

Математические методы в физике.
Курс для аспирантов ФТИ им.
А.Ф.Иоффе.
Практикум.

А.Н. Баженов

22 апреля 2023 г.

Глава 1

Интервальный анализ

1.1 Особенность матрицы

1.2 Решение линейной задачи о допусках

Глава 2

Анализ данных с интервальной неопределённостью

2.1 Оценка постоянной величины для интервальной выборки

2.1.1 Постановка задачи и предварительные оценки

Предметная область относится к физике полупроводников — исследованиям фотоэлектрических характеристик испытываемого датчика, проводимым специалистами лаборатории фотоэлектрических преобразователей Физико-технического института им. А. Ф. Иоффе [23]. Развёрнутое описание задачи дано в статье [5].

Данные выборки. Имеется выборка данных X_1 с интервальной неопределённостью. Число отсчётов в выборке равно 200.

На Рис. 2.1 представлены сырые данные с прибора [23].

Для дальнейшей работы используется модель данных с уравновешенным интервалом погрешности.

$$x = \hat{x} + \epsilon; \quad \epsilon = [-\epsilon, \epsilon] \quad \text{для некоторого } \epsilon > 0, \quad (2.1)$$

Здесь \hat{x} —данные прибора, $\epsilon = 10^{-4}$ — погрешность прибора [23].

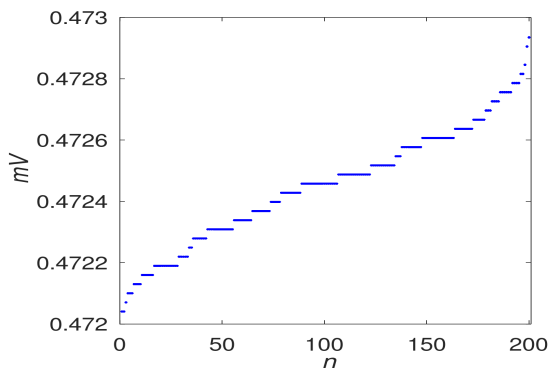


Рис. 2.1: Данные выборки X_1 [23].

Диаграмма рассеяния. Привести диаграмму рассеяния выборки с учётом погрешности прибора.

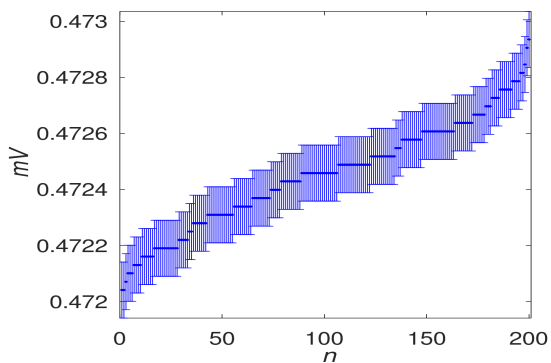


Рис. 2.2: Диаграмма рассеяния выборки X_1 с уравновешенным интервалом погрешности (2.1).

Оценки исходной выборки. Вычислим базовые оценки исходной выборки. Эти выборки несовместны и их информационные множества

пусты. Внешние оценки найдем как

$$\underline{J} = \min_{1 \leq k \leq n} \underline{x}_k, \quad \overline{J} = \max_{1 \leq k \leq n} \overline{x}_k. \quad (2.2)$$

Вычисления дают следующие результаты:

$$J_1 = 0.47194, \quad \text{wid } J_1 = 0.47304. \quad (2.3)$$

Верхние и нижние вершины оценок J_1 совпадают с границами отображения на рис. 2.2.

2.1.2 Меры совместности выборок.

Оценим выборку с помощью набора мер совместности [4].

Сначала значения мер совместности качества возьмём в исходном, ненормированном виде:

1. Размер максимальной клики $\max \mu_j$
2. Величина коэффициента вариабельности по Оскорбину k_O
3. Мера совместности Жаккара J_i

Вычисление моды выборки и максимальной клики. Имеет смысл распространить понятие моды на обработку интервальных данных, где она будет обозначать интервал тех значений, которые наиболее часты, т. е. встречаются в интервалах обрабатываемых данных наиболее часто. Фактически, это означает, что точки из моды интервальной выборки накрываются наибольшим числом интервалов этой выборки. Ясно, что по самому своему определению понятие моды имеет наибольшее значение (и наибольший смысл) лишь для накрывающих выборок. Иначе, если выборка ненакрывающая, то смысл «частоты» тех или иных значений в пределах рассматриваемых интервалов этой выборки в значительной мере теряется, хотя и не обесценивается.

