

Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной
физики

**Отчет по лабораторной работе №3
“Интервальный анализ”**

Выполнили студент группы 5030102/10201:

Умнов Сергей Алексеевич

Преподаватель:

Баженов Александр Николаевич

Санкт-Петербург
2024

Содержание

1	Постановка задачи	2
2	Теория	2
2.1	Интервальная мода	2
2.2	Интервальная медиана Крейновича	3
2.3	Интервальная медиана Пролубникова	3
2.4	Коэффициент Жаккара	3
3	Реализация	3
3.1	Алгоритм	3
4	Анализ и обработка данных	4
5	Нахождение констант	8
6	Результаты	9
7	Выводы	13

1 Постановка задачи

Даны 2 интервальных выборки

$$\mathbf{X} = \{\mathbf{x}_i\}, \quad (1)$$

$$\mathbf{Y} = \{\mathbf{y}_i\}. \quad (2)$$

Взять \mathbf{X}, \mathbf{Y} из файлов данных, задав $\text{rad}\mathbf{x} = \text{rad}\mathbf{y} = \frac{1}{2^N}B$, $N = 14$.

Файлы данных:

- *-0.205_lvl_side_a_fast_data.bin*
- *0.225_lvl_side_a_fast_data.bin*

Связь кодов данных и B :

$$V = N/16384 - 0.5$$

Сделать оценки констант a, t в уравнениях:

$$\mathbf{X} + a = \mathbf{Y}, \quad (3)$$

$$t\mathbf{X} = \mathbf{Y}, \quad (4)$$

Метод решения:

$$\hat{a} = \operatorname{argmax} F(a, \mathbf{X}, \mathbf{Y}), \quad (5)$$

где F — функционал.

В качестве функционала взять варианты:

$$J_i(a, \mathbf{X}, \mathbf{Y}), \quad (6)$$

$$J_i(a, \text{mode}\mathbf{X}, \text{mode}\mathbf{Y}), \quad (7)$$

$$J_i(a, \text{med}_K\mathbf{X}, \text{med}_K\mathbf{Y}), \quad (8)$$

$$J_i(a, \text{med}_P\mathbf{X}, \text{med}_P\mathbf{Y}), \quad (9)$$

где J_i — коэффициент Жаккара, mode — интервальная мода, $\text{med}_K, \text{med}_P$ — интервальные медианы Крейновича и Пролубникова.

Сделать точечные и интервальные оценки, задавшись уровнем α .

2 Теория

2.1 Интервальная мода

Пусть имеется интервальная выборка

$$\mathbf{X} = \{\mathbf{x}_i\}.$$

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

Для каждого интервала \mathbf{z}_i подсчитываем число μ_i интервалов из выборки \mathbf{X}_i , включающих \mathbf{z}_i . Максимальные $\mu_i = \max \mu$ достигаются для индексного множества K . Тогда можно найти интервальную моду как мультиинтервал.

$$\text{mode}\mathbf{X} = \bigcup_{k \in K} \mathbf{z}_k. \quad (10)$$

2.2 Интервальная медиана Крейновича

Пусть дана выборка $\mathbf{X} = \{\mathbf{x}_i\}$.

Пусть $\underline{c} = \{\underline{x}_i\}$, $\bar{c} = \{\bar{x}_i\}$ — конфигурация точек, составленных, соответственно, из левых и правых концов интервалов из \mathbf{X} .

Тогда медианой Крейновича $\text{med}_K \mathbf{X}$ интервальной выборки \mathbf{X} — это интервал

$$\text{med}_K = [\text{med}_c, \text{med}\bar{c}]. \quad (11)$$

2.3 Интервальная медиана Пролубникова

Медиана Пролубникова $\text{med}_P \mathbf{X}$ выборки \mathbf{X} — это интервал \mathbf{x}_m , для которого половина интервалов из \mathbf{X} лежит слева, а половина — справа.

