Санкт-Петербургский политехнический университет
Петра Великого
Физико-механический институт
Высшая школа прикладной математики и вычислительной физики

Отчет по лабораторной работе №3
по дисциплине
"Интервальный анализ"

Обработка константы

Выполнил студент:

Кромачев Максим Александрович
группа: 5030102/10201

Проверил:
доцент
Баженов Александр Николаевич

Санкт-Петербург
2024

Содержание

1 Теоретическое обоснование

Необходимые формулы:

- Выборочная медиана
$$med\ x = \begin{cases} x_{(l+1)} & \text{при } n = 2l+1 \\ \frac{x_{(l)}+x_{(l+1)}}{2} & \text{при } n = 2l \end{cases}$$

- Ширина интервала:
$$\text{wid} = \overline{a} - \underline{a}, \quad \text{где } [\underline{a}, \overline{a}] \text{ — интервал.} \tag{1}$$

- Середина интервала:
$$\text{mid} = \frac{\underline{a} + \overline{a}}{2}, \quad \text{где } [\underline{a}, \overline{a}] \text{ — интервал.} \tag{2}$$

- Радиус интервала:
$$\text{rad} = \frac{\text{wid}}{2} = \frac{\overline{a} - \underline{a}}{2}. \tag{3}$$

- Минимум по включению:
$$a \wedge b := \inf \subseteq \{a, b\} = [\max\{\underline{a}, \underline{b}\}, \min\{\overline{a}, \overline{b}\}] \tag{4}$$

- Максимум по включению:
$$a \wedge b := \sup \subseteq \{a, b\} = [\min\{\underline{a}, \underline{b}\}, \max\{\overline{a}, \overline{b}\}] \tag{5}$$

- *Модой* интервальной выборки назовём совокупность интервалов пересечения наибольших совместных подвыборок рассматриваемой выборки. 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
$$\text{mode } \mathbf{X} = \bigcup_{k \in K} \mathbf{z}_k.$$

- *Интервальная медиана Крейновича*

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

  Тогда медианой Крейновича $\text{med}_K\mathbf{X}$ интервальной выборки $\mathbf{X}$ — это интервал
$$\text{med}_K = [\text{med } \underline{c}, \text{med } \overline{c}].$$
Сложность вычисления $\text{med}_k X$ составляет $O(N)$. При определении медианы для интервальных данных мы руководствуемся принципом соответствия. Этот принцип можно сформулировать так: если ширина интервалов стремится к нулю, то значения медианы, рассчитанные для интервальных данных, будут приближаться к значениям медианы для точечных данных, к которым стремятся эти интервалы.

- *Интервальная медиана Пролубникова*

  Для того, чтобы распространить определение медианы точечных данных на интервальные данные, требуется задать для них линейный порядок ($\leq$) и частоту. Возможные способы задания отношения порядка на IR определены стандартом 1788 IEEE. Говорят, что неравенство ($\mathbf{a} \leq \mathbf{b}$) выполняется:

1. в сильном смысле, если $(\forall a \in \mathbf{a})(\forall b \in \mathbf{b})(a \le b)$;
2. в слабом смысле, если $(\exists a \in \mathbf{a})(\exists b \in \mathbf{b})(a \le b)$;
3. в $\forall\exists$-смысле, если $(\forall a \in \mathbf{a})(\exists b \in \mathbf{b})(a \le b)$;
4. в $\exists\forall$-смысле, если $(\exists a \in \mathbf{a})(\forall b \in \mathbf{b})(a \le b)$;
5. в центральном смысле, если $\frac{\underline{a}+\overline{a}}{2} \le \frac{\underline{b}+\overline{b}}{2}$.

Для элементов выборки $\tilde{\mathbf{X}}$, сформированной на основе выборки $X = \{x_i\}_{i=1}^N$ и содержащей данные с интервальными неопределённостями, возникшими в результате группировки, можно определить линейный порядок, используя любое из пяти вышеуказанных отношений порядка на IR. То есть, если $i \ne j$, то либо $\tilde{x}_i \le \tilde{x}_j$, либо $\tilde{x}_i \ge \tilde{x}_j$ для любого из этих отношений порядка. Это позволяет рассматривать множество $\tilde{\mathbf{X}}$ для сгруппированных данных как интервальный вариационный ряд.

Медиана Пролубникова $\mathrm{med}_P\mathbf{X}$ 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, $\mathbf{x}_m \ne \mathbf{x}_{m+1}$ 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 $\mathbf{x}_{m+1}$:

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

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

$$\mathrm{Ji}(\mathbf{x},\mathbf{y}) = \frac{\mathrm{wid}(x \wedge y)}{\mathrm{wid}(x \vee y)} = \frac{\min\{\overline{\mathbf{x}},\overline{\mathbf{y}}\} - \max\{\underline{\mathbf{x}},\underline{\mathbf{y}}\}}{\max\{\overline{\mathbf{x}},\overline{\mathbf{y}}\} - \min\{\underline{\mathbf{x}},\underline{\mathbf{y}}\}}.$$

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

$$\mathrm{Ji}(\mathbf{X}) = \frac{\min\overline{\mathbf{x}_i} - \max\underline{\mathbf{x}_i}}{\max\overline{\mathbf{x}_i} - \min\underline{\mathbf{x}_i}}.$$

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

$$\mathrm{Ji}_k(\mathbf{X},\mathbf{Y}) = \frac{\min\{\overline{\mathbf{x}_k},\overline{\mathbf{y}_k}\} - \max\{\underline{\mathbf{x}_k},\underline{\mathbf{y}_k}\}}{\max\{\overline{\mathbf{x}_k},\overline{\mathbf{y}_k}\} - \min\{\underline{\mathbf{x}_k},\underline{\mathbf{y}_k}\}},\ k \in 1,2,\dots,|\mathbf{X}|.$$