Для вычисления моды интервальной выборки можно применить графовые алгоритмы, которые рассматривались в предшествующем §?? и предназначены для нахождения максимальной совместной подвыборки. Мода является пересечением интервалов такой максимальной подвыборки, и если максимальных подвыборок имеется более одной, то мода будет объединением их пересечений, т. е. мультиинтервалом. Но

можно предложить и более простой алгоритм вычисления моды интервальной выборки, который не опирается на нахождение максимальных совместных подвыборок.

Используем алгоритм для нахождения моды интервальной выборки [6]. Псевдокод специализированного алгоритма для нахождения моды выборки интервальных измерений и её частоты приведён в Табл. 2.1.

Ключевым в алгоритме Табл. 2.1 является формирование множества *элементарных подинтервалов измерений* из упорядоченных вершин (концов интервалов) $\underline{x}_1, \bar{x}_1, \underline{x}_2, \bar{x}_2, \dots, \underline{x}_n, \bar{x}_n$ исходной выборки \mathbf{X} .

Отметим также, что мода интервальной выборки — это интервал или мультиинтервал, который не обязан совпадать с каким-либо из интервалов обрабатываемой выборки.

Пример 2.1.1. Рассмотрим пример вычисления моды интервальной выборки из 4 элементов

$$\mathbf{X} = \{ [1, 4], [5, 9], [1.5, 4.5], [6, 9] \}. \quad (2.4)$$

В соответствии с алгоритмом Табл. 2.1, проверим совместность \mathbf{X} . Пересечение элементов выборки пусто

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

Таким образом, необходимо выполнить шаги алгоритма Табл. 2.1 после ключевого слова ELSE.

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

$$\mathbf{z} = \{ [1, 1.5], [1.5, 4], [4, 4.5], [4.5, 5], [5, 6], [6, 9], [9, 9] \}. \quad (2.5)$$

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

$$\{1, 2, 1, 0, 1, 2, 2\}. \quad (2.6)$$

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

$$\text{mode } \mathbf{X} = \bigcup_{k \in K} \mathbf{z}_k = [1.5, 4] \cup [6, 9] \cup [9, 9] = [1.5, 4] \cup [6, 9]. \quad (2.7)$$

■

Проведём вычисление моды выборки \mathbf{X}_1 по алгоритму Табл. 2.1.

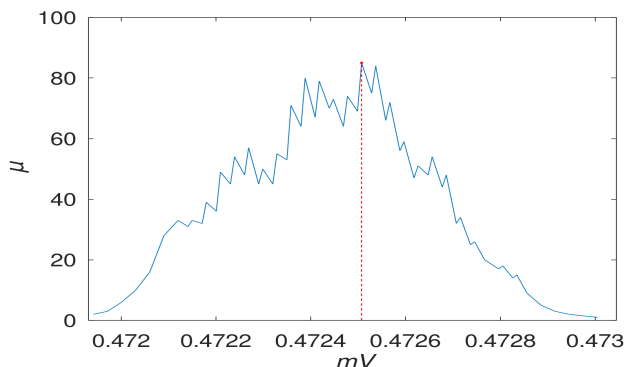


Рис. 2.3: График частот при вычислении моды выборки \mathbf{X}_1

По результатам вычисления моды выборки находим размер максимальной клики $\max \mu_j$:

$$\max \mu_j(\mathbf{X}_1) = 85 \quad (2.8)$$

$$\text{mode}(\mathbf{X}_1) = [0.4725, 0.47251]. \quad (2.9)$$

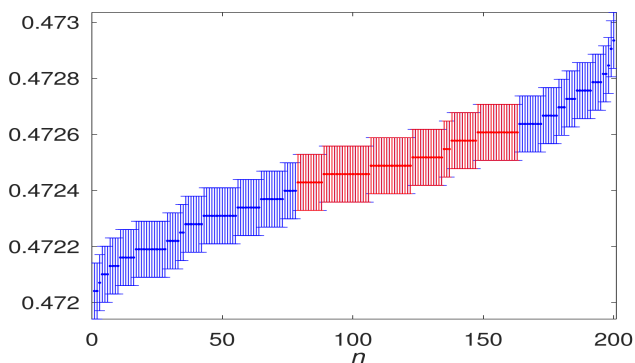


Рис. 2.4: Элементы выборки \mathbf{X}_1 , в которые входит мода.