В ситуации, когда имеются два элемента подинтервала \mathbf{x}_m и \mathbf{x}_{m+1} , расположенных посередине вариационного ряда, $\mathbf{x}_m \neq \mathbf{x}_{m+1}$ медиана может быть определена естественным обобщением взятия полусуммы точечных значений, расположенных посередине ряда из точечных значений, в случае интервальной выборки взятие полусуммы интервалов \mathbf{x}_m и \mathbf{x}_{m+1} :

$$\text{med}_P \mathbf{X} = (\mathbf{x}_m + \mathbf{x}_{m+1})/2. \quad (12)$$

2.4 Коэффициент Жаккара

Коэффициент Жаккара для двух интервалов $\mathbf{x} \in \mathbb{IR}$ и $\mathbf{y} \in \mathbb{IR}$:

$$\text{Ji}(\mathbf{x}, \mathbf{y}) = \frac{\text{wid}(x \wedge y)}{\text{wid}(x \vee y)} = \frac{\min\{\bar{x}, \bar{y}\} - \max\{\underline{x}, \underline{y}\}}{\max\{\bar{x}, \bar{y}\} - \min\{\underline{x}, \underline{y}\}}. \quad (13)$$

Коэффициент Жаккара для множества интервалов $\mathbf{X} \in \mathbb{IR}^n$:

$$\text{Ji}(\mathbf{X}) = \frac{\min \bar{x}_i - \max \underline{x}_i}{\max \bar{x}_i - \min \underline{x}_i}. \quad (14)$$

Коэффициент Жаккара для двух множеств интервалов $\mathbf{X} \in \mathbb{IR}^n$ и $\mathbf{Y} \in \mathbb{IR}^n$:

$$\text{Ji}_k(\mathbf{X}, \mathbf{Y}) = \frac{\min\{\bar{x}_k, \bar{y}_k\} - \max\{\underline{x}_k, \underline{y}_k\}}{\max\{\bar{x}_k, \bar{y}_k\} - \min\{\underline{x}_k, \underline{y}_k\}}, \quad k = 1, 2, \dots, |\mathbf{X}|. \quad (15)$$

3 Реализация

Лабораторная работа выполнена на языке программирования Python. В ходе работы были также использованы библиотеки `numpy` и `intervalpy`

Ссылка на GitHub репозиторий: <https://github.com/vainmoon/interval-analysis>

3.1 Алгоритм

Для поиска параметров, при которых функционал достигал наибольших значений, был использован алгоритм тернарного поиска с заданной точностью $\varepsilon = 10^{-3}$ на участках, где функции вели себя как унимодальные.

Его же будем использовать для расчёта интервальной оценки параметров a, t с уровнем значимости $\alpha = 0.05$

4 Анализ и обработка данных

Данные представляют собой 100 фреймов на 8 каналах по 1024 пикселя.

Рассмотрим распределение данных фрейма в выборках **X** и **Y** для 1 канала и 1 пикселя

