.10)$$

Положительность выражения (5.10) равнозначна непустоте допускового множества, а интервалы \mathbf{x} , доставляющие его положительность (5.11), образуют само допусковое множество

$$\text{Tol}_{JK}(\mathbf{x}; \mathbf{A}, \mathbf{b}) \geq 0. \quad (5.11)$$

Приведем численный пример.

Пример 11 (Допусковое множество ИСЛАУ и график функционала Tol_{JK})

Рассмотрим ИСЛАУ [31]:

$$\begin{pmatrix} [2, 5] & [1, 2] \\ [-7, -5] & [6, 7] \end{pmatrix} \mathbf{x} = \begin{pmatrix} [2.1153, 4.8847] \\ [6.1153, 8.8847] \end{pmatrix}.$$

С помощью программ пакета И.Шарой IntLinInc2D на языке Octave [38] найдем допусковое множество Tol , и вычислим область значений $\text{JK}_A = 1$ (5.10). Они представлены на Рис. 5.1.

Рис. 5.1 демонстрирует равнозначность вычисления допускового множества с помощью пакета И.Шарой IntLinInc2D [38] и несимметричной меры совместности. ■

Нахождение управляемого множества решений ИСЛАУ. Аналогично, для нахождения управляемого множества можно использовать как распознающий функционал конструкцию

$$\text{Ctl}_{JK}(\mathbf{x}; \mathbf{A}, \mathbf{b}) = \min_i \{ \text{JK}_b((\mathbf{A}\mathbf{x}, \mathbf{b})_i) - 1 \}. \quad (5.12)$$

Положительность выражения (5.12) равнозначна непустоте управляемого множества,

$$\text{Ctl}_{JK}(\mathbf{x}; \mathbf{A}, \mathbf{b}) \geq 0, \quad (5.13)$$

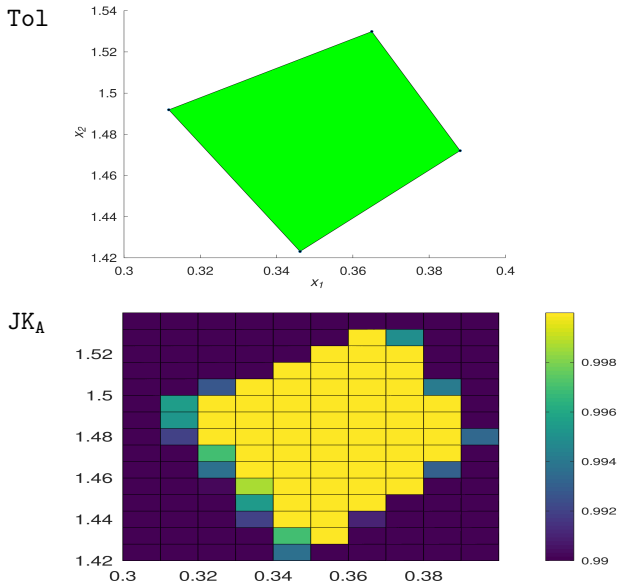


Рис. 5.1. Допусковое множество и график функционала J_{K_A} .

а интервалы \mathbf{x} , доставляющие его положительность, образуют само управляемое множество.

Нахождение объединённого множества решений ИСЛАУ.

Для объединённого множества решений имеем

$$\text{Uni}_{JK}(\mathbf{x}; \mathbf{A}, \mathbf{b}) = \min_i \{JK((\mathbf{A}\mathbf{x}, \mathbf{b})_i)\}, \quad (5.14)$$

и условие на знак

$$\text{Uni}_{JK}(\mathbf{x}; \mathbf{A}, \mathbf{b}) \geq 0, \quad (5.15)$$

В данном случае используется симметричная мера совместности JK.

Пример 12 (Объединённое множество ИСЛАУ и график функционала Uni_{JK})

Снова рассмотрим ИСЛАУ:

$$\begin{pmatrix} [2, 5] & [1, 2] \\ [-7, -5] & [6, 7] \end{pmatrix} \mathbf{x} = \begin{pmatrix} [2.1153, 4.8847] \\ [6.1153, 8.8847] \end{pmatrix}.$$

С помощью программ пакета И.Шарой IntLinInc2D [38] найдём объединённое множество решений Uni , и вычислим область значений $JK \geq 0$ (5.14). Они представлены на Рис. 5.2.

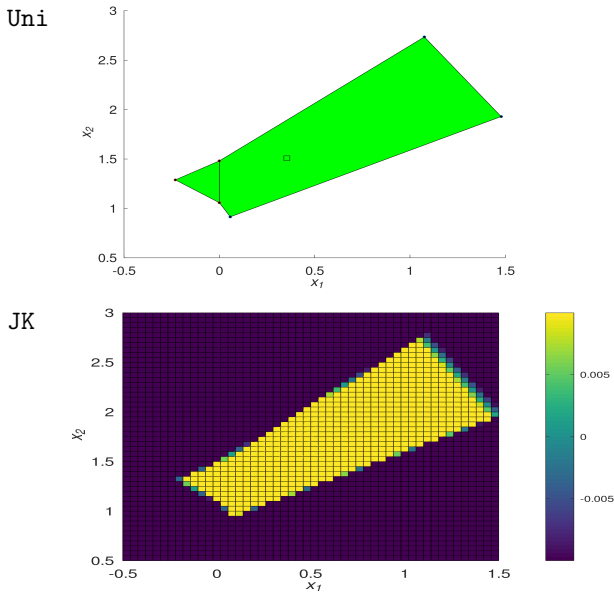


Рис. 5.2. Объединённое множество и график функционала JK.

Рис. 5.2 демонстрирует равнозначность вычисления допускового множества с помощью пакета И.Шарой IntLinInc2D [38] и симметричной меры совместности. ■

Дополнительно заметим, что использование мер совместности выполняется в арифметике Каухера. Поэтому возможны постановки задач смешанного типа, так что по разным уравнениям можно ставить условия, отвечающим как допусковому, так управляемому множествам решений.

Глава 6

Применения меры покрытия в интервальном анализе данных.

6.1 Характеризация области значения переменной

Значительная часть понятий, используемых в интервальном анализе данных, имеет аналогии в традиционной статистике. Таковы, например, мода, медиана, и т.п. Согласно принципу соответствия, при переходе от интервальных данных к точечным, оценки, полученные в интервальном подходе, должны переходить в соответствующие оценки для вещественных данных.

Вместе с тем, интервальный характер объектов может порождать конструкции, которые не имеют непосредственного аналога для точечных данных или имеют другую специфику. Рассмотрим пример интервальных данных 3.2.1. из книги [1], представленный на Рис. 6.1 .

В эксперименте [32] температура в установке измерялась двумя датчиками. Сама температура служила параметром, характеризующим условия проведения испытаний. Таким образом, этот параметр необходимо представить численно для проведения различных вычислений.

Помимо этого, исследователю желательно иметь критерий, чис-

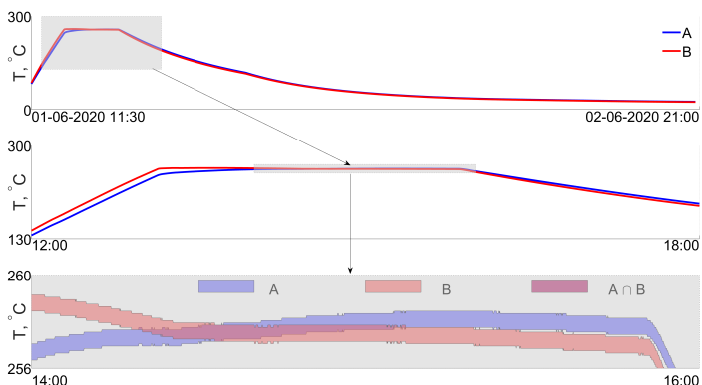


Рис. 6.1. Измерения температуры двумя датчиками.

На верхнем рисунке — данные за весь период проведения эксперимента, около 2-х суток. Серым залит временной диапазон среднего графика. На среднем рисунке — данные в окрестности стационарного режима, серая заливка — временной диапазон нижнего графика. На нижнем рисунке — данные за 2 часа проведения испытаний. Совместность показаний обоих датчиков, т. е. $A \cap B \neq \emptyset$, имеет место около 1 часа.

