Тема X-1. Обработка и анализ данных с интервальной неопределённостью.

А.Н. Баженов

Санкт-Петербургский политехнический университет Петра Великого

a_bazhenov@inbox.ru

13.09.2022

Интервальный анализ и его методы

Интервал — замкнутый отрезок вещественной оси, а **интервальная неопределенность** – состояние неполного знания об интересующей нас величине, когда известна лишь ее принадлежность **некоторому интервалу**.

Интервальный анализ — отрасль математического знания, исследующая задачи с интервальными неопределенностями и методы их решения.

Поиск **множества**, удовлетворяющего **постановке задачи**.

Понятие интервала

Интервалом $[a, b]$ вещественной оси R называется множество всех чисел, расположенных между заданными числами включая их самих, т.е.

$$[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

При этом a и b называются концами интервала.

«. . . В большинстве случаев некорректно говорить о «решении интервальных уравнений» (систем уравнений, неравенств и т. п.) вообще.
Правильнее вести речь о решении тех или иных **постановок задач**, связанных с интервальными уравнениями (системами уравнений, неравенств и т. п.). В свою очередь, формулировка постановки интервальной задачи подразумевает указание, по крайней мере, **множества решений задачи и способа его оценивания**».
С.П.Шарый. Конечномерный интервальный анализ, 2022

Обработка и анализ данных с интервальной неопределённостью.

План

Общий план

- Общие понятия
- Обработка постоянной величины
- Задача восстановления зависимостей
- Задача классификации данных

Теория:
А.Н. Баженов, С.И. Жилин, С.И. Кумков, С.П. Шарый. Обработка и анализ данных с интервальной неопределённостью. РХД. Серия «Интервальный анализ и его приложения». Ижевск. 2022. с.270.

Общие понятия.

Отношения между интервалами.

Интервалы являются множествами, составленными из вещественных чисел, и неудивительно, что большую роль для них играют теоретико-множественные отношения и операции (объединение, пересечение и др.). Особенно важно отношение включения одного интервала в другой:

$$\boldsymbol{a} \subseteq \boldsymbol{b} \quad \text{равносильно тому, что} \quad \underline{\boldsymbol{a}} \geq \underline{\boldsymbol{b}} \quad \text{и} \quad \overline{\boldsymbol{a}} \leq \overline{\boldsymbol{b}}. \tag{1}$$

Отношение включения является частичным порядком и превращает множество интервалов в частично упорядоченное множество, важную и хорошо изученную математическую структуру.

Отношения между интервалами.

Помимо порядка по включению на множестве интервалов огромную роль играют также другие отношения, которые обобщают хорошо известный порядок «$\leq$» на вещественной оси $\mathbb{R}$.
Фундаментальным фактом является то, что порядок «$\leq$» между вещественными числами может быть обобщен на интервалы многими осмысленными способами (и даже бесконечно большим числом способов). Значительная часть получающихся при этом отношений на $\mathbb{IR}$ не являются полноценными порядками.

Отношения между интервалами.

Помимо порядка по включению на множестве интервалов огромную роль играют также другие отношения, которые обобщают хорошо известный порядок «$\leq$» на вещественной оси $\mathbb{R}$.
Фундаментальным фактом является то, что порядок «$\leq$» между вещественными числами может быть обобщен на интервалы многими осмысленными способами (и даже бесконечно большим числом способов). Значительная часть получающихся при этом отношений на $\mathbb{IR}$ не являются полноценными порядками.

Отношения между интервалами.

Важную роль играет следующее упорядочение

Definition

Для интервалов $\mathbf{a}, \mathbf{b} \in \mathbb{IR}$ условимся считать, что $\mathbf{a}$ *не превосходит* $\mathbf{b}$ и писать «$\mathbf{a} \leq \mathbf{b}$» тогда и только тогда, когда $\underline{\mathbf{a}} \leq \underline{\mathbf{b}}$ и $\overline{\mathbf{a}} \leq \overline{\mathbf{b}}$.
Интервал называется *неотрицательным*, т. е. «≥ 0», если неотрицательны оба его конца. Интервал называется *неположительным*, т. е. «≤ 0», если неположительны оба его конца.