На рис. 2.4 показаны элементы выборки \mathbf{X}_1 , в которые входит мода.

Таблица 2.1: Алгоритм для нахождения моды
интервальной выборки

Вход

Интервальная выборка $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ длины n .

Выход

Мода $\text{mode } \mathbf{X}$ выборки \mathbf{X} и её частота μ .

Алгоритм

$\mathbf{I} \leftarrow \bigcap_{i=1}^n \mathbf{x}_i$;

IF $\mathbf{I} \neq \emptyset$ THEN

$\text{mode } \mathbf{X} \leftarrow \mathbf{I}$;

$\mu \leftarrow n$

ELSE

 помещаем все концы $\underline{\mathbf{x}}_1, \overline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \overline{\mathbf{x}}_2, \dots, \underline{\mathbf{x}}_n, \overline{\mathbf{x}}_n$

 интервалов рассматриваемой выборки \mathbf{X} в один

 массив $Y = (y_1, y_2, \dots, y_{2n})$;

 упорядочиваем элементы в Y по возрастанию значений;

 порождаем интервалы $\mathbf{z}_i = [y_i, y_{i+1}]$, $i = 1, 2, \dots, 2n - 1$

 (назовём их *элементарными подинтервалами измерений*);

 для каждого \mathbf{z}_i подсчитываем число μ_i интервалов

 из выборки \mathbf{X} , включающих интервал \mathbf{z}_i ;

 вычисляем $\mu \leftarrow \max_{1 \leq i \leq 2n-1} \mu_i$;

 выбираем номера k интервалов \mathbf{z}_k , для которых μ_k

 равно максимальному, т. е. $\mu_k = \mu$, и формируем

 из таких k множество $K = \{k\} \subseteq \{1, 2, \dots, 2n - 1\}$;

$\text{mode } \mathbf{X} \leftarrow \bigcup_{k \in K} \mathbf{z}_k$

END IF

Варьирование неопределённости измерений. Один из приёмов выявления достижения совместности выборки интервальных наблюдений основан на представлении о причине несовместности как недооценённой величины неопределённости [24, 25]. Закономерным шагом в этом случае становится поиск некоторой минимальной коррекции величин неопределённости интервальных наблюдений, необходимой для обеспечения совместности задачи построения зависимости. Если величину коррекции каждого интервального наблюдения $\mathbf{y}_i = [\hat{y}_i - \epsilon_i, \hat{y}_i + \epsilon_i] = \langle \hat{y}_i, \epsilon_i \rangle$ выборки \mathcal{S}_n выражать коэффициентом его уширения $w_i \geq 1$, а общее изменение выборки характеризовать суммой этих коэффициентов, то минимальная коррекция выборки в виде вектора коэффициентов $w^* = (w_1^*, \dots, w_n^*)$, необходимая для совместности задачи построения зависимости $y = f(x, \beta)$ может быть найдена решением задачи условной оптимизации

$$\text{найти } \min_{w, \beta} \sum_{i=1}^n w_i \quad (2.10)$$

при ограничениях

$$\begin{cases} \hat{y}_i - w_i \epsilon_i \leq f(x_i, \beta) \leq \hat{y}_i + w_i \epsilon_i, \\ w_i \geq 1, \end{cases} \quad i = 1, \dots, n. \quad (2.11)$$

Результирующие значения коэффициентов w_i^* , строго превосходящие единицу, указывают на наблюдения, которые требуют уширения интервалов неопределённости для обеспечения совместности данных и модели. Именно такие наблюдения заслуживают внимания при анализе данных на выбросы. Значительное количество подобных наблюдений может говорить либо о неверно выбранной структуре зависимости, либо о том, что величины неопределённости измерений занижены во многих наблюдениях (например, в результате неверной оценки точности измерительного прибора).