ленно характеризующий валидность рабочего параметра. Конкретно в представленном исследовании необходимо выбрать область данных, или отрезок времени, в течение которого проведение эксперимента осмыслено и надёжно. Попробуем охарактеризовать «надёжность» интересующего нас параметра численно. Причём желательно, чтобы этот параметр характеризовал не только «надёжную» область, но и все остальные. Вернёмся к примеру.

Пример 13 (Характеризация области значения переменной) — пример 3.2.1. книги [1]

График значений функционала ЖК для данных Рис. 6.1 представлен на Рис. 6.2. Для большей части данных величина ЖК отрицательна, что отвечает несовместности показаний датчиков на этапах разогрева и остывания. Как видно и из Рис. 6.1, значительную часть времени показания однотипных датчиков несовместны.

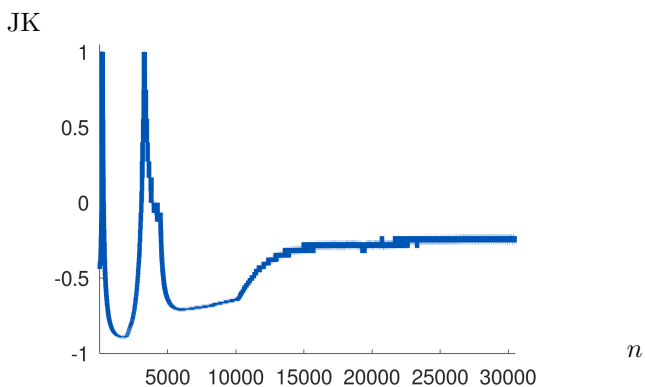


Рис. 6.2. График значений функционала JK для данных Рис. 6.1.

На относительно небольшом отрезке времени, JK имеет значения около нуля и положительные. Рассмотрим этот отрезок подробнее, ограничив его температурами от 256 до 260 С. Для этого, из более чем 30000 отсчетов выберем интервал [2500, 4500].

Температура, С

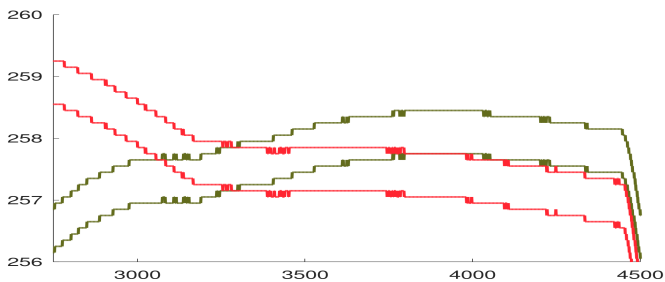


Рис. 6.3. Графики температуры для области стабилизации.

Как видно из Рис. 6.3, даже и в узком диапазоне область перечисления коридоров данных невелика, примерно в области [3000, 3500] замеров.

Результаты вычисления величины JK в этом диапазоне представлено на Рис. 6.4. Сопоставление графиков температуры и JK в области стабилизации даёт основание позиционировать область параметров с положительными или немного отрицательными значениями функционала JK как отвечающими наиболее представительному участку вре-

мени для проведения достоверных измерений.

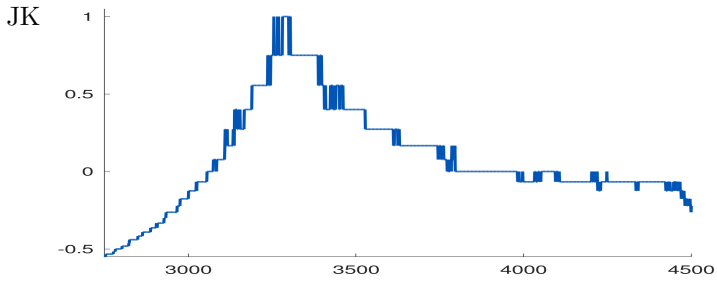


Рис. 6.4. Графики ЖК для области стабилизации.



Интересно сопоставить полученный результат с подобным вычислением для контрольного измерения температуры в лаборатории.

Пример 14 (Характеризация области значения переменной — контрольные измерения.) Контрольные измерения также проводились довольно заметное время, с захватом ночного. Графики температуры представлены Рис. 6.5. В отличие от предыдущего примера,

Температура, С

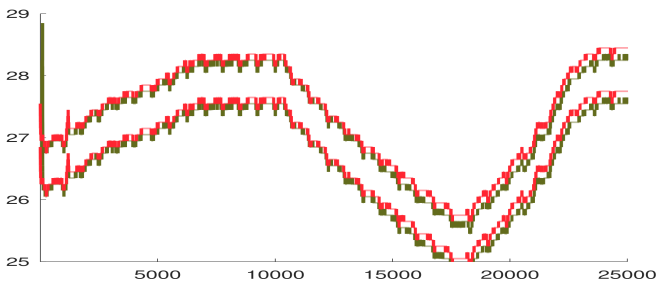


Рис. 6.5. Графики температуры для контрольного измерения.

в течение всего эксперимента значения датчиков совместны в течение всего эксперимента.

Значения ЖК для контрольного измерения близки к единице в первой половине эксперимента и снижаются к концу измерений.

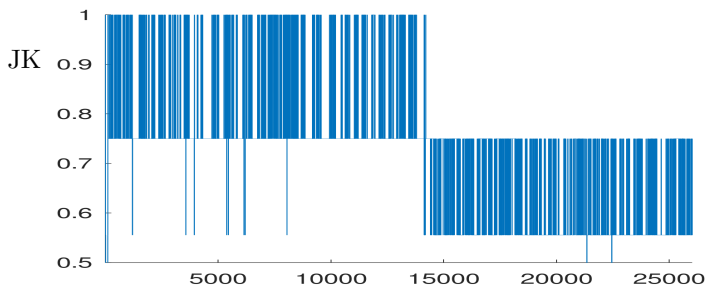


Рис. 6.6. Графики температуры и ЖК для контрольного измерения.



Подводя итог рассмотрению, можно констатировать, что величину и стабильность ЖК можно использовать как индикатор надёжности измерений при использовании нескольких датчиков или измерительных каналов. Например, постепенное снижение величины ЖК говорит о снижении надёжности при весьма длительных измерениях.

6.2 Оценка степени совместности данных при восстановлении зависимостей

Восстановление зависимостей — типичная задача при обработке экспериментальных данных [1]. Получение необходимых для этого надёжных результатов требует многократного измерения одного и того же процесса в некотором диапазоне параметров. В ряде случаев желательно измерение зависимости при переборе входного параметра как в прямом, так и обратном направлении. Если такие данные имеются, встаёт вопрос о степени их согласованности. Рассмотрим пример.

Пример 15 (Оценка степени совместности данных при восстановлении зависимостей) Рассмотрим набор данных движения вала «вперёд-назад» под управлением шагового двигателя [33]. В таблице 6.1 приведены показания энкодера (датчика угла).

Конечной целью измерений является построение линейной зависимости типа

$$\mathbf{X}\beta = \mathbf{Y}, \quad (6.1)$$

где \mathbf{X} — данные контролера управления шагового двигателем, β — параметры модели, \mathbf{Y} — данные датчика.

Код управления	Данные энкодера
0	30
64	30
128	26
192	24
256	17
320	11
384	7
448	0
384	6
320	7
256	11
192	14
128	20
64	25
0	29

Таблица 6.1. Выборка данных движения «вперёд-назад». Точка останова соответствует коду управления 448.

Вообще говоря, данные датчика \mathbf{Y} являются твинами §3.1 (3.1). Каждому значению управляющего воздействия соответствует 4 числовых значения выходного параметра. Например, первый элемент выборки можно представить как $\mathbf{Y}_1 = [[28, 30], [29, 31]]$.

Поскольку на настоящее время нет доступных средств решения твинных уравнений, запишем данные движения «вперёд-назад» из Табл. 6.1 в виде векторов с учётом погрешности энкодера

$$\mathbf{Y}_{up} = \{ [28, 30], [24, 26], [19, 21], [13, 15], [10, 12], [6, 8], [5, 7] \}, \quad (6.2)$$

$$\mathbf{Y}_{back} = \{ [29, 31], [29, 31], [25, 27], [23, 25], [16, 18], [10, 12], [6, 8] \}. \quad (6.3)$$

Ни верхнем графике Рис. 6.7 представлены векторы $\mathbf{Y}_{up}, \mathbf{Y}_{back}$. На

нижнем графике представлены элементы вектора JK , вычисленного по формуле (2.22).