Теоретико-множественные операции между интервалами.

Если интервалы $\boldsymbol{a}$ и $\boldsymbol{b}$ имеют непустое пересечение, т. е. $\boldsymbol{a} \cap \boldsymbol{b} \neq \varnothing$, то можно дать простые выражения для результатов теоретико-множественных операций пересечения и объединения через концы этих интервалов

$$\boldsymbol{a} \cap \boldsymbol{b} = \left[\max\{\underline{\boldsymbol{a}}, \underline{\boldsymbol{b}}\}, \min\{\overline{\boldsymbol{a}}, \overline{\boldsymbol{b}}\}\right], \quad \boldsymbol{a} \cup \boldsymbol{b} = \left[\min\{\underline{\boldsymbol{a}}, \underline{\boldsymbol{b}}\}, \max\{\overline{\boldsymbol{a}}, \overline{\boldsymbol{b}}\}\right]. \quad (2)$$

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

Теоретико-множественные операции между интервалами.

Обобщением операций пересечения и объединения являются операции взятия минимума и максимума относительно включения «$\subseteq$»:

$$\boldsymbol{a} \wedge \boldsymbol{b} = \left[\max\{\underline{\boldsymbol{a}}, \underline{\boldsymbol{b}}\}, \min\{\overline{\boldsymbol{a}}, \overline{\boldsymbol{b}}\}\right], \quad \boldsymbol{a} \vee \boldsymbol{b} = \left[\min\{\underline{\boldsymbol{a}}, \underline{\boldsymbol{b}}\}, \max\{\overline{\boldsymbol{a}}, \overline{\boldsymbol{b}}\}\right]. \quad (3)$$

Они также понадобятся нам при обработке интервальных измерений.

Первая из этих операций, «$\wedge$», не всегда выполнима во множестве обычных интервалов, но это затруднение преодолевается путём расширения множества интервалов специальными элементами — неправильными интервалами.

Измерения

Definition

Измерением (замером, наблюдением) будем называть измеренное значение величины.

По способу получения результата измерения все процессы измерения разделяются на *прямые*, *косвенные* и *совокупные*.

Измерения и их результаты

- Погрешности квантования
- Неопределённость измерения нуля
- Агрегирование результатов многократных наблюдений

Агрегирование результатов многократных наблюдений.

Во многих практических ситуациях измерение интересующей нас величины выполняется для надёжности многократно. Тем не менее, повторные измерения над одними и теми же явлениями не показывают разумное (в пределах точности измерений) совпадение результатов.

Приняв все необходимы меры предосторожности, обеспечив постоянные условия измерения, мы всё равно не получаем разумно согласующихся друг с другом результатов.

Скажем, в промышленности, как бы тщательно ни был отрегулирован измерительный прибор, колебания в его показаниях не могут быть уменьшены ниже некоторого предела.

Агрегирование результатов многократных наблюдений.

В этих условиях результатом серии повторяющихся измерений можно взять интервал от минимального до максимального из полученных результатов, т. е. агрегировать (объединить) результаты отдельных измерений.

Математически, если результаты повторных измерений величины равны $x_1, x_2, \dots, x_n$, то интервальным результатом следует взять

$$\boldsymbol{x} = \Big[\min_{1 \le i \le n} x_i, \max_{1 \le i \le n} x_i \Big].$$

Будем называть этот способ получения интервального результата измерения *агрегированием*.

Используя операции взятия *интервальной оболочки множества* и *максимума по включению* этот результат можно записать следующим равносильным образом:

$$\boldsymbol{x} = \square\{x_1, x_2, \ldots, x_n\}$$

или

$$\boldsymbol{x} = \bigvee_{1 \le i \le n} x_i .$$

Эти представления хороши тем, что могут быть обобщены на более сложные случаи.