Рис. 1: Гистограмма распределения 1 пикселя и 1 канала в выборке X

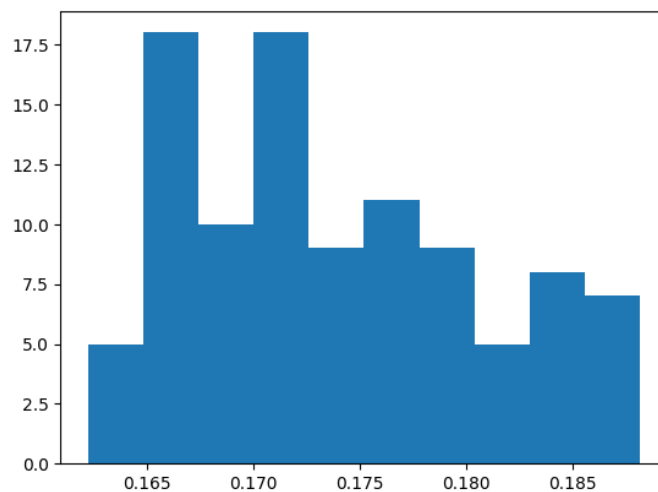


Рис. 2: Гистограмма распределения 1 пикселя и 1 канала в выборке Y

Рассмотрим графики рассеяния в выборках X и Y

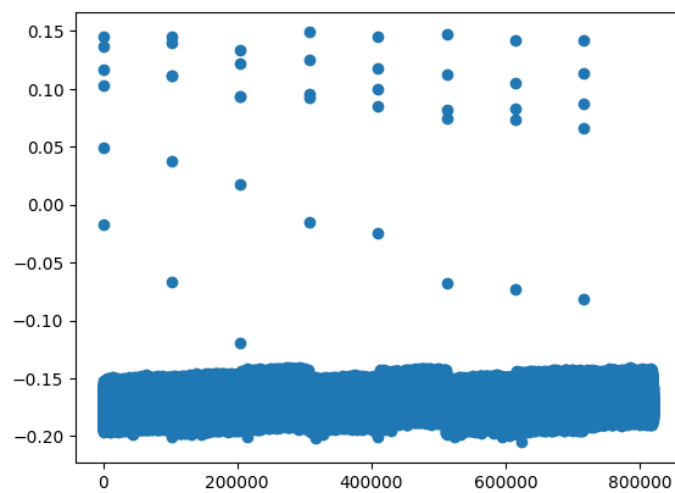


Рис. 3: График рассеяния данных в выборке X

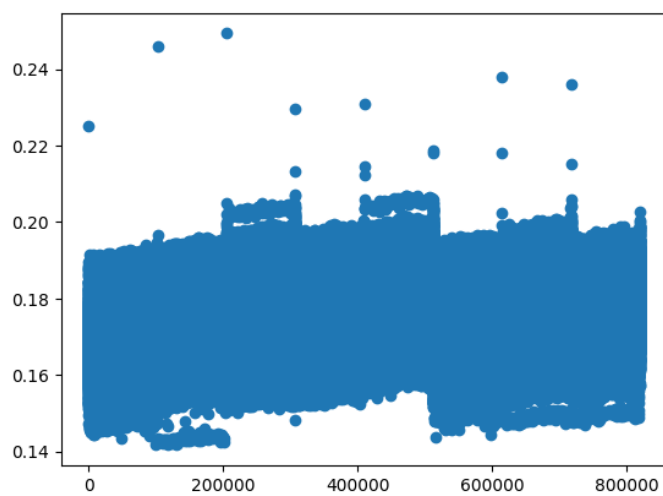


Рис. 4: График рассеяния данных в выборке Y

Очевидно, для данных требуется усреднение, чтобы уменьшить вариативность данных.