Данные датчика

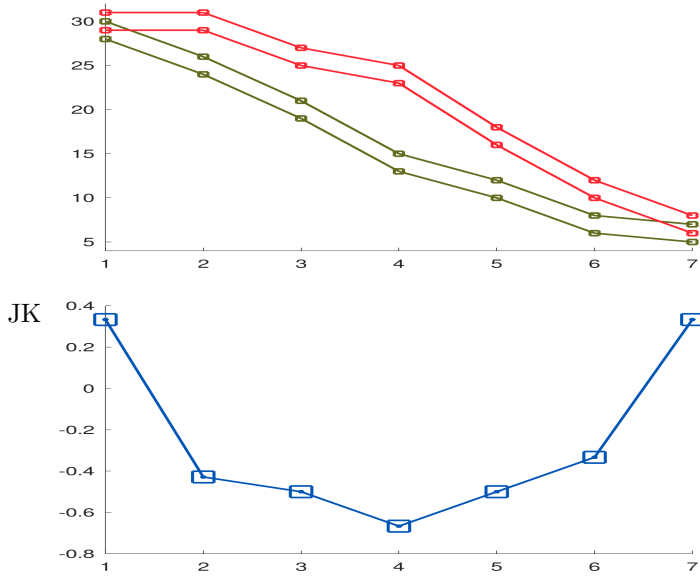


Рис. 6.7. Графики данных датчика и вектор JK .

Как видно из данных, положительные значения имеют первый и последний элемент вектора JK_k . Наименьшее значение имеет измерение в середине рабочего диапазона $\text{JK}_4 = -0.67$. ■

Глава 7

Построение гистограмм интервальных данных

Рассмотрим вопрос построения гистограмм интервальных данных. Напомним, что гистограмма — способ представления табличных данных в виде столбчатой диаграммы. В случае построения гистограмм вещественных данных подсчитывается количество попаданий данных выборки в интервалы множества-классификатора, или бины.

В случае интервальных данных интервал представителя множества исходной выборки может попасть не в один бин (интервал) множества-классификатора. Необходимо придать процессу классификации однозначность.

В таком случае могут оказаться полезными несимметричные варианты (2.16)

$$s_x(\mathbf{x}, \mathbf{y}) = \frac{\min\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\} - \max\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\}}{\text{wid } x}.$$

Здесь \mathbf{x} — представители множества исходной выборки \mathbf{X} , \mathbf{y} — представители множества-классификатора (бины) \mathbf{Y} .

Определим условие принадлежности элемента \mathbf{x} бину \mathbf{y} как условие вхождения более половины интервала \mathbf{x} в интервал \mathbf{y} :

$$s_x(\mathbf{x}, \mathbf{y}) > \frac{1}{2}, \quad \mathbf{x} \in \mathbf{X}, \mathbf{y} \in \mathbf{Y}. \quad (7.1)$$

Приведем численный пример.

Пример 16 (Гистограмма интервальной выборки) Рассмотрим пример выборки интервальных данных 3.7.1. из книги [1].

$$\mathbf{X} = \{ [1.5, 3.5], [0.5, 2.5], [1.5, 3.5], [2, 4] \}. \quad (7.2)$$

Пусть множество-классификатор \mathbf{Y} выбрано как

$$\mathbf{Y} = \{ [0.5, 2.25], [2.25, 4] \}. \quad (7.3)$$

Применим условие (7.1) к выборке (7.2) с классификатором (7.3).

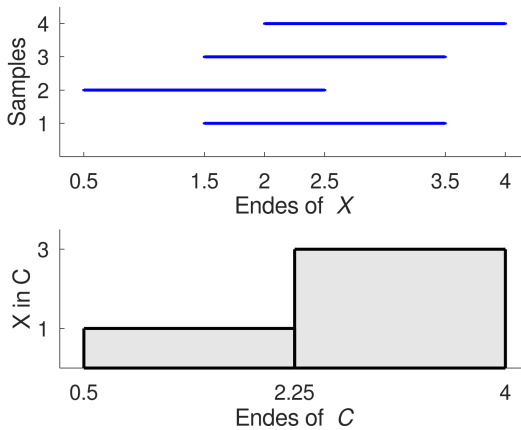


Рис. 7.1. Диаграмма рассеяния и гистограмма интервальной выборки (7.2).

Частоты гистограммы равны

$$N_{\mathbf{X} \in \mathbf{Y}} = \{ 1, 3 \}. \quad (7.4)$$

Диаграмма рассеяния выборки \mathbf{X} и её гистограмма по \mathbf{Y} приведены на рисунке 7.1. ■

Обсуждение вопроса классификации элементов интервальных множеств проведено в книгах [1], [9]. Удобный способ работы с выяснением взаимного отношения элементов интервальных выборок связан с аппаратом интервальных графов и связанных с ними матриц отношений [36], [7].

В случае построения гистограмм интервальных выборок формируется матрица C , элементы которой равны

$$c_{ij} = \begin{cases} 1, & s_x(\mathbf{x}_i, \mathbf{y}_j) > \frac{1}{2}, \\ 0, & \text{иначе.} \end{cases}$$

Производя суммирование элементов c_{ij} построчно, получаем массив частот, образующий гистограмму, представленную на рисунке 7.1.

Размер и строение множества классификатора \mathbf{Y} не является однозначным и зависит от приложения. В частности, в случае вычисления интервальной моды, \mathbf{Y} образуется из вершин элементов \mathbf{X} и проверяется уже вхождение \mathbf{y} в элементы \mathbf{x} . В этом случае условие (7.1) переходит в

$$s_y(\mathbf{x}, \mathbf{y}) = 1.$$

Глава 8

Регуляризация данных на основе меры совместности

Рассмотрим ситуацию, когда имеющаяся выборка не является накрывающей, что было рассмотрено в §2. В книге [1] дан развёрнутый анализ понятия накрытия и рассмотрены различные варианты того как могут соотноситься измерения и истинное (или иное) значение.

Ценность свойства накрытия столь велика, что нередко для его достижения прибегают к специальным приёмам в процессе предобработки данных. Например, там, где это имеет смысл, можно несколько расширить полученные интервалы результатов измерений вокруг базового измеренного значения, чтобы новые расширенные интервалы были гарантированно накрывающими, т. е. содержащими истинные значения измеряемых величин.

Так как проверка того факта, что измерение или выборка являются накрывающими, весьма непросто, можно привлекать косвенный признак накрытия — свойство совместности выборки. Эта подмена особенно привлекательна потому, что многие методы обработки интервальных данных позволяют точно указать величину расширения исследуемых интервалов, необходимую для достижения совместности выборки.

Рассматриваемый в данном разделе приём регуляризации выборки основан на подходе варьирования неопределённости, представленном в §2.2.

8.1 Нахождение оптимального параметра по мере покрытия

Мера покрытия может быть применена для нахождения оптимального параметра. Пусть имеется известная функциональная зависимость

$$F(\mathbf{X}, p) = 0 \quad (8.1)$$

для интервальной выборки \mathbf{X} с параметром p .

Если выборку \mathbf{X} можно модифицировать с использованием параметра p , то в качестве критерия оптимальности можно использовать максимум меры покрытия

$$\hat{p} = \arg \max_p \text{JK}(\mathbf{X}_p) \longrightarrow \max, \quad (8.2)$$

здесь \hat{p} — оптимальная оценка, \mathbf{X}_p — выборка \mathbf{X} , при значении параметра p .

В качестве естественных примеров трансформации выборки можно привести сдвиг нулевого значения всех элементов выборки

$$\mathbf{X}_p = \mathbf{X} + p, \quad (8.3)$$

и масштабное изменение всех элементов выборки

$$\mathbf{X}_p = p \cdot \mathbf{X}. \quad (8.4)$$

К примеру, при наличии опорной выборки \mathbf{X}_0 может потребоваться оценка параметра p при возможном сдвиге нулевого значения или масштабном изменении всех элементов выборки \mathbf{X} . В таком случае необходимо решить задачу оптимизации

$$\hat{p} = \arg \max_p \text{JK}(\mathbf{X}_p, \mathbf{X}_0) \longrightarrow \max. \quad (8.5)$$

8.1.1 Применение коэффициента Жаккара для регуляризации. Случай совместной выборки.