Модель погрешности наблюдения.

Интервалы в результатах измерений могут возникать различным способом. Они могут получаться сразу, в виде готовых интервалов, но могут возникать в результате коррекции точечных результатов.

Один из распространённых способов получения интервальных результатов в первичных измерениях — это «обинтерваливание» точечных значений, когда к точечному *базовому значению* $\mathring{x}$ прибавляется *интервал погрешности* $\boldsymbol{\epsilon}$:

$$\boldsymbol{x} = \mathring{x} + \boldsymbol{\epsilon} \tag{4}$$

Модель погрешности наблюдения.

Интервал погрешности, вообще говоря, может быть произвольным, но если он уравновешен, то есть

$$\boldsymbol{\epsilon} = [-\epsilon, \epsilon],$$

то это можно трактовать, как отсутствие систематических погрешностей в прямом измерении.

Твины.

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»).

Твины были введены в научный оборот в начале 80-х годов XX века в работах испанских исследователей.
Развёрнутый анализ дан в диссертации В.М.Нестерова. Твинные арифметики и их применение в методах и алгоритмах двустороннего интервального оценивания. – Санкт-Петербург, 1999.
http://www.nsc.ru/interval/Library/InteDiss/Nesterov-disser-1999.pdf

Твины.

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

$$\boldsymbol{X} = [\boldsymbol{a}, \boldsymbol{b}] = \big[\,[\underline{\boldsymbol{a}}, \overline{\boldsymbol{a}}], [\underline{\boldsymbol{b}}, \overline{\boldsymbol{b}}]\,\big]. \tag{5}$$

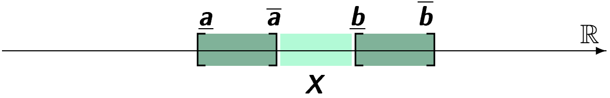

Рис.: Твины на вещественной оси.

На рисунке твин $\boldsymbol{X}$ представлен в графической форме. Концы твина, т. е. интервалы $\boldsymbol{a}$ и $\boldsymbol{b}$, даны более тёмной заливкой, чем остальная часть твина.

Твины.

Твин является множеством всех интервалов, больших или равных $[\underline{a}, \overline{a}]$ и меньших или равных $[\underline{b}, \overline{b}]$, и точное определение зависит от смысла, который вкладывается в понятия «больше или равно», «меньше или равно».

Поскольку интервалы могут быть упорядочены различными способами, то существуют различные виды твинов. Двум основным частичным порядкам на $\mathbb{IR}$ и $\mathbb{KR}$, «$\subseteq$ » и «$\leq$», соответствуют два основных типа твинов. Разработаны различные операции с твинами, а также способы оценок значений функций от них.

Пример

Измерение температуры термометром сопротивления.

В повседневной лабораторной и промышленной практике широко применяются термометры сопротивления.
Один из типов таких датчиков, платиновый термометр Pt100, имеет номинальное сопротивление 100 Ом при температуре $0^\circ C$ и систематическую погрешность

$$\Delta t = \pm 0.35\ ^\circ C.$$

Пример

Пусть измеряемая температура находится в диапазоне [19.5, 20.5] $^\circ C$, которую представим как интервал $\boldsymbol{t}$:

$$\boldsymbol{t} = [19.5, 20.5]\ ^\circ C. \tag{6}$$

Аналогично рассмотренному выше примеру, представим границы $\underline{\boldsymbol{t}}$, $\overline{\boldsymbol{t}}$ интервала $\boldsymbol{t}$ как интервалы. С учётом систематической погрешности твин температур $\boldsymbol{T}$, даваемый датчиком, составит

$$\boldsymbol{T} = \Big[\,[19.15, 19.85],\ [20.15, 20.85]\,\Big]\ ^\circ C. \tag{7}$$

Графическое представление твина $\boldsymbol{T}$ (7) дано на рисунке 2.