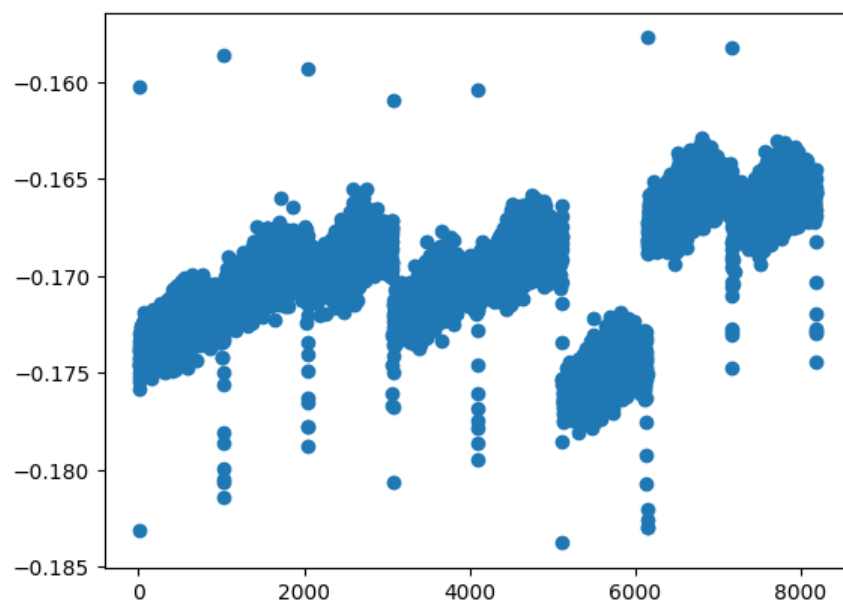


Рис. 5: Усредненные данные в выборке X

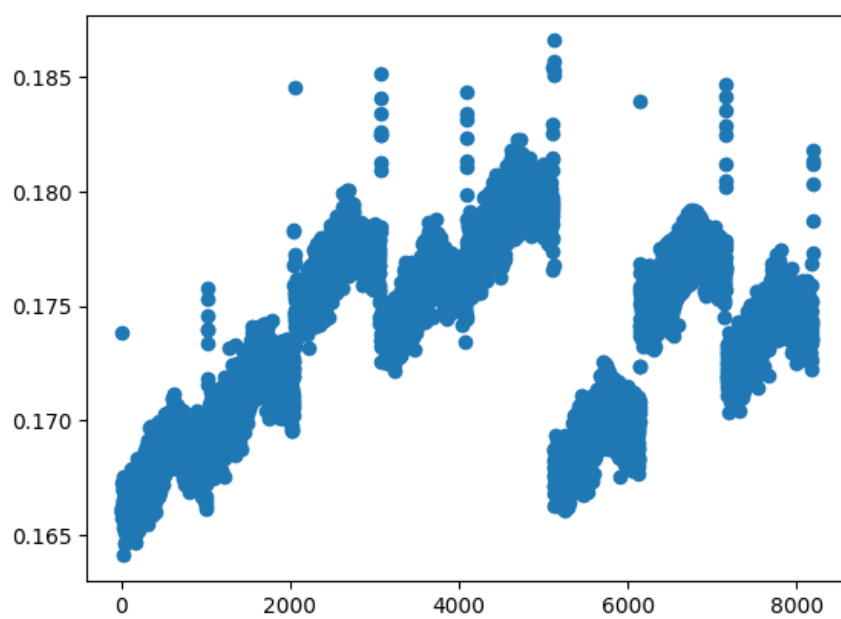


Рис. 6: Усредненные данные в выборке Y

Теперь наблюдается распределение данных по 8 имеющимся каналам.
Рассмотрим расположение данных в 1-ом канале:

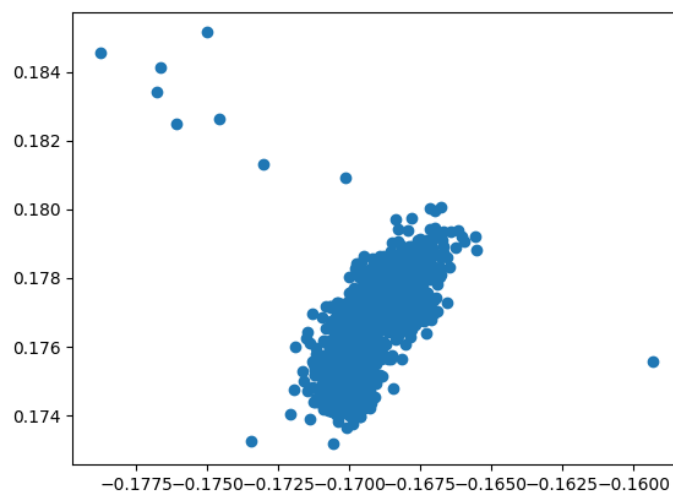


Рис. 7: Необработанные данные $Y(X)$

Удалим выбросы данных, используя правило трёх сигм.

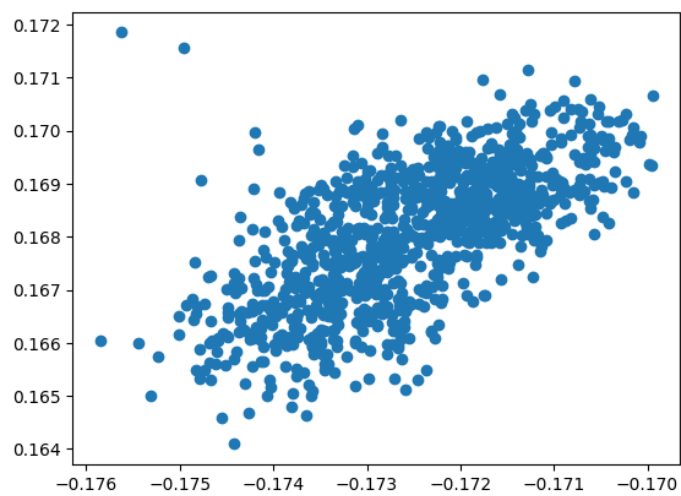


Рис. 8: Обработанные данные $Y(X)$

5 Нахождение констант

Константы будем находить при помощи тернарного поиска, поскольку (как будет видно далее из графиков) функция зависимости функционала от параметра на рассматриваемом участке ведёт себя как унимодальная, а значит, у неё есть единственный максимум / минимум. Интервалы для поиска возьмём сознательно больше значений крайних левых и правых точек.

6 Результаты

Для функционала со стандартными интервальными выборками (6):