Следует отметить значительную гибкость языка неравенств. Он даёт возможность переформулировать и расширять систему ограничений (2.11) для учёта специфики данных и задачи при поиске допустимой коррекции данных, приводящей к разрешению исходной несовместности. Например, если имеются основания считать, что величина неопределённости некоторой группы наблюдений одинакова и при коррекции должна увеличиваться синхронно, то система ограничений (2.11) может быть пополнена равенствами вида

$$w_{i_1} = w_{i_2} = \dots = w_{i_K},$$

где i_1, \dots, i_K — номера наблюдений группы. В случае, когда в надёжности каких-либо наблюдений исследователь уверен полностью, при решении задачи (2.10)–(2.11) соответствующие им величины w_i можно положить равными единице, т.е. запретить варьировать их неопределённость.

Задача поиска коэффициентов масштабирования величины неопределённости (2.10)–(2.11) сформулирована для распространённого случая уравновешенных интервалов погрешности и подразумевает синхронную подвижность верхней и нижней границ интервалов неопределённости измерений \mathbf{y}_i при сохранении базовых значений интервалов \hat{y}_i неподвижными. При необходимости постановка задачи легко обобщается. Например, если интервалы наблюдений не уравновешены относительно базовых значений (то есть $\mathbf{y}_i = [\hat{y}_i - \epsilon_i^-, \hat{y}_i + \epsilon_i^+]$ и $\epsilon^- \neq \epsilon^+$), то границы интервальных измерений можно варьировать независимо, масштабируя величины неопределённости ϵ_i^- и ϵ_i^+ с помощью отдельных коэффициентов w_i^- и w_i^+ :

$$\text{найти} \quad \min_{w^-, w^+, \beta} \quad \sum_{i=1}^n (w_i^- + w_i^+) \quad (2.12)$$

при ограничениях

$$\left\{ \begin{array}{l} \hat{y}_i - w_i^- \epsilon_i^- \leq f(x_i, \beta) \leq \hat{y}_i + w_i^+ \epsilon_i^+, \\ w_i^- \geq 1, \\ w_i^+ \geq 1, \end{array} \quad i = 1, \dots, n. \right. \quad (2.13)$$

Для линейной по параметрам β зависимости $y = f(x, \beta)$ задача (2.10)–(2.11) представляет собою задачу линейного программирования, для решения которой широко доступны хорошие и апробированные программы в составе библиотек на различных языках программирования, в виде стандартных процедур систем компьютерной математики, а также в виде интерактивных подсистем электронных таблиц.

Индекс Жаккара Для описания выборок, помимо оценок их размеров, желательно иметь дополнительную информацию о мере сходства элементов выборки. В различных областях анализа данных, биологии, информатике, в науках о Земле часто используют различные меры сходства множеств (см. [22]). Мету сходства между объектами A и B можно определить как двухместную вещественнозначную функцию $S(A, B)$, которая обладает следующими свойствами:

- ограниченность: $0 \leq S(A, B) \leq 1$;
- симметричность: $S(A, B) = S(B, A)$;
- неразличимость совпадающих элементов: $S(A, B) = 1 \Leftrightarrow A = B$;
- монотонность: $A \subseteq B \subseteq C \Rightarrow S(A, B) \geq S(A, C)$.

Таким образом, значение 1 этой меры соответствует совпадению множеств A и B , а значение 0 означает их полное несходство. Отметим, что существуют и иные системы аксиом сходства. В компьютерных приложениях (обработка изображений, машинное обучение) меру сходства множеств обозначают как *IoU* (*Intersection over Union*). В математике и её приложениях для подобных конструкций часто используется термин *индекс Жаккара*, по имени исследователя, впервые предложившего эту меру.

В процессе развития интервального анализа были введены различные определения и конструкции оценки меры совместности интервальных объектов. Вместе с тем в практике обработки данных часто необходимо оперировать относительными величинами. В частности, это нужно в связи с необходимостью сопоставления допусков и размеров деталей, погрешности измерителей и значений измеряемых величин и т. п. [?].

Представим обобщение меры Жаккара на выборки интервалов [5]. В качестве числовой характеристики степени совпадения двух интервалов \mathbf{x} , \mathbf{y} рассмотрим величину

$$Ji(\mathbf{x}, \mathbf{y}) := \frac{\text{wid}(\mathbf{x} \wedge \mathbf{y})}{\text{wid}(\mathbf{x} \vee \mathbf{y})}. \quad (2.14)$$

В выражении (2.14) используется ширина интервала (см. стр. ??), а вместо операций пересечения и объединения множеств — операции взятия точной нижней грани “ \wedge ” (инфимума, см. (??)) и точной верхней грани “ \vee ” (супремума, см. (??)) относительно включения для двух величин в полной интервальной арифметике Каухера. В обозначении $Ji(\mathbf{x}, \mathbf{y})$ буква J указывает на фамилию «Jaccard», а i — на интервальность его применения.¹ В общем случае инфимум по включению в числителе выражения (2.14) может быть неправильным интервалом, и его ширина тогда отрицательна.