Пример

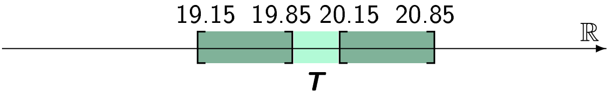

Рис.: Температура как твин.

Форма записи температуры в виде твина $\boldsymbol{T}$ (7) выразительно и полно представляет информацию об измеряемых данных. В случае, если концы интервала в выражении (6) могут меняться независимо, возможны различные ситуации. В частности, может реализоваться ситуация, подобная рассмотренной выше для твина $\boldsymbol{R}_2$. Также может оказаться, что значения температур для левого конца будут выше, чем для правого.

Мультиинтревалы.

В ряде разделов науки и техники имеют место ситуации, когда исследуемая величина содержится в неодносвязной области.

Мультиинтервал — это объединение конечного числа несвязных интервалов числовой оси (Рис. 3).

Рис.: Мультиинтервал в $\mathbb{R}$.

Мультиинтревалы.

Между мультиинтервалами также могут быть определены арифметические операции «по представителям», аналогично тому, как это делается на множестве интервалов.

Мультиинтервальная арифметика применяется редко ввиду серъёзных ограничений, которые возникают при алгебраических операциях с мультиинтервальными величинами и вычислительных сложностей. Тем не менее, сама по себе идея мультиинтервалов содержательна и полностью отметать её не стоит.

Ряд научных и технических примеров возниконовения мультиинтервалов приводится в материале А.Н.Баженов. Естественнонаучные и технические применения интервального анализа: учебное пособие. `https://elib.spbstu.ru/dl/5/tr/2021/tr21-169.pdf/info`.

Пример

Рассмотрим задачу калибровки временной шкалы прибора. Для этого на прибор подаётся гармонический сигнал. В силу того, что на промышленно выпускаемых генераторах положительный и отрицательный фронт имеет разную длительность, необходимо различать эти части временной шкалы.
На рисунке 4 черным цветом показан гармонический сигнал и выделены соответственно красным и синим цветом области положительной и отрицательной производной сигнала. Эти области образуют мультиинтревалы. Они преобразуются при изменении калибровочного сигнала.

При изменении частоты составляющие мультиинтревалов расширяются или сужаются. При изменении фазы происходит их сдвиг.

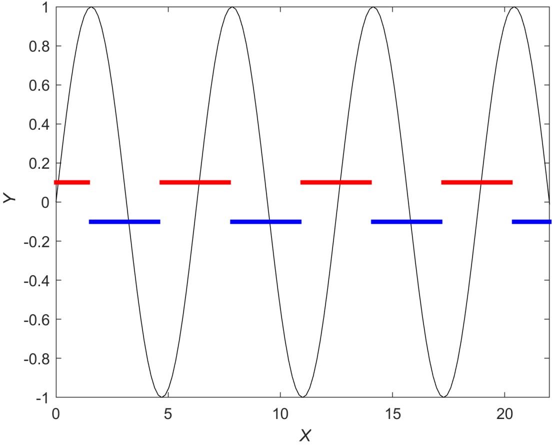

Рис.: Мультиинтревалы фаз гармонического сигнала.

Погрешность измерений

На практике измерения и наблюдения, как правило, подвержены неизбежным внешним влияниям, выполняющие их средства измерений и приборы не вполне точны и т. п., что в целом приводит к отличию измеренного значения от истинного (идеального) значения физической величины.

По отношению к неточным измерениям иногда используют термин «зашумлённые» (зашумлённые данные и т. п.), особенно, когда проводится целая серия таких измерений или наблюдений. Чтобы количественно охарактеризовать неточности измерений, вводится понятие *погрешности*.