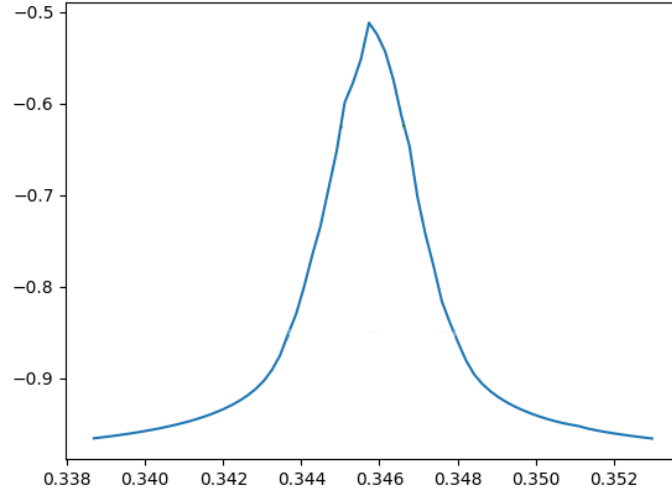


Рис. 9: Зависимость функционала от параметра a

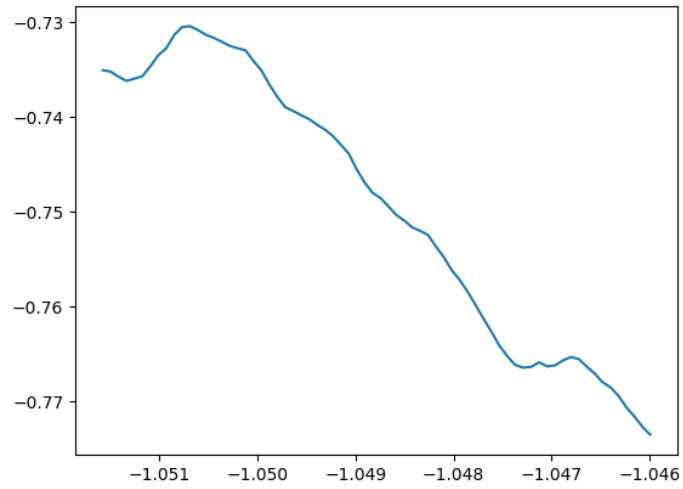


Рис. 10: Зависимость функционала от параметра t

$$\hat{a} = 0.34600 \pm 0.0005, \quad F_1(\hat{a}) = -0.94918,$$

$$\hat{t} = -1.05068 \pm 0.0005, \quad F_1(\hat{t}) = -0.92734.$$

Для функционала с интервальными модами (7):

$$\begin{aligned}\hat{a} &= 0.34811 \pm 0.0005, \quad F_2(\hat{a}) = -0.25437, \\ \hat{t} &= -1.04832 \pm 0.0005, \quad F_2(\hat{t}) = -0.92750.\end{aligned}$$

Для функционала с интервальными медианами Крейновича (8):

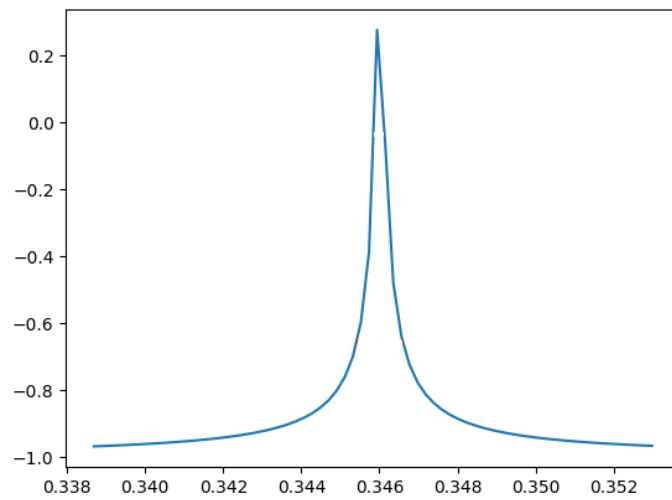


Рис. 11: Зависимость функционала от параметра a

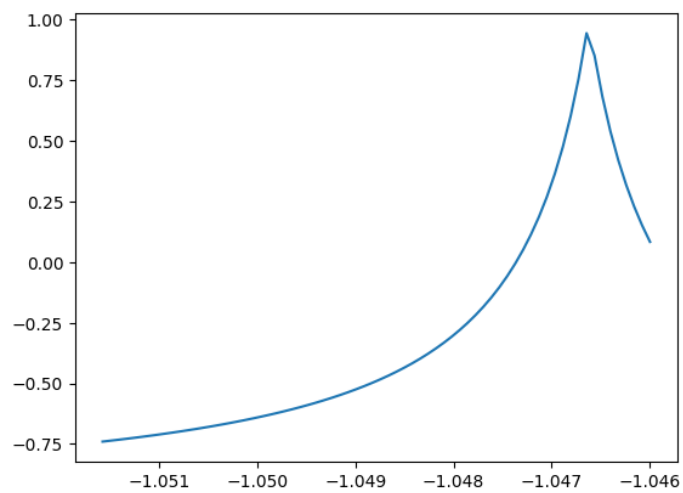


Рис. 12: Зависимость функционала от параметра t

$$\hat{a} = 0.34589 \pm 0.0005, \quad F_3(\hat{a}) = -0.00184,$$

$$\hat{t} = -1.04527 \pm 0.0005, \quad F_3(\hat{t}) = 0.63020.$$

Для функционала с интервальными медианами Пролубникова (9):

