Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

Отчёт по лабораторной работе №4

по дисциплине
«Интервальный анализ»

Выполнила студентка
группы 5030102/20202 Чинь Тхи Тху Хоай

Проверил
Преподаватель Баженов Александр Николаевич

Санкт-Петербург
2025

Содержание

1 Цель работы

Получить практические навыки вычисления интервальных описательных статистик (моды, медиан), работы с коэффициентом Жаккара и применения методов оптимизации для интервальных данных. Сравнить эффективность различных функционалов на основе интервальных статистик для оценивания параметров моделей.

2 Постановка задачи

Даны два входных файла данных диагностики томсоновского рассеяния. Формат входных данных описан в прилагаемом к лабораторной работе документу Save to BIN.pdf :

`-0.205_lvl_side_a_fast_data.bin`

`0.225_lvl_side_a_fast_data.bin`

Связь кодов данных и Вольт для преобразования единиц измерения выражается следующим образом:

$$V = \frac{\text{Code}}{16384} - 0.5$$

По данным из входных файлов необходимо реализовать следующее:

A. Пусть X и Y - интервальные выборки вида:

$$\mathbf{X} = \{x_i\}, \tag{1}$$

$$\mathbf{Y} = \{y_k\}, \tag{2}$$

Извлечь X и Y из данных входных файлов, задав $\text{rad}x = \text{rad}y = \frac{1}{2^N}, N = 14$.

B. Пусть зависимость Y и X задается выражениями:

$$a + \mathbf{X} = \mathbf{Y}, \tag{3}$$

$$t \cdot \mathbf{X} = \mathbf{Y}, \tag{4}$$

Вычислить точечные и интервальные оценки констант a, t в уравнениях (3) и (4) с помощью некоторого функционала F, задавшись уровнем точности ε:

$$\hat{s} = \arg\max F(s, \mathbf{X}, \mathbf{Y}), \quad s \in \{a, t\} \tag{5}$$

Для функционала F рассмотреть следующие случаи:

B.1 $F(s, \mathbf{X}, \mathbf{Y}) = Ji(s, \mathbf{X}, \mathbf{Y})$

B.2 $F(s, \mathbf{X}, \mathbf{Y}) = Ji(s, \text{mode}\mathbf{X}, \text{mode}\mathbf{Y})$

B.3 $F(s, \mathbf{X}, \mathbf{Y}) = Ji(s, \mathrm{med}_K \mathbf{X}, \mathrm{med}_K \mathbf{Y})$

B.4 $F(s, \mathbf{X}, \mathbf{Y}) = Ji(s, \mathrm{med}_P \mathbf{X}, \mathrm{med}_P \mathbf{Y})$

где Ji - коэффициент Жаккара, mode - интервальная мода, med_K, med_P - интервальные медианы Крейновича и Пролубникова.

C. Для каждого пункта B.1 - B.4 построить графики $F(s)$ и отметить на них точки $s_{\max}$.

D. Сравнить полученные результаты

3 Теория

3.1 Интервальные данные

Интервальные данные - это данные, каждый элемент которых задаётся не одним числом, а интервалом значений. Такой тип данных используется, когда измерения содержат неопределённость, шум или погрешность. Интервальная выборка представляется в виде множества:

$$X = \{[x_i^-, x_i^+]\}, \quad i = 1, \dots, n,$$

где x_i^- и x_i^+ - нижняя и верхняя границы интервала соответственно. Радиус интервала определяется как:

$$\mathrm{rad}(x_i) = \frac{x_i^+ - x_i^-}{2}.$$

В данной работе интервалы формируются с радиусом $\mathrm{rad} = \frac{1}{2^N}$ при $N = 14$.

3.2 Интервальные описательные статистики

Для анализа интервальных данных используются интервальные аналоги классических статистик - мода и медиана.