Погрешность измерений — вещественная арифметика

Погрешность измерения — это отклонение результата измерения от истинного значения измеряемой величины. Математически погрешность равна алгебраической разности измеренного значения и истинного значения величины.

Если это истинное значение x^* и результат измерения $\tilde{x}$ — вещественные числа, то погрешностью является разность $\tilde{x} - x^*$.

Погрешность измерений — интервальная арифметика

Если истинное значение и результат измерения — интервалы $\boldsymbol{x}^*$ и $\tilde{\boldsymbol{x}}$ соответственно, то погрешность $\boldsymbol{\Delta}$ определяется как *алгебраическая разность*

$$\boldsymbol{\Delta} \;=\; \tilde{\boldsymbol{x}} \ominus \boldsymbol{x}^* \tag{8}$$

в полной интервальной арифметике Каухера.
Напомним, что обычное интервальное вычитание, которое обозначается традиционным знаком «$-$» и является интервальным расширением вычитания, не является операцией, алгебраически обратной сложению и для нашей цели непригодно.

Формула (8) справедлива и в том случае, когда истинное значение величины x^* — точечное, а результат её измерения $\tilde{\boldsymbol{x}}$ интервальный. При этом в (8) полагаем $\boldsymbol{x}^* = [x^*, x^*]$.

Расстояние на множестве интервалов.

Расстояние между интервалами $\boldsymbol{a}$ и $\boldsymbol{b}$ из $\mathbb{IR}$ или $\mathbb{KR}$ определяется как

$$\operatorname{dist}(\boldsymbol{a}, \boldsymbol{b}) \; = \; \max\Big\{|\underline{\boldsymbol{a}} - \underline{\boldsymbol{b}}|, |\overline{\boldsymbol{a}} - \overline{\boldsymbol{b}}|\Big\}. \tag{9}$$

Оно обладает всеми свойствами абстрактного расстояния (метрики) и ещё некоторыми хорошими свойствами в связи с интервальными арифметическими операциями. Кроме того, легко убедиться, что

$$\operatorname{dist}(\boldsymbol{a}, \boldsymbol{b}) \; = \; |\boldsymbol{a} \ominus \boldsymbol{b}|.$$

Эта формула является полным аналогом расстояния между точками вещественной оси, как модуля их разности, т. е. $|a - b|$.

Расстояние на множестве интервалов.

Рассмотрим интервал $[3, 5]$ и точку 3.6 внутри него. Расстояние от этой точки, отождествляемой с вырожденным интервалом $[3.6, 3.6]$, до данного интервала равно

$$\operatorname{dist}(3.6, [3, 5]) = \max\{|3.6 - 3|, |3.6 - 5|\} = 1.4.$$

Рассмотрим дуальный интервал к интервалу $[3, 5]$. Это интервал $\operatorname{dual}[3, 5] = [5, 3]$. Расстояние его до исходного интервала равно

$$\operatorname{dist}([3, 5], [5, 3]) = 2.$$

Расстояние важно для определения отклонения интервалов друг от друга и, как следствие, для определения погрешности интервальных измерений.

Погрешность измерений

Абсолютной погрешностью измерения назовём модуль (абсолютное значение) погрешности.

Для интервальных измерений абсолютная погрешность равна модулю интервала разности $\tilde{x} \ominus \boldsymbol{x}$, и, как легко видеть, она равна расстоянию (9) между измеренным и истинным значениями величины.