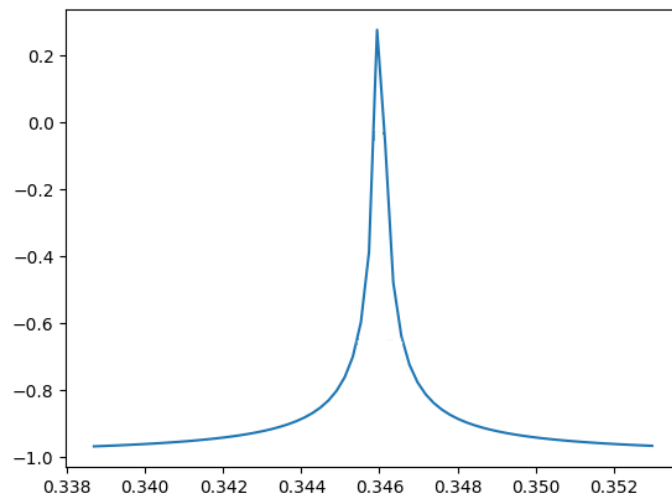


Рис. 13: Зависимость функционала от параметра a

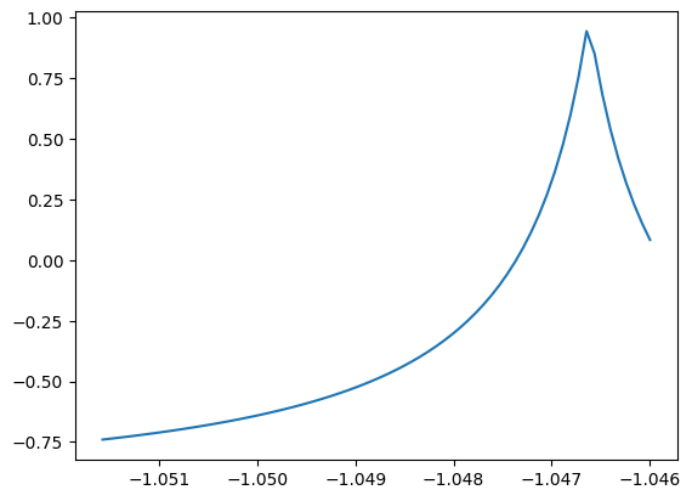


Рис. 14: Зависимость функционала от параметра t

$$\hat{a} = 0.34589 \pm 0.0005, \quad F_4(\hat{a}) = -0.12457,$$

$$\hat{t} = -1.04527 \pm 0.0005, \quad F_4(\hat{t}) = 0.63021.$$

7 Выводы

В процессе выполнения лабораторной работы были рассмотрены подходы к оценке параметров в уравнениях, содержащих интервальные данные. На основе использования различных функционалов, включая коэффициент Жаккара, были определены оптимальные значения параметров \hat{a} и \hat{t} для уравнений вида $\mathbf{X} + a = \mathbf{Y}$ и $t\mathbf{X} = \mathbf{Y}$.

На основании полученных результатов можно сделать следующие выводы:

1. Значения параметров \hat{a} и \hat{t} зависят от выбранного функционала. Это подчеркивает необходимость тщательного подхода к выбору критерия оптимальности при решении задач интервального анализа.
2. Наиболее устойчивые значения были получены при использовании функционала с интервальными медианами Крейнвича (8). Оптимальное значение параметра \hat{t} в этом случае демонстрирует положительное значение коэффициента Жаккара, что свидетельствует о хорошем уровне совпадения интервалов.
3. Использование интервальной моды и медиан (в интерпретации Крейнвича и Пролубникова) в качестве статистических характеристик позволило повысить точность оценок параметров, что подчеркивает их значимость для анализа данных с интервальной природой.

Таким образом, выполненная работа показала практическую применимость и эффективность методов интервального анализа для задач параметрической оценки, а также акцентировала внимание на важности выбора адекватных инструментов и методов для обработки данных.