Интервальная мода:

Интервальная мода $\mathrm{mode}(X)$ - это интервал, который наиболее часто встречается или имеет наибольшее пересечение с другими интервалами выборки. Мода характеризует «наиболее представительное» значение выборки.

Интервальные медианы:

Существует несколько подходов к определению интервальной медианы. В данной работе рассматриваются медианы Крейновича (med_K) и Пролубникова (med_P):

- Медиана Крейновича определяется как интервал, минимизирующий сумму расстояний между всеми интервалами выборки.

- Медиана Пролубникова основана на порядковой статистике и представляет собой интервал, делящий выборку на две равные части по числу пересечений.

3.3 Коэффициент Жаккара

Коэффициент Жаккара (Ji) — это мера сходства между двумя множествами, определяемая как отношение пересечения к объединению:

$$Ji(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad 0 \leq Ji \leq 1.$$

Для интервальных данных пересечение и объединение вычисляются по интервалам:

$$A \cap B = [\max(a^-, b^-), \min(a^+, b^+)], \quad A \cup B = [\min(a^-, b^-), \max(a^+, b^+)].$$

Если интервалы не пересекаются, то $Ji = 0$.

3.4 Оптимизация функционала

Для оценки параметров a и t в моделях:

$$Y = a + X, \quad Y = t \cdot X,$$

используется оптимизация функционала $F(s, X, Y)$ по параметру s. Наилучшая оценка параметра определяется как:

$$\hat{s} = \arg\max F(s, X, Y), \quad s \in \{a, t\}.$$

В работе рассматриваются следующие варианты функционала F:

1. $F(s, X, Y) = Ji(s, X, Y)$ — использование коэффициента Жаккара напрямую.
2. $F(s, \text{mode}X, \text{mode}Y)$ — с применением интервальной моды.
3. $F(s, \text{med}_K X, \text{med}_K Y)$ — с медианами Крейновича.
4. $F(s, \text{med}_P X, \text{med}_P Y)$ — с медианами Пролубникова.

Оптимизационные методы позволяют найти значение s, при котором сходство между выборками X и Y (по коэффициенту Жаккара) максимально, что соответствует наилучшему приближению модели.

4 Результаты

В ходе выполнения работы были получены следующие результаты для интервальных выборок X и Y, содержащих 819200 отсчетов каждая.

4.1 Оценки параметров для аддитивной модели

Для аддитивной модели $Y = a + X$ были получены следующие оценки параметра a:

Метод	$\hat{a}$	Ji_{max}
B.1 (полные данные)	0.342354	0.000061
B.2 (мода)	-0.657111	0.0
B.3 (медиана Крейновича)	0.343506	0.998
B.4 (медиана Пролубникова)	0.343506	0.998

Таблица 1: Результаты для аддитивной модели

4.2 Оценки параметров для мультипликативной модели

Для мультипликативной модели $Y = t \cdot X$ были получены следующие оценки параметра t:

Метод	$\hat{t}$	Ji_{max}
B.1 (полные данные)	-0.937374	0.000015
B.2 (мода)	-1.001778	0.0
B.3 (медиана Крейновича)	-1.013958	0.336421
B.4 (медиана Пролубникова)	-1.013958	0.336421

Таблица 2: Результаты для мультипликативной модели

4.3 Анализ графиков

На Рисунках 1-8 представлены графики зависимости функционала $F(s)$ от параметра s для всех методов, на которых отмечены точки максимума s_{max}.