Пример 17 (Применение коэффициента Жаккара для регуляризации. Сравнение рабочей выборки с эталонной)

Исходные данные. Имеется 2 выборки данных с интервальной неопределённостью [42]. Одна из них, \mathbf{X}_2 , относится к эталонному датчику. Другая выборка, \mathbf{X}_1 , соответствует исследуемому датчику. Число отсчетов в выборках равно 200. Калибровка исследуемого датчика производится по эталону. Зависимость между квантовыми эффективностями датчиков предполагается постоянной для каждой пары наборов измерений:

$$\frac{x_2^i}{x_1^i} = \text{const}, \quad i = 1, 2, \dots, 200. \quad (8.6)$$

Диаграммы рассеяния интервальных измерений величины \mathbf{X}_1 и \mathbf{X}_2 приведены на Рис. 8.1.

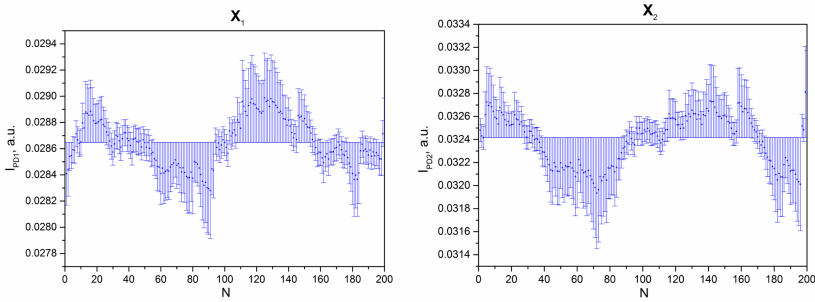


Рис. 8.1. Диаграммы рассеяния интервальных измерений величины \mathbf{X}_1 и \mathbf{X}_2 [6].

Представление данных. Произведём «обинтервализацию» точечных значений, как и ранее, добавив к точечному базовому значению \hat{x} , которое считывается по показаниям измерительного прибора, интервал погрешности ϵ :

$$\mathbf{x}^i = \hat{x}^i + \epsilon^i. \quad (8.7)$$

Интервал погрешности в конкретном случае уравновешен, т. е.

$$\epsilon^i = [-\epsilon^i, \epsilon^i], \quad \epsilon^i = 10^{-4}, \quad i = 1, 2, \dots, 200.$$

Исследование оптимальной коррекции данных. Для проведения регуляризации, зададимся моделью масштабного изменения всех элементов выборки \mathbf{X}_1

$$\mathbf{X}_1 = \mathbf{R} \cdot \mathbf{X}_2. \quad (8.8)$$

Оценка оптимальности произведём по критерию максимума коэффициента Жаккара

$$R_{21}^{opt} = \arg \max_{R_{21}} \text{JK}(\mathbf{X}') \longrightarrow \max, \quad (8.9)$$

здесь \mathbf{X}' — объединение выборок $\mathbf{R} \cdot \mathbf{X}_1$ и \mathbf{X}_2

$$\mathbf{X}' = \mathbf{R} \cdot \mathbf{X}_1 \cup \mathbf{X}_2. \quad (8.10)$$

Интервальная оценка R_{21} соответствует информационному множеству задачи (8.11)

$$R_{21} : \text{JK}(\mathbf{X}') \geq 0. \quad (8.11)$$

Перебор значения множителя коррекции данных производился в диапазоне от 1.0 (нет коррекции) до 1.3 (грубая оценка). Результаты расчётов представлены на Рис. 8.2.

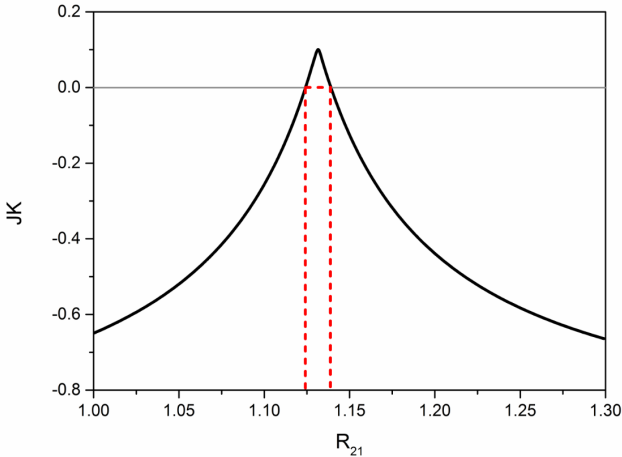


Рис. 8.2. Индекс Жаккара, показано информационное множество значений R_{21} по формуле (8.9).

На графике выделена область, соответствующая условию (8.11). Оптимальное значение R в постановке (8.9) получилось равным

$$R_{21}^{opt} = 1.132,$$

а информационное множество значений (8.11)

$$\mathbf{R}_{21} = [1.12, 1.14].$$

Гистограмма значений объединенной выборки при показана на Рис. 8.3.

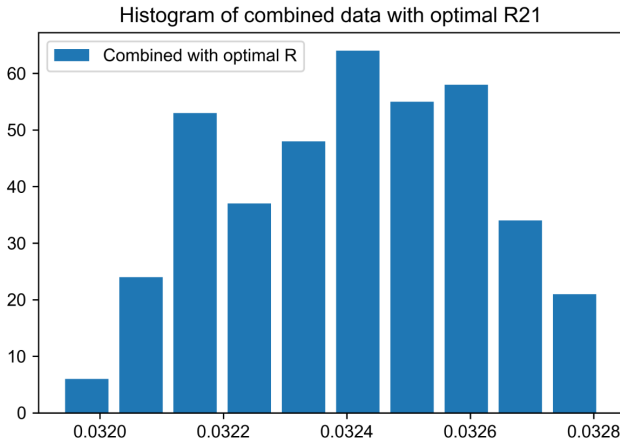


Рис. 8.3. Гистограмма значений объединённой выборки.

Распределение частот гистограмм близко к симметричному и уни-модальному, что положительно свидетельствует о корректности процедуры оптимизации. ■

В приведённом примере конечная выборка \mathbf{X}' (8.10) совместна. Все $2N = 200$ замеров имеют непустые пересечения. Такая ситуация имеет место далеко не всегда. Далее мы рассмотрим более сложный пример, в котором придётся привлечь дополнительные понятия для анализа проблемы и рассмотрения результатов обработки.

8.1.2 Применение коэффициента Жаккара для регуляризации. Случай несовместной выборки.

Пример 18 (Применение коэффициента Жаккара для регуляризации. Сравнение двух выборок)

Постановка задачи. Обработка двух выборок с интервальной неопределённостью. Пример относится к ядерной физике [39] и подробно рассмотрен в публикации [7].

Исходные данные. Данные представлены в 2-5 столбцах таблицы

Номер замера	Compton	std Compton	Peak	std Peak
1	2	3	4	5
1	-3.1	2.7	-4.4	2.7
2	-0.2	2.1	-3.4	1.9
3	-4.0	2.1	-6.9	2.4
4	-2.1	2.5	-1.2	2.4
5	-3.7	1.9	-1.0	2.7
6	-1.7	3.7	-10.8	3.5
7	-5.7	2.8	-10.2	2.8
8	-2.8	1.9	-6.3	2
9	-8.0	4.0	-10.4	4.1
10	-2.1	3.9	0.6	3.4
11	-3.6	2.6	-1.8	2
12	-7.2	2.5	-6.6	2.1
13	—	—	-4.9	2.1
14	—	—	-6.0	2.4
15	—	—	-4.0	2.7

Таблица 8.1. Данные, часть Таблицы 1 из работы [39].

В таблице 8.1 основные данные измерения содержатся в столбцах Peak — средние значения и std Peak — оценки ошибки. В столбцах BG и std BG приведены данные, которые можно использовать для коррекции систематических ошибок. В первом столбце дан условный номер эксперимента.

Как видно из таблицы 8.1, измеряемая величина и фон имеют различные абсолютные значения и знаки. Это говорит о сложном характере отношения сигнала и фона. В рамках исследования знак фона рассматриваться не будет.

Представление данных. В первую очередь представим данные таким образом, чтобы применить понятия статистики данных с интер-

вальной неопределённостью.