¹Можно встретить также другие обозначения для этой конструкции, в частности, «JK», как это сделано в работе [5].

Рассмотренная мера обобщает обычное понятие меры совместности на различные типы взаимной совместности интервалов. Если пересечение интервалов \mathbf{x}, \mathbf{y} пусто, т. е. $\mathbf{x} \cap \mathbf{y} = \emptyset$, то $\mathbf{x} \wedge \mathbf{y}$ — неправильный интервал и числитель формулы (2.14) имеет отрицательное значение. В предельном случае несовпадающих вещественных вырожденных интервалов $\mathbf{x} = x$ и $\mathbf{y} = y$, $x \neq y$, имеем

$$Ji(x, y) = -1.$$

В целом получаем

$$-1 \leq Ji(\mathbf{x}, \mathbf{y}) \leq 1. \quad (2.15)$$

Таким образом, величина Ji непрерывно описывает ситуации от полной несовместности вещественных значений $x \neq y$ до полного перекрытия интервалов $\mathbf{x} = \mathbf{y}$. Следует заметить, что в отличие от случая вещественных величин, для которых индекс Жаккара может принимать только два значения, 0 и 1, формула (2.14) даёт характеристику различных отношений сходства интервалов с помощью непрерывного ряда значений между -1 и 1 .

Мера совместности, введённая для двух интервалов в форме (2.14), допускает естественное обобщение на случай интервальной выборки $\mathbf{X} = \{\mathbf{x}_i\}$, $i = 1, 2, \dots, n$. Определим меру $Ji(\mathbf{X})$ для этой выборки как

$$Ji(\mathbf{X}) = \frac{\text{wid}(\bigwedge_i \mathbf{x}_i)}{\text{wid}(\bigvee_i \mathbf{x}_i)}. \quad (2.16)$$

Видно, что выражение (2.16) переходит в случае интервальной выборки из двух элементов в выражение (2.14).

В связи несовместностью выборки будем использовать следующую меру, которая имеет место и в случае несовместных выборок.

$$\rho(\text{mode}(\mathbf{X})) = \frac{\text{wid}(\text{mode}(\mathbf{X}))}{\text{wid}(\bigvee_i \mathbf{x}_i)} = 0.0655. \quad (2.17)$$

Назовём конструкцию (2.17) *относительная ширина моды*. В отличие от минимума по включению, мода выборки всегда является правильным интервалом. В целом получаем

$$0 \leq \rho(\text{mode}(\mathbf{X})) \leq 1. \quad (2.18)$$

Литература

- [1] А.Н. Баженов, С.И. Жилин, С.И. Кумков, С.П. Шарый. «Обработка и анализ данных с интервальной неопределённостью». 2023.
- [2] Баженов А.Н. Интервальный анализ. Основы теории и учебные примеры: учебное пособие. — СПб. 2020 – 78 с. <https://elib.spbstu.ru/dl/2/s20-76.pdf/info>
- [3] Баженов А.Н.. Обобщение мер совместности для анализа данных с интервальной неопределённостью. – СПб., 2022. – 80 с. <https://elib.spbstu.ru/dl/5/tr/2022/tr22-142.pdf/info>
- [4] Баженов А.Н.. Примеры обработки измерений постоянной величины в анализе данных с интервальной неопределённостью. – СПб., 2023. <https://elib.spbstu.ru/dl/5/tr/2023/tr23-29.pdf/info>
- [5] Баженов А.Н., Тельнова А.Ю. Обобщение коэффициента Жаккара для анализа данных с интервальной неопределённостью // Измерительная техника. – 2022. № 12, С. 12-15.
- [6] Баженов А.Н. Введение в анализ данных с интервальной неопределённостью : учеб. пособие / А.Н. Баженов - СПб. : ПОЛИТЕХ-ПРЕСС, 2022. - 92 с. Табл.3. Ил.36. Библиогр.: 48 назв. ISBN 978-5-7422-7910-5 Санкт-Петербургский политехнический университет Петра Великого, 2022 doi:10.18720/SPBPU/2/id22-247
- [7] Оскорбин Н.М., Максимов А.В., Жилин С.И. Построение и анализ эмпирических зависимостей методом центра неопределённости // Известия Алтайского государственного университета. – 1998. – №1. – С. 35–38.
- [8] Примеры анализа интервальных данных в Octave <https://github.com/szhilin/octave-interval-examples>
- [9] Шарый С.П. Конечномерный интервальный анализ. – ФИЦ ИВТ: Новосибирск, 2022. Электронная книга, доступная на <http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf>