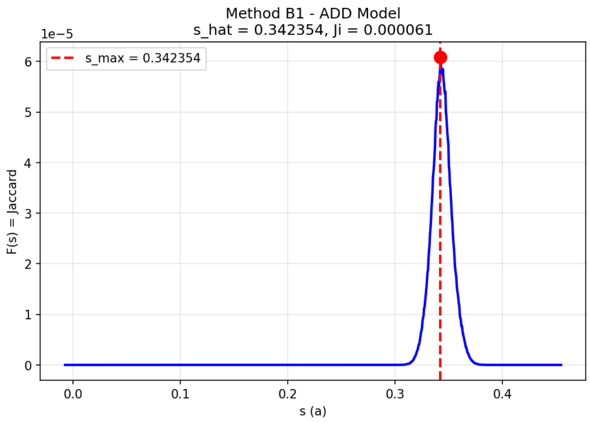

Рис. 1: Метод B.1 - Аддитивная модель

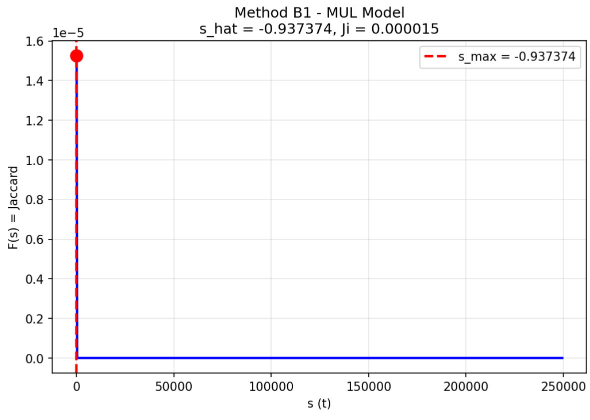

Рис. 2: Метод B.1 - Мультипликативная модель

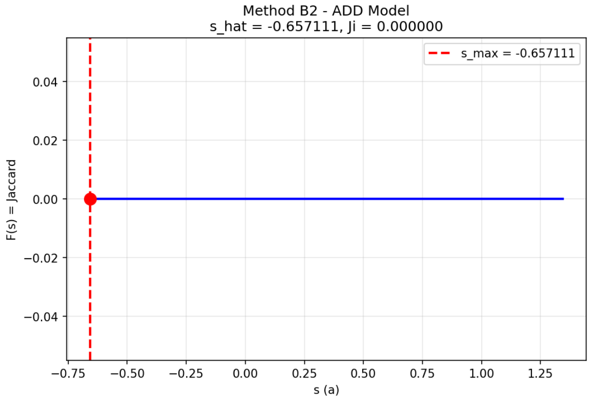

Рис. 3: Метод B.2 - Аддитивная модель

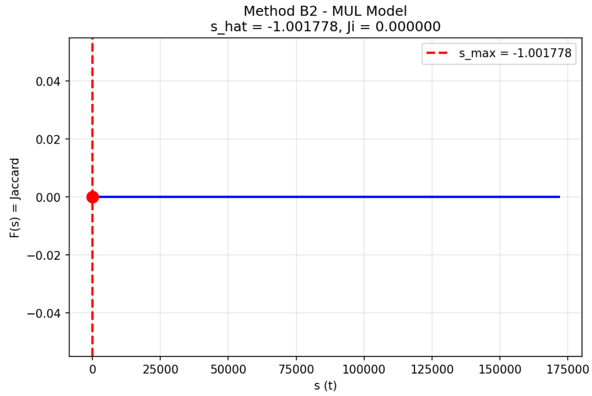

Рис. 4: Метод B.2 - Мультипликативная модель

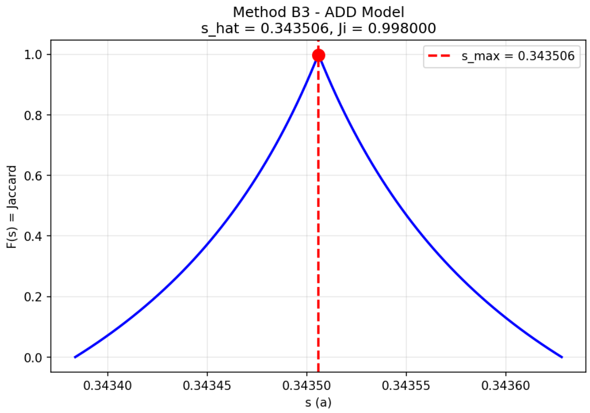