Один из распространённых способов получения интервальных результатов в первичных измерениях — это «обинтерваливание» точечных значений, когда к точечному *базовому значению* \hat{x} , которое считается по показаниям измерительного прибора, прибавляется *интервал погрешности* ϵ :

$$\mathbf{x} = \hat{x} + \epsilon. \quad (8.12)$$

Интервал погрешности, вообще говоря, может быть произвольным, но если он уравновешен, т. е.

$$\epsilon = [-\epsilon, \epsilon] \quad \text{для некоторого } \epsilon > 0,$$

то иногда для прямых измерений это можно трактовать, как отсутствие систематических погрешностей.

Согласно терминологии интервального анализа, рассматриваемая выборка — это вектор интервалов, или интервальный вектор $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$. Для того, чтобы придать данным таблицы 8.1 необходимую форму, примем, что в качестве элементов \mathbf{x} будут выступать данные

$$\text{mid } \mathbf{x}_k = \text{Peak}(k), \quad \epsilon_k = \text{rad } \mathbf{x}_k = \text{std Peak}(k), \quad k = 1, 2, \dots, 15. \quad (8.13)$$

Из таблицы 8.1 и Рис. 8.4 видно, что элементы выборки *неравноширинными*, поскольку величина неопределённости $\text{rad } \mathbf{x}_k$ меняется в зависимости от измерения выборки, $k = 1, \dots, n$.

Проведём вычисление некоторых величин, характеризующих выборку интервальных данных. Подробно этот анализ представлен в публикации [7].

Информационным множеством в случае оценивания единичной физической величины по выборке интервальных данных будет также интервал, который называют *информационным интервалом*. Неформально говоря, это интервал, содержащий значения оцениваемой величины, которые «совместны» с измерениями выборки («согласуются» с данными этих измерений).

Найдём значение интервальной моды, см. §3.2. Для исходной выборки (см. столбцы 4-5 таблицы 8.1) имеем

$$\text{mode } \mathbf{X} = \{[-6.7, -6.3], [-5.3, -4.5], -3.6, -2.8\}. \quad (8.14)$$

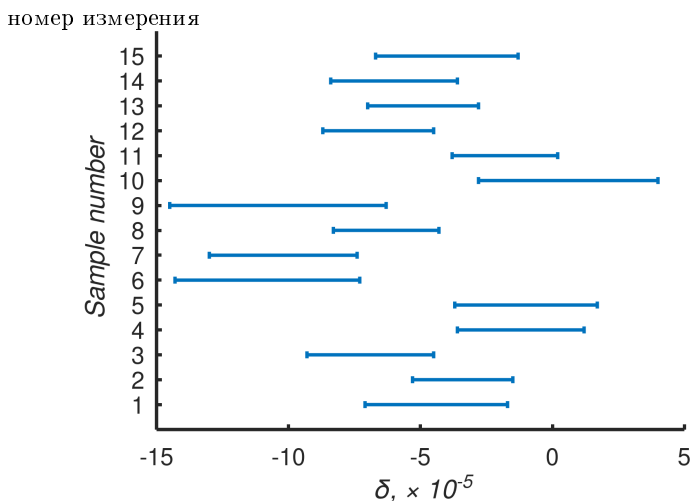


Рис. 8.4. Диаграмма рассеяния интервальных измерений величины [39] для пика полного поглощения, Табл. 8.1.

Поскольку выборка \mathbf{X} несовместна, необходимо дать численные характеристики этого факта. Значение меры совместности

$$JK(\mathbf{X}) = -0.249. \quad (8.15)$$

Вариация неопределенности по Оскорбину, код на языке Octave [30]

$$k = 1.75. \quad (8.16)$$

Граф совместности выборки. Взаимное отношение интервальных измерений в выборке в смысле их совместности друг с другом удобно описывать *графом совместности выборки*, а для исследования структуры выборки прибегать к развитым инструментам теории графов [36].

Напомним, что граф представляет собой математическую абстракцию системы произвольной природы, в которой между составляющими её объектами возникают парные связи. Граф есть совокупность двух множеств — множества объектов, называемого *множеством вершин*, и множества их пар, именуемого *множеством рёбер*. Наглядно граф представляется диаграммой, где вершины отображаются в виде точек, а рёбра — в виде отрезков (или дуг), соединяющих пары вершин.

В графе совместности выборки $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ множество вершин образовано интервалами выборки \mathbf{x}_i , а множество рёбер составлено из пар интервалов $(\mathbf{x}_j, \mathbf{x}_k)$, которые имеют непустое пересечение, то есть совместны друг с другом. Графы, определенные таким способом, в теории графов принято называть *интервальными графами* или *графами интервалов*. Они неплохо изучены и находят много приложений [36].

Поскольку в совместной выборке любые два интервала дают непустое пересечение, то в соответствующем ей графе совместности рёбрами соединены все возможные пары вершин. В теории графов такие графы называются *полными*. Очевидно, что графы совместности для несовместных выборок полными не являются. В них отсутствуют рёбра, соединяющие пары вершин, соответствующие непересекающимся интервалам.

Предметом интереса в несовместных выборках являются максимальные совместные подвыборки. Термин «максимальная» здесь означает, что такую подвыборку невозможно расширить какими-либо дополнительными интервальными измерениями выборки, при сохранении свойства её совместности. При этом несовместная выборка может содержать несколько максимальных совместных подвыборок различной или одинаковой длины.

Нетрудно заметить, что в графе совместности совместным подвыборкам соответствуют полные подграфы, а в случае, когда совместная подвыборка максимальна, — максимальный полный подграф. В теории графов максимальный полный подграф имеет специальное название — «клика».

Следует заметить, что граф совместности выборки может содержать и одиночные вершины, не соединённые ребрами с какими-то другими. Такие изолированные вершины также представляют собой максимальные полные подграфы (клики) и соответствуют интервальным измерениям выборки, не имеющим пересечений с прочими интервалами. Такие вершины и измерения могут рассматриваться как кандидаты в выбросы.

В случае выборки \mathbf{X} имеем 4 набора совместных подмножеств с номерами замеров из Табл. 8.1:

$$\begin{aligned} K_1 &= \{ 1, 3, 8, 9, 12, 13, 14, 15 \}, \\ K_2 &= \{ 1, 2, 3, 8, 12, 13, 14, 15 \}, \\ K_3 &= \{ 1, 2, 4, 5, 11, 13, 14, 15 \}, \\ K_4 &= \{ 1, 2, 4, 5, 10, 11, 13, 15 \}. \end{aligned}$$

Таким образом, размер ни одной из совместных подвыборок не превышает 8, чуть более половины из полной выборки размером 15.

При этом измерения 6 и 7 не входят ни в одну из совместных подвыборок. Однако, нет оснований исключать их из общего набора данных. Как показано в публикации [7], на основании методики *анализа статусов* измерений — см. [1], измерения 6 и 7 нельзя отнести к *выбросам*.

Итоги предварительного анализа данных. Просуммируем факты, выявленные при обработке выборки \mathbf{X} , столбцы 2-3 Табл. 8.1:

- В вычислении физической величины используется часть данных;
- Распределение величины δ мультимодально, максимальное значение частоты покрытия равно 8, что заметно меньше полного числа элементов выборки 15 ;
- Недооценена систематическая погрешность данных, что приводит к необходимости существенной вариации неопределённости измерений (8.16).

Привлечение дополнительных данных из столбцов 2-3. Существуют физические соображения [7], говорящие о наличии связи выборок столбцов 2-3 и 4-5.

$$\delta_0 = k_{BG} \cdot \delta_1. \quad (8.17)$$

Здесь δ_0, δ_1 — данные из 2 и 4 столбцов, k_{BG} - неизвестная мультипликативная поправка.

Исследование оптимальной коррекции данных области комптоновского рассеяния. Перебор значения множителя коррекции данных столбца 2 от 1 (нет коррекции) до 2 (весьма грубая граница) с шагом 0.05.

Результаты расчётов представлены на Рис. 8.5.