Пример

Рассмотрим для примера ситуацию, когда истинное значение измеряемой величины, скажем, массы какого-то груза, является интервалом $[3, 4]$ кг, а её измерение дало интервал $[3, 5]$ кг. Тогда его погрешность равна

$$[3, 5] \text{ кг} \ominus [3, 4] \text{ кг} = [0, 1] \text{ кг}. \tag{10}$$

Пример

Если в результате измерения мы получим вещественное значение 3.8 кг, которое отождествляется с интервалом $[3.8, 3.8]$ кг, то его погрешность

$$[3.8, 3.8] \text{ кг} \ominus [3, 4] \text{ кг} = [0.8, -0.2] \text{ кг} \tag{11}$$

— неправильный интервал.
Может показаться, что он бессмыслен с физической точки зрения, но это поспешный вывод. Ситуация здесь совершенно аналогична, например, тому, как при измерении положительных физических величин (массы, плотности, давления и т. п.) мы получаем отрицательную погрешность, если измеренное значение приближает истинное значение снизу.
Абсолютная погрешность измерения равна 1 в случае (10) и 0.8 в случае (11).

Накрывающие и ненакрывающие измерения

Если результат измерения — точечная величина, то для неё возможны только два исхода проведения измерения: либо она получается равной истинному значению интересующей нас физической величины, либо не равной ей. Как говорят математики и программисты, исход измерения является «булевозначным», «да» или «нет».

При этом ясно, что в случае измерения непрерывных физических величин равенство является исключительным событием и почти никогда не достигается. Если же оно по каким-то причинам произошло, то является неустойчивым к сколь угодно малым возмущениям или же погрешностям в вычислительных алгоритмах.

Накрывающие и ненакрывающие измерения

Принципиально другая ситуация возникает, если результат измерения может быть интервалом.

Интервал по своей сути является двусторонней «вилкой» значений, и принадлежность ей истинного значения — это уже не исключительное событие. Оно, как правило, устойчиво к возмущениям и погрешностям обработки. Как следствие, для теории обработки интервальных данных фундаментальный характер имеют следующие определения:

Definition

Накрывающее измерение (накрывающий замер) — это интервальная оценка неизвестной истинной величины, гарантированно ее содержащая.
Измерение, не являющееся накрывающим, будем называть *ненакрывающим* (Рис. 5 и Рис. 6).

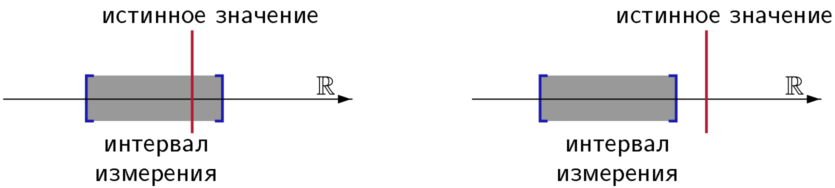

Рис.: Накрывающее (слева) и ненакрывающее (справа)

измерения точечного истинного значения некоторой физической величины.

Definition

Накрывающая выборка — совокупность накрывающих измерений, т. е. выборка, в которой все измерения (наблюдения) являются накрывающими. Напротив, выборка называется *ненакрывающей*, если хотя бы одно из входящих в неё измерений — ненакрывающее.

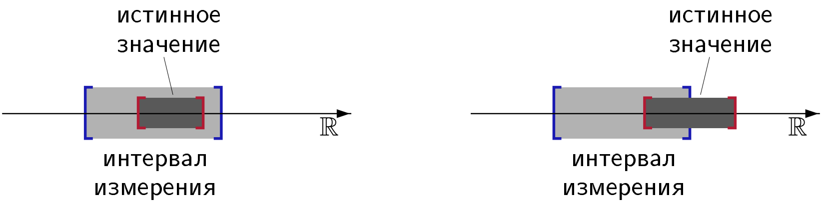

Рис.: Накрывающее (слева) и ненакрывающее (справа)

измерения интервального истинного значения некоторой физической величины.

Информационное множество

Неформально говоря, *информационное множество* — это множество параметров задачи, которые совместны с данными измерений в рамках выбранной модели их обработки.

Информационное множество

Аналогом «информационного множества» может отчасти служить понятие *доверительного интервала* оцениваемой случайной величины в традиционной вероятностной статистике.