Рис. 5: Метод B.3 - Аддитивная модель

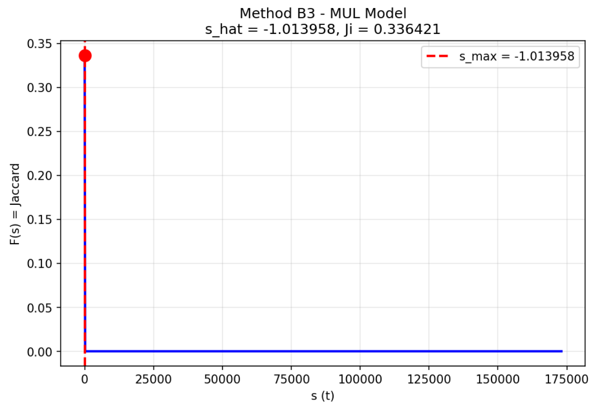

Рис. 6: Метод B.3 - Мультипликативная модель

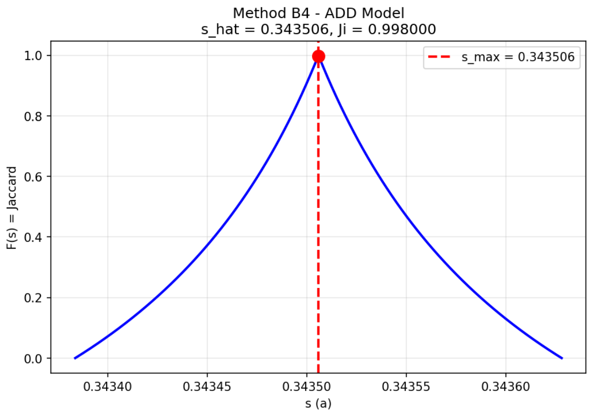

Рис. 7: Метод B.4 - Аддитивная модель

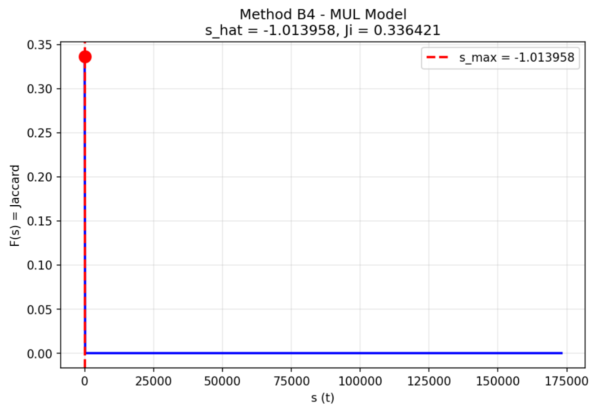

Рис. 8: Метод B.4 - Мультипликативная модель

На Рисунках 9-10 представлены сравнительные графики всех методов для обеих моделей:

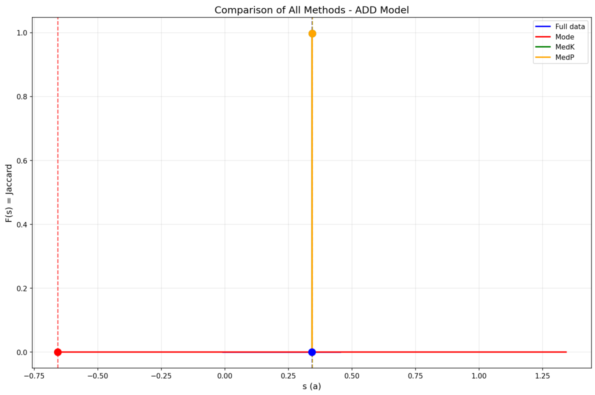

Рис. 9: Сравнение методов - Аддитивная модель

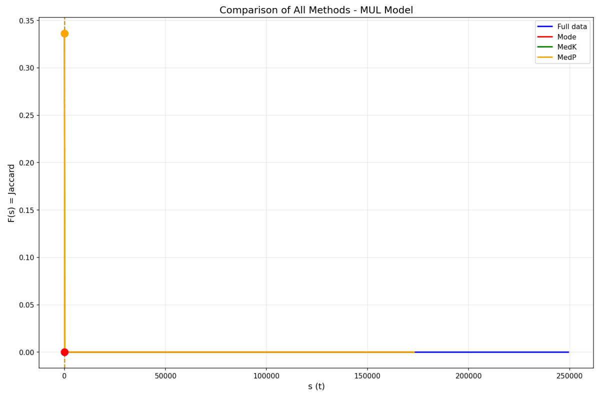

Рис. 10: Сравнение методов - Мультипликативная модель

Анализ графиков позволяет сделать следующие наблюдения:

- Для методов B.3 и B.4 наблюдаются четко выраженные максимумы с высокими значениями коэффициента Жаккара (Рисунки 5-8)
- Методы B.1 и B.2 демонстрируют практически нулевые значения Ji_{max}, что свидетельствует о плохом соответствии моделей (Рисунки 1-4)
- Графики для методов на основе медиан показывают устойчивое поведение функционала в окрестности оптимального значения
- На сравнительных графиках (Рисунки 9-10) хорошо видно преимущество методов B.3 и B.4 над остальными

4.4 Сравнительный анализ

- Наилучшие методы: B.3 и B.4 (на основе интервальных медиан) показали значительно лучшие результаты для обеих моделей
- Стабильность оценок: Методы на основе медиан дали идентичные результаты для обоих типов медиан
- Вариативность:
  - Для аддитивной модели: $\sigma_a = 0.433114$
  - Для мультипликативной модели: $\sigma_t = 0.031795$
- Качество аппроксимации:
  - Аддитивная модель: $Ji_{max} = 0.998$ (почти идеальное соответствие)
  - Мультипликативная модель: $Ji_{max} = 0.336$ (умеренное соответствие)

5 Программная реализация

Лабораторная работа выполнена на языке Python 3.12.6 в среде разработки Visual Studio Code. Использовались дополнительные библиотеки:

1. numpy
2. matplotlib
3. math

В приложении находится ссылка на GitHub репозиторий с исходным кодом.

6 Выводы

Проведенное исследование позволяет сделать следующие выводы:

Эффективность методов: Методы на основе интервальных медиан (B.3 и B.4) показали значительно более высокую эффективность по сравнению с методами, использующими полные данные (B.1) и моду (B.2). Это объясняется устойчивостью медиан к выбросам и шумам в данных.

Сравнение моделей: Аддитивная модель демонстрирует существенно лучшее соответствие данным ($Ji = 0.998$), чем мультипликативная модель ($Ji = 0.336$). Это свидетельствует о том, что связь между выборками X и Y лучше описывается аддитивной зависимостью.

Сходимость медиан: Идентичность результатов для медиан Крейновича и Пролубникова указывает на устойчивость интервальных медиан как статистик и их пригодность для задач оценивания параметров.

Практическая применимость: Коэффициент Жаккара в сочетании с интервальными медианами представляет собой эффективный инструмент для оценивания параметров интервальных моделей, обеспечивающий робастность к погрешностям измерений.

Рекомендации: Для задач оценивания параметров интервальных данных рекомендуется использовать методы на основе интервальных медиан в сочетании с оптимизацией функционала Жаккара, как показавшие наилучшие результаты в данном исследовании.

Таким образом, работа подтвердила практическую ценность интервальных статистик и коэффициента Жаккара для решения задач параметрического оценивания в условиях неопределенности измерений.

7 Приложение

Код программы GitHub URL:
`https://github.com/Akira1707/IntervalAnalysis/tree/main/Lab%204`