2 Постановка задачи

Даны 2 интервальных выборки

$$\mathbf{X} = \{x_i\}, \tag{6}$$

$$\mathbf{Y} = \{y_k\}, \tag{7}$$

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

$$0.205\_lvl\_side\_a\_fast\_data.bin$$

$$0.227\_lvl\_side\_a\_fast\_data.bin$$

Формат файлов — Save to BIN.pdf.
Связь кодов данных и В:

$$V = Code/163840.5.$$

Сделать оценки констант a, t в уравнениях.

$$a + \mathbf{X} = \mathbf{Y}, \tag{8}$$

$$t * \mathbf{X} = \mathbf{Y}, \tag{9}$$

Метод решения:

$$\hat{a} = \text{argmax} F(a, \mathbf{X}, \mathbf{Y}), \tag{10}$$

где F — функционал.
В качестве функционала взять варианты:

$$\text{Ji}(a, \mathbf{X}, \mathbf{Y}), \tag{11}$$

$$\text{Ji}(a, \text{mode}\mathbf{X}, \text{mode}\mathbf{Y}), \tag{12}$$

$$\text{Ji}(a, \text{med}_K\mathbf{X}, \text{med}_K\mathbf{Y}), \tag{13}$$

$$\text{Ji}(a, \text{med}_P\mathbf{X}, \text{med}_P\mathbf{Y}), \tag{14}$$

где Ji — коэффициент Жаккара, mode — интервальная мода, med_K, med_P — интервальные медианы Крейновича и Пролубникова.
Сделать точечные и интервальные оценки, задавшись уровнем α.

3 Описание работы

Лабораторная работа выполнена на языке программирования Python в среде разработки VSCode. В ходе работы были использованы следующие библиотеки: numpy, intvalpy, pandas. GitHub репозиторий: `https://github.com/kromachmax/Intervalka`

4 Описание алгоритма

Для поиска максимума функционалов будем использовать метод дихотомии для интервала, где функционал унимодален
Вход:
$f(x)$-унимодальная на $[a_0, b_0]$ функция;
a_0, b_0 - крайние точки интервала неопределенности;
ε - точность, с которой необходимо найти максимум функции $f(x)$
Выход:
x_* - точка максимума функции $f(x)$;
$f(x_*)$ - значение функции $f(x)$ в точке максимума

5 Результаты

Дальше в таблице *const* - это константа, для которой ищем оценку.
Возъмём уровень $\alpha = 0.001$

$\text{Ji}(a/t, \mathbf{X}, \mathbf{Y})$		
const	Оценка	Значение функционала
$\hat{a}$	0.3409 ± 0.001	-0.7867
$\hat{t}$	-1.0502 ± 0.001	-0.8610

$\text{Ji}(a/t, \text{mode}\mathbf{X}, \text{mode}\mathbf{Y})$		
const	Оценка	Значение функционала
$\hat{a}$	0.3402 ± 0.001	-0.25437
$\hat{t}$	-0.9987 ± 0.001	-0.92750

$\text{Ji}(a/t, \text{med}_K\mathbf{X}, \text{med}_K\mathbf{Y})$		
const	Оценка	Значение функционала
$\hat{a}$	0.3436 ± 0.001	0.1351
$\hat{t}$	-1.0144 ± 0.001	0.7746

$\text{Ji}(a/t, \text{med}_P\mathbf{X}, \text{med}_P\mathbf{Y})$		
const	Оценка	Значение функционала
$\hat{a}$	0.3436 ± 0.001	0.1351
$\hat{t}$	-1.0144 ± 0.001	0.7746

5.1 Графики

Продемонстрируем результаты вычислений на графиках.

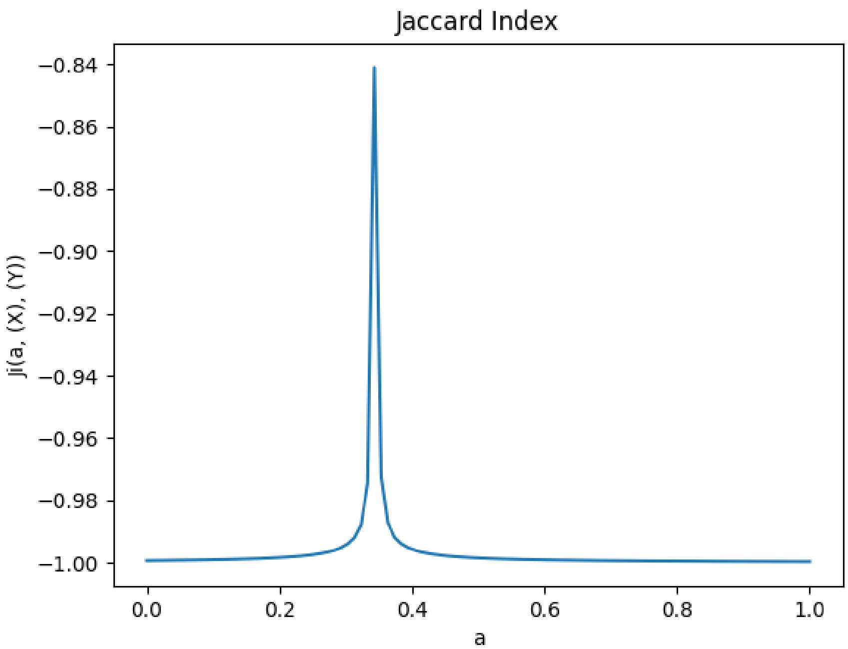

Рис. 1: Коэффициент Жаккара для a, функционал (6)