Оценка оптимальности произведена по критерию максимума коэффициента Жаккара

$$k_{BG}^{opt} = \arg \max_{k_{BG}} \text{JK} \longrightarrow \max. \quad (8.18)$$

Оптимальное значение множителя коррекции данных области комптоновского рассеяния по (8.18)

$$k_{BG}^{opt} = 1.57. \quad (8.19)$$

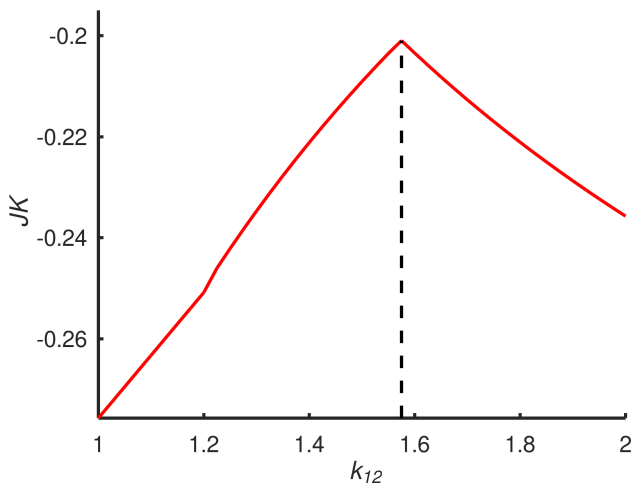


Рис. 8.5. Индекс Жаккара, показан оптимум по формуле (8.17).

Диаграмма рассеяния общего массива данных. На Рис. 8.6 приведена диаграмма рассеяния для исходной выборки (столбцы 4-5) и скорректированных данных (столбцы 2-3), Табл. 8.1.

Вычисление моды интервальной выборки. Вычисление моды интервальной выборки показало, что распределение величины μ_i по всем данным имеет максимум со значением $\max \mu_i = 18$.

Максимальное значение частоты покрытия $\max \mu_i$ равно 18, что составляет 67% от полного числа элементов выборки, 27. Величина моды

$$\text{mode } \mathbf{X} = [-6.7, -6.3]. \quad (8.20)$$

В случае объединённой выборки \mathbf{X} имеем максимальное совместное подмножество с номерами замеров из Табл. 8.1 с номерами:

$$K = \{ 1, 3, 8, 9, 12, 13, 14, 15\}, \\ \{ 16, 18, 19, 20, 21, 22, 23, 24, 25, 26\}.$$

Приём варьирования неопределённости. Проведём варьирование ширин интервалов для достижения совместной выборки.

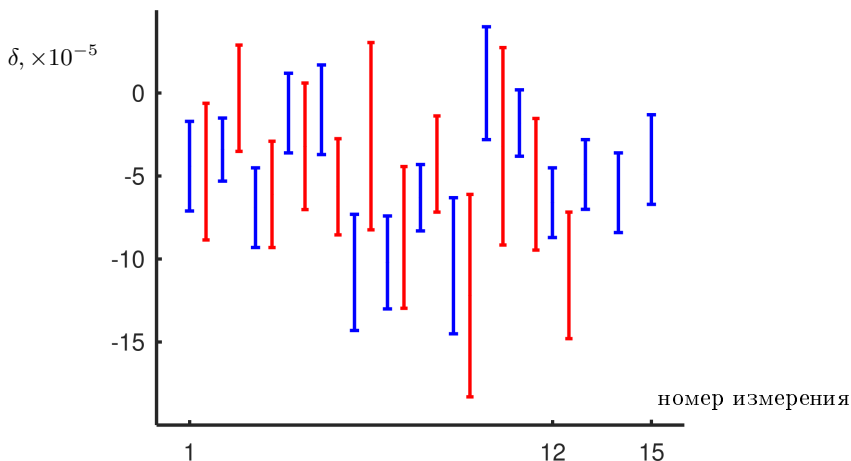


Рис. 8.6. Диаграмма рассеяния интервальных измерений величины для исходной выборки (столбцы 4-5) и скорректированных данных (столбцы 2-3).

Получим значения подобно (8.16)

$$x_{MM} = \text{oskorbin_center} = -5.1, \quad k = 1.57. \quad (8.21)$$

На Рис. 8.8 дано графическое представление интервальных данных и результаты обработки по методике [41]. Красным цветом даны данные объединённой выборки пика полного поглощения и скорректированных данных области комптоновского рассеяния. Черным цветом показаны расширенные согласно (2.1) интервалы.

Анализ данных с расширенной выборкой. Как и в случае исходной выборки, просуммируем факты, выявленные при обработке расширенной выборки X , столбцы 2-3 и 4-5 Табл. 8.1:

- В вычислении физической величины используются все данные, 27 против 15;
- Распределение величины δ унимодально, максимальное значение частоты покрытия равно 18;

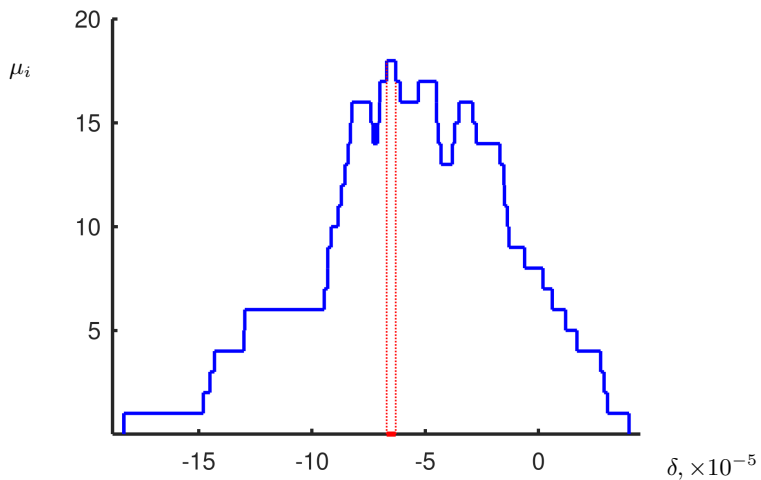


Рис. 8.7. Мода интервальной выборки величины δ для всех данных [39].

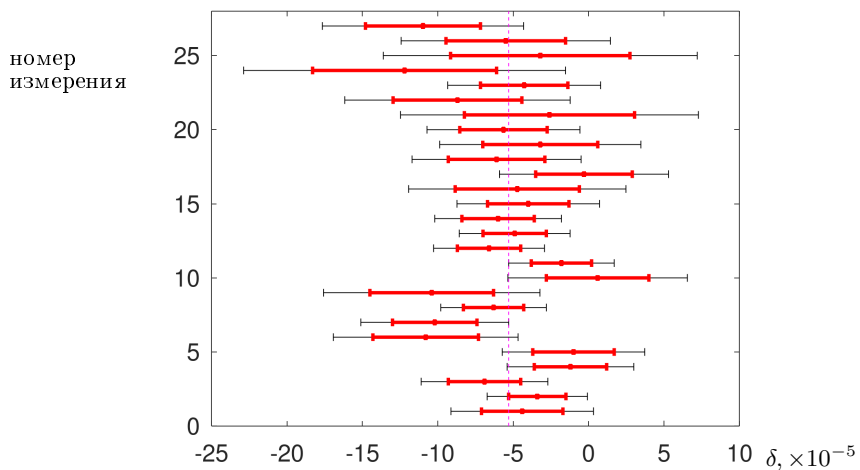


Рис. 8.8. Графическое представление интервальных данных и результаты обработки по методике [41].

- Добавление данных не увеличило несовместности, необходимость вариации неопределённости измерений (8.21) обусловлена подвыборкой данных столбцов 2-3 Табл. 8.1.



8.2 Характеристика вариабельности меры сходства

Вернёмся к началу §8.1 «Нахождение оптимального параметра по мере покрытия».

Если выборку \mathbf{X} можно модифицировать с использованием параметра p , то в качестве критерия оптимальности было предложено использовать максимум меры покрытия (8.2)

$$\hat{p} = \arg \max_p \text{JK}(\mathbf{X}_p) \longrightarrow \max,$$

где \hat{p} — оптимальная оценка, \mathbf{X}_p — выборка \mathbf{X} , при значении параметра p .

Интервальная оценка параметра p соответствует информационному множеству задачи (8.11)

$$p : \text{JK}(\mathbf{X}, p) \geq 0.$$

Рис. 8.5 и Рис. 8.2 представляют характерный вид зависимости $\text{JK}(\mathbf{X}, p)$. Подобная зависимость наблюдается и в других численных экспериментах.

С.И.Жилиным было предложено характеризовать вариабельность меры сходства множеств через «остроту» кривой $\text{JK}(\mathbf{X}, p)$ в области её максимума \hat{p} :

$$\text{EVJK} = \max \left| \frac{\partial \text{JK}(\mathbf{X}_p)}{\partial p} \right|_{p=\hat{p}} \quad (8.22)$$

Впервые идея использования свойств распознающих функционалов для характеристики вариабельности оценки вектора параметров $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m)$ в линейной функции (задача интервальной регрессии), которая получена с помощью метода максимума совместности, была предложена в работе [43].

В книге [1] были предложены новые меры вариабальности для сильного и слабого согласования. В них входят аргумент и значение максимума используемых функционалов следующего вида

$$\max_{\mathbb{R}^n} \text{Tol} \cdot \left\| \arg \max_{\mathbb{R}^n} \text{Tol} \right\|,$$

где Tol и $\arg \max_{\mathbb{R}^n} \text{Tol}$ — конкретные значения в случае сильной зависимости

В духе упомянутых конструкций (8.22) можно записать как

$$\text{EVJK} = \frac{\max |\text{JK}(\mathbf{X}_p)|_{p=\hat{p}}}{\text{wid } p}, \quad (8.23)$$

в случае выполнения условия положительности $\text{JK}(\mathbf{X}_p)|_{p=\hat{p}}$.

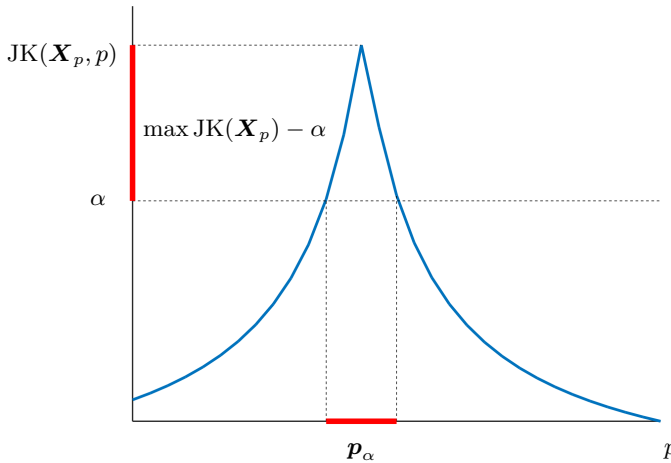


Рис. 8.9. Элементы характеристика вариабальности множества совместности

Если максимальное значение коэффициента совместности неположительно, в качестве знаменателя выражения (8.23) можно взять ширину интервала решения уравнения

$$p_\alpha : \text{JK}(\mathbf{X}_p, p) = \alpha, \quad \alpha < \max \text{JK}(\mathbf{X}_p), \quad (8.24)$$

а в числителе использовать величину $\max \text{JK}(\mathbf{X}_p - \alpha$. В таком случае

$$\text{EVJK} = \frac{\max \text{JK}(\mathbf{X}_p) - \alpha}{\text{wid } \mathbf{p}_\alpha}. \quad (8.25)$$

При $\alpha = 0$, (8.25) переходит в (8.23). В качестве предложения, можно назвать конструкцию (8.25) *характеристика вариабельности множества совместности*.

Исследование содержательности конструкций (8.23) и (8.25) и возможной сферы их применения требует дополнительных исследований.

Заключение

В пособии представлен в значительной степени оригинальный материал. Предложенные меры и различные математические конструкции могут быть использованы для характеристики совместности отдельных интервалов, составных интервальных объектов и интервальных множеств.

Помимо собственно характеристики, меры совместности могут быть применены в вычислительных схемах. Привлекательность мер совместности состоит в простоте их вычисления, поскольку используются только операции взятия минимумов и максимумов.

Существенным достоинством мер совместности, обеспеченным применением полной интервальной арифметики Каухера, является возможность работы с несовместными выборками интервальных данных. Именно в анализе данных с интервальной неопределённостью автору представляется наиболее вероятное использование мер совместности.

Первые опыты применения методики показали, что применение могут иметь как симметричная, так и несимметричные формы меры совместности. Практика покажет, какие из них наиболее полезны в зависимости от области применения.

Автор выражает благодарность за консультации, обсуждения и полезные замечания участникам Всероссийского интервального вебинара, на котором обсуждалась значительная часть материала пособия, С.И. Жилину, С.И. Кумкову, А.В. Пролубникову, С.П. Шарому, Е.В. Чаусовой. Также автор благодарен А.А. Карповой за плодотворное сотрудничество в области интервального анализа.

Литература

- [1] А.Н. БАЖЕНОВ, С.И. ЖИЛИН, С.И. КУМКОВ, С.П. ШАРЫЙ. Обработка и анализ данных с интервальной неопределённостью. 2022
- [2] А.Н. БАЖЕНОВ. Интервальный анализ. Основы теории и учебные примеры: учебное пособие. Санкт-Петербург, 2020.
<https://elib.spbstu.ru/dl/2/s20-76.pdf/info>
- [3] А.Н. БАЖЕНОВ. Естественнонаучные и технические применения интервального анализа. Санкт-Петербург, 2021.
<https://elib.spbstu.ru/dl/5/tr/2021/tr21-169.pdf/info>
- [4] ШАРЫЙ С.П. Конечномерный интервальный анализ. – ФИЦ ИВТ: Новосибирск, 2021. Электронная книга, доступная на <http://www.nsc.ru/interval/Library/InteBooks/SSharyBook.pdf>
- [5] А.Н. БАЖЕНОВ. Введение в анализ данных с интервальной неопределённостью: учебное пособие. Санкт-Петербург, 2022.
- [6] А.Н. БАЖЕНОВ, А.Ю. ТЕЛЬНОВА. Обобщение коэффициента Жаккара для анализа данных с интервальной неопределённостью. Измерительная техника. 2022
- [7] A.N. Bazhenov, O.M Skrekel. A revision of data processing of the circular polarization of the γ -quanta in the $np \rightarrow d\gamma$ reactions with polarized neutrons.
NIM A (to be submitted)
- [8] ШАРЫЙ С.П. Сильная согласованность в задаче восстановления зависимостей при интервальной неопределённости данных // Вычислительные технологии. – 2017. – Т. 2, №2. – С. 150–172.
- [9] NGUYEN H.T., KREINOVICH V., WU B., XIANG G. Computing Statistics under Interval and Fuzzy Uncertainty. Applications to Computer Science and Engineering. – Springer, Berlin-Heidelberg, 2012.
- [10] TUKEY J.W. The Future of Data Analysis // Annals of Mathematical Statistics – 1962. – Vol. 33, Issue 1. – P. 1–67

- [11] КОЭФФИЦИЕНТ ЖАККАРА
https://en.wikipedia.org/wiki/Jaccard_index
 JACCARD P. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines // Bull. Soc. Vaudoise sci. Natur. 1901. V. 37. Bd. 140. S. 241–272.
- [12] Б.И. СЁМКИН. О СВЯЗИ МЕЖДУ СРЕДНИМИ ЗНАЧЕНИЯМИ ДВУХ МЕР ВКЛЮЧЕНИЯ И МЕРАМИ СХОДСТВА. Бюллетень Ботанического сада-института ДВО РАН, 2009. Вып. 3. С. 91-101
- [13] Ю.А. ШРЕЙДЕР. Равенство, сходство, порядок: Популярное введение в теорию бинарных отношений. С примерами из математической лингвистики № 248. Изд. 2 URSS. 2021. 256 с. ISBN 978-5-9710-8453-2.
- [14] KEARFOTT R. B., NAKAO M. T., NEUMAIER A., RUMP S. M., SHARY S. P., VAN HENTENRYCK P. Standardized notation in interval analysis // Вычислительные технологии. 2010. Т. 15. № 1. С. 7–13.
- [15] IEEE Std 1788-2015 – IEEE standard for interval arithmetic
<https://standards.ieee.org/standard/1788-2015.html>
- [16] ШАРЫЙ С.П. Метод максимума согласования для восстановления зависимостей по данным с интервальной неопределённостью // Известия Академии Наук. Теория и системы управления. – 2017. – №6. – С. 3–19.
- [17] ШАРЫЙ С. П. Выявление выбросов в методе максимума согласования при анализе интервальных данных // Сборник трудов Всероссийской конференции по математике с международным участием «МАК-2018». – Барнаул : Изд-во АлтГУ, 2018. – С. 215–218.
 Доступна на <http://elibrary.asu.ru/handle/asu/6303>
- [18] ШАРЫЙ С.П. О мере варибельности оценки параметров в статистике интервальных данных // Вычислительные технологии. – 2019. – Т. 24, №5. – С. 90–108.
- [19] S.P. SHARY Numerical computation of formal solutions to interval linear systems of equations. <https://arxiv.org/abs/1903.10272v1>
- [20] ШАРЫЙ С.П. Задача восстановления зависимостей по данным с интервальной неопределённостью // Заводская лаборатория. Диагностика материалов. – 2020. – Т. 86, №1. – С. 62–74. DOI: 10.26896/1028-6861-2020-86-1-62-74
- [21] ZHILIN, S.I. On fitting empirical data under interval error // Reliable Computing. – 2005. – Vol. 11. – P. 433–442. DOI: 10.1007/s11155-005-0050-3
- [22] ZHILIN S.I. Simple method for outlier detection in fitting experimental data under interval error // Chemometrics and Intelligent Laboratory Systems. – 2007. – Vol. 88, No. 1. – P. 60-68 DOI: 10.1016/j.chemolab.2006.10.004

- [23] Кумков С.И. Обработка экспериментальных данных ионной проводимости расплавленного электролита методами интервального анализа // Расплавы. – 2010. – №3. – С. 79–89.
- [24] KUMKOV, S.I., MIKUSHINA, YU. V. Interval approach to identification of catalytic process parameters // *Reliable Computing*. – 2013. – Vol. 19. – P. 197–214.
- [25] В.И. ЛЕВИН Сравнение интервальных чисел и оптимизация систем с интервальными параметрами, *Автомат. и телемех.*, 2004, № 4, 133–142; *Autom. Remote Control*, 65:4 (2004), 625–633
- [26] KABIR, S., WAGNER, C., HAVENS, T. C., ANDERSON, D. T. AND AICKELIN, U. Novel Similarity Measure for Interval-Valued Data Based on Overlapping Ratio. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2017)*, IEEE. <https://doi.org/10.1109/FUZZIEEE.2017.8015623>.
- [27] T. WILKIN AND G. BELIAKOV, "The Mode of Interval-Valued Data," 2019 *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2019, pp. 1-6, doi: 10.1109/FUZZ-IEEE.2019.8858850.
- [28] Shaily Kabir, Christian Wagner, Timothy C. Havens, Derek T. Anderson, Uwe Aickelin: Novel similarity measure for interval-valued data based on overlapping ratio. *FUZZ-IEEE 2017*: 1-6
- [29] SHAILY KABIR, CHRISTIAN WAGNER, ZACK ELLERBY Towards Handling Uncertainty-at-Source in AI - A Review and Next Steps for Interval Regression. *CoRR abs/2104.07245* (2021)
- [30] С.И.Жилин. Примеры анализа интервальных данных в Octave. Сборник jupyter-блокнотов с примерами анализа интервальных данных. <https://github.com/szhilin/octave-interval-examples>
- [31] А.Н. БАЖЕНОВ, А.А. КАРПОВА Интервальный анализ для исследователей. РХД, Ижевск. 2022.
- [32] А.Н. БАЖЕНОВ, А.Н. КОВАЛЬ, С.Ю. ТОЛСТЯКОВ, Е.Е. МУХИН, А.М. ДМИТРИЕВ, Д.С. САМСОНОВ Стенд для термовакуумных механических испытаний. // *Приборы и техника эксперимента*, 2021, т.1 с.: 151–152.
- [33] Н.В. ЕРМАКОВ, А.Н. БАЖЕНОВ, А.Н. СМИРНОВ, С.Ю. ТОЛСТЯКОВ. Стенд для испытаний шаговых двигателей. // *Приборы и техника эксперимента*. 2022.
- [34] HU C., HU Z.H. On statistics, probability, and entropy of interval-valued datasets // Lesot M.J. et al. (eds) *Information Processing and Management*

- of Uncertainty in Knowledge-Based Systems. IPMU 2020. Communications in Computer and Information Science, vol 1239. – Cham: Springer, 2020.
- [35] НЕСТЕРОВ В.М. Твинные арифметики и их применение в методах и алгоритмах двустороннего интервального оценивания. дисс. д.ф.-м.н. Санкт-Петербург, Санкт-Петербургский институт информатики и автоматизации РАН, 1999, 234 с.
- [36] Робертс Ф.С., Дискретные математические модели с приложениями к социальным, биологическим и экологическим задачам. Изд. "Наука" 1986, 496 с.
- [37] ШАРЫЙ С.П. Алгебраический подход к анализу линейных статических систем с интервальной неопределённостью. // Известия РАН. Теория и системы управления. – 1997. – № 3. – С. 51-61.
- [38] ШАРАЯ И.А. Пакет `IntLinInc2D` для визуализации множеств решений интервальных линейных систем с двумя неизвестными. – Программное обеспечение, доступное на <http://www.nsc.ru/interval/sharaya/>. Описание <http://www.nsc.ru/interval/Programing/MCodes/IntLinInc2D.pdf>
- [39] A.N.Bazhenov, L.A.Grigor'eva, V.V.Ivanov, E.A.Kolomensky, V.M.Lobashev, V.A.Nazarenko, A.N.Pirozhkov, Yu.V.Sobolev. Circular polarization of γ -quanta in $np \rightarrow d\gamma$ reactions with polarized neutrons *Physics Letters B* Volume 289, Issues 1–2, 3 September 1992, Pages 17-21
- [40] С.И.Жилин. Библиотека полной интервальной арифметики `kinterval` в среде `Octave`. Частное сообщение.
- [41] ОСКОРБИН Н.М. Некоторые задачи обработки информации в управляемых системах // Синтез и проектирование многоуровневых иерархических систем. Материалы конференции. – Барнаул: Алтайский государственный университет, 1983.
- [42] М.З.Шварц, рабочие материалы
- [43] ШАРЫЙ С.П. О мере вариабельности оценки параметров в статистике интервальных данных // Вычислительные технологии. – 2019. – Т. 24, №5. – С. 90–108.

Предметный указатель

- t*-норма, 19
- IoU — Intersection over Union, 19
- Mathematica, 11
- Octave, 11, 25, 33, 45, 65
- Python, 11
- аксиома ограниченности, 19
- аксиома рефлексивности, 19
- аксиома симметрии, 19
- аксиома транзитивности, 19
- анализ данных, 11
- базовое значение, 64
- гистограммы интервальных данных, 55
- граф совместности выборки, 65
- диаграмма рассеяния, 25, 33, 56, 64
- интервал погрешности, 64
- интервальная арифметика Каухера (полная), 14, 21, 47
- интервальная арифметика Кэхэна, 27
- интервальная арифметика классическая, 13
- интервальный метод Ньютона, 38
- клика, 66
- коэффициент (индекс) Жаккара для выборки интервальных величин, 25
- коэффициент (индекс) Жаккара для интервальных величин, 21
- коэффициент (индекс) Жаккара для интервальных величин — свойства, 22
- коэффициент (индекс) Жаккара для интервальных величин, альтернативная форма, 23
- коэффициент (индекс) Жаккара для интервальных величин, несимметричная форма, 24, 28, 40, 45, 47, 55
- коэффициент (индекс) Жаккара для множеств, 19
- мера вариабельности, 71
- мера совместности интервалов, 21
- меры совместности для неограниченных интервалов, 27
- меры совместности несимметричные, 24, 28
- множества АЕ-решений, 42
- множество АЕ-решений типа $\mathcal{A}\beta$, 43
- множество решений допусковое, 44
- множество решений объединённое, 44
- множество решений управляемое, 44
- мода выборки, 33
- мультимодальная выборка, 33, 67
- накрывающая выборка, 17, 62

накрывающее измерение, 17
ненакрывающая выборка, 17
ненакрывающее измерение, 17
неопределённость, 12
неравноширинные измерения, 19, 64
оператор Кравчика для нелинейных
уравнений, 39
оператор Ньютона, 38
приём варьирования
 неопределённости, 18, 58

равноширинные измерения, 64
распознающий функционал, 44
распознающий функционал
 допускового множества, 45
распознающий функционал
 объединённого множества,
 46
распознающий функционал
 управляемого множества,
 45

соответствия принцип, 17, 34
твин, 30
твин Нестерова, 31
твина относительная ширина, 31
теорема об аналитической
 характеризации
 АЕ-решений ИСЛАУ, 43
униmodalная выборка, 33, 68, 69
характеризация области значения
 переменной, 49
характеристика вариабельности
 множества совместности,
 73
частота моды, 33