В определение доверительного интервала входит дополнительный параметр — *уровень статистической значимости*, без которого понятие становится бессодержательным из-за неограниченности носителей большинства вероятностных распределений, но смысл доверительного интервала примерно соответствует «информационному множеству».

Информационное множество

Далее для обозначения различных информационных множеств мы будем использовать прописную греческую букву

$$\Omega$$

(«омега»), добавляя к ней при необходимости параметры, обозначающие контекст задачи.

Так как информационное множество может быть достаточно произвольным множеством в пространстве параметров и не обязательно является интервалом, интервальным вектором или интервальной матрицей, мы не выделяем его символ жирным шрифтом.

Принцип соответствия

Принцип соответствия в методологии науки — это утверждение, что любая новая научная теория должна включать старую теорию и её результаты как частный предельный случай.

Мы будем использовать принцип соответствия, как инструмент проверки «разумности» и адекватности наших конструкций, понятий и методов обработки данных с интервальными неопределённостями, который позволяет отсекать заведомо «неразумные».

Выбросы и промахи

Выбросами или *промахами* в метрологии называются такие измерения, результаты которых не привносят информацию об исследуемом объекте в рамках его принятой модели.

Другое популярное определение выбросов или промахов состоит в том, что это результаты измерений, которые для данных условий резко отличаются от остальных результатов общей выборки.

Выбросы и промахи

Что считать выбросом (промахом) в случае интервальных результатов измерений? Прежде всего, не стоит связывать выбросы со свойством измерений быть накрывающими или ненакрывающими.
Более точно, из того, что интервальное измерение не является накрывающим, не следует, что оно представляет выброс или промах.

Отождествление выбросов (промахов) со свойством ненакрывания противоречит принципу соответствия, сформулированному в предыдущем параграфе.
В самом деле, при стремлении ширины интервальных измерений к нулю они переходят в точечные измерения, которые, как правило, всегда ненакрывающие. Тем не менее, различение для них выбросов (промахов) от этого не исчезает.

Измерение физической величины (константы).

Физическая величина взята в качестве примера. Данные могут быть любой природы: из наук о Земле, биологии, науках об обществе, экономики, etc.

Измерение физической величины — пример.

Проведём рассмотрение обработки данных физического эксперимента по измерению константы. В качестве источника данных будем использовать публикацию [?], представляющую результаты измерения циркулярной поляризации гамма-кванта в реакции захвата поляризованного нейтрона протоном.

Приведём часть данных таблицы 1 из публикации [?].
В таблице 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.

Исходные данные. Величина $\delta \times 10^5$.

Номер замера	Peak	std Peak	BG	std BG
1	-4.4	2.7	4.2	6.7
2	-3.4	1.9	-3.2	4.8
3	-6.9	2.4	12.1	9
4	-1.2	2.4	12.4	7.2
5	-1.0	2.7	9.4	5.1
6	-10.8	3.5	1	12.4
7	-10.2	2.8	-0.6	6.1
8	-6.3	2	3.9	4.3
9	-10.4	4.1	10.3	10
10	0.6	3.4	-4.8	10.6
11	-1.8	2	4.6	4.2
12	-6.6	2.1	-5.7	4.6
13	-4.9	2.1	13	3
14	-6.0	2.4	8.4	4.6
15	-4.0	2.7	10.6	5.5

Таблица: Данные таблицы 1 для величины $\delta \times 10^5$ [?].

В первую очередь представим данные таким образом, чтобы применить понятия статистики данных с интервальной неопределённостью.
Согласно терминологии интервального анализа, рассматриваемая выборка — это вектор интервалов, или интервальный вектор $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$.
Для того, чтобы придать данным таблицы 1 необходимую форму, примем, что в качестве элементов $\boldsymbol{x}$ будут выступать данные

$$\mathrm{mid}\,\boldsymbol{x}_k = \mathtt{Peak}(k), \quad \mathtt{rad}\boldsymbol{x}_k = \mathtt{std}\ \mathtt{Peak}(k), \quad k = 1, 2, \dots, 15.$$

Для наглядного представления выборки часто рисуют образующие её интервалы в виде графика, изображённого на Рис. 7, который по статистической традиции мы будем называть *диаграммой рассеяния*.