- [10] ШРЕЙДЕР Ю.А. Равенство, сходство, порядок. – М.: Наука, 1971. 256 с.
- [11] KEARFOTT, R.B., NAKAO, M., NEUMAIER, A., RUMP, S., SHARY, S.P., VAN HENTENRYCK, P. Standardized notation in interval analysis // Вычислительные Технологии. – 2010. – Т. 15, №1. – С. 7–13.
- [12] 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
- [13] E. GARDENES, A.TREPAT, J.M. JANER Approaches to simulation and to the linear problem in the SIGLA system. // Gardenes, E., Trepata, A., and Janer, J. M.: Approaches to Simulation and to the Linear Problem in the SIGLA System, Freiburger Interval-Berichte 81 (8) (1981), pp. 1-28.
- [14] NESTEROV, V.M. Interval and Twin Arithmetics. Reliable Computing 3, 369–380 (1997). <https://doi.org/10.1023/A:1009945403631>
- [15] В.М. НЕСТЕРОВ Твинные арифметики и их применение в методах и алгоритмах двустороннего интервального оценивания. дисс. д.ф.-м.н. г.Санкт-Петербург, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 1999, с. 234.
- [16] ОСКОРБИН Н. М. Построение и анализ эмпирических зависимостей методом центра неопределенности / Н.М. Оскорбин, А.В. Максимов, С.И. Жилин. — Барнаул : Изв. Алтайского гос. ун-та, 1998. — № 1. — С. 35-38.
- [17] С.И.Жилин Интервальная арифметика Каухера <https://github.com/szhilin/kinterval>
- [18] А.Андросов Библиотека `intvalpy` <https://github.com/AndrosovAS/intvalpy>,
- [19] А. ЖАВОРОНКОВА Арифметика твинов Нестерова. <https://github.com/Zhavoronkova-Alina/twin>
- [20] Т. ЯВОРУК Вычисления с изотопами `MendeleevTwin` — <https://github.com/Tatiana655/MendeleevTwin>
- [21] Tables and charts for isotope-abundance variations and atomic weights of selected elements: 2016 (Ver. 1.1, May 2018) <https://www.sciencebase.gov/catalog/item/580e719ae4b0f497e794b7d8>
- [22] Коэффициент Жаккара. URL: https://en.wikipedia.org/wiki/Jaccard_index
- [23] М.З. ШВАРЦ <https://github.com/AlexanderBazhenov/Solar-Data>

- [24] Жилин С.И. Нестатистические методы и модели построения и анализа зависимостей. – Барнаул, 2004. – Диссертация на соискание учёной степени канд. физ.-мат. наук по специальности 05.13.01 «системный анализ, управление и обработка информации». Доступна на <http://www.nsc.ru/interval/Library/AppIDiss/Zhilin.pdf>
- [25] 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.
- [26] J.-R. Yu, G.-H. Tzeng, Han-Lin Li. General piecewise necessity regression analysis based on linear programming Fuzzy Sets and Systems Volume 105, Issue 3, 1 August 1999, Pages 429-436.
- [27] J.-R. Yu, G.-H. Tzeng, Han-Lin Li. General fuzzy piecewise regression analysis with automatic change-point detection. Fuzzy Sets and Systems Volume 119 Issue 2 April 16, 2001 pp 247-257 [https://doi.org/10.1016/S0165-0114\(98\)00384-4](https://doi.org/10.1016/S0165-0114(98)00384-4)
- [28] Circular polarization of γ -quanta in the $np \rightarrow d\gamma$ reactions with polarized neutrons / A.N.Bazhenov [et al.] // Physics Letters B. 3. — September 1992. Vol. 289. — No. 1-2. — P. 17-21.