Диаграмма рассеяния интервальных измерений.

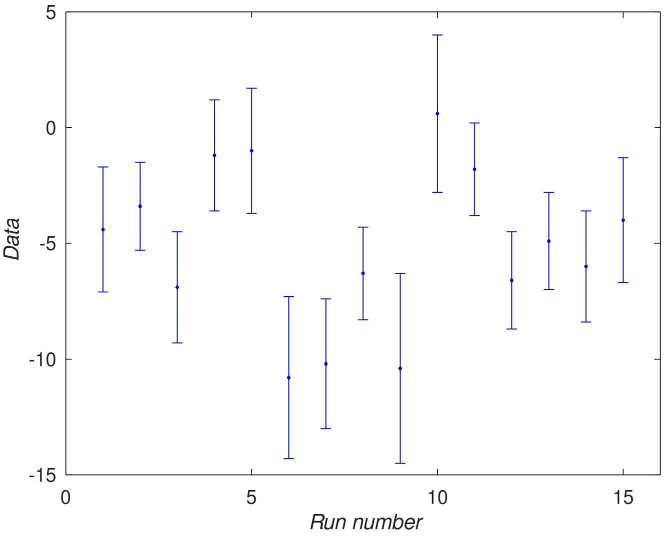

Рис.: Диаграмма рассеяния интервальных измерений [?].

Диаграмма рассеяния интервальных измерений.

Из таблицы 1 и Рис. 7 видно, что элементы выборки *неравноширинные*, поскольку величина неопределённости $\mathtt{rad}\boldsymbol{x}_k$ меняется в зависимости от измерения выборки, $k = 1, \dots, n$.

Информационное множество.

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).

Конкретный смысл, вкладываемый в понятия «совместные» или «согласующиеся», будет различен для разных ситуаций. В частности, он зависит от того, является ли выборка интервальных данных накрывающей или нет.

Совместность выборки

Важным внутренним свойством интервальной выборки, характеризующим согласование её данных между собой, является понятие совместности.

Definition

Выборка $\{\,\mathbf{x}_k\}_{k=1}^{n}$ называется *совместной*, если пересечение всех интервалов составляющих её измерений непусто, т. е.

$$\bigcap_{1\leq k\leq n} \mathbf{x}_k \neq \varnothing.$$

В противном случае, если пересечение всех интервалов $\mathbf{x}_k$, $k = 1, \dots, n$, является пустым, то выборка называется *несовместной*.

Совместность выборки

Свойство совместности характеризует саму выборку и, строго говоря, не связано напрямую с её свойством быть накрывающей выборкой, т. е. с включением ею истинного значения измеряемой величины. Выборка может быть совместной, но ненакрывающей. Но если выборка накрывающая, то она обязана быть совместной. Эквивалентная формулировка этого свойства: если выборка несовместна, то она и ненакрывающая.

Основываясь на этих соображениях, в практической обработке результатов измерений трудный анализ накрытия выборкой истинного значения часто заменяют анализом её совместности, так как это удобнее и нагляднее (хотя и не вполне строго).

Совместность выборки

Если обрабатываемая выборка несовместна, то это может вызываться следующими причинами:

(а) неверно заданным значением неопределённости измерений $\mathrm{rad}\,\boldsymbol{x}_k$ для каких-то $k \in \{1, 2, \ldots, n\}$, которое занижено в сравнении с фактическим значением неопределённости;

(б) наличием в этой выборке выбросов (промахов), т. е. сбойных измерений;

(в) невыполнением условий на измеряемую физическую величину (её непостоянство и т. п.).

Диаграмма совместность—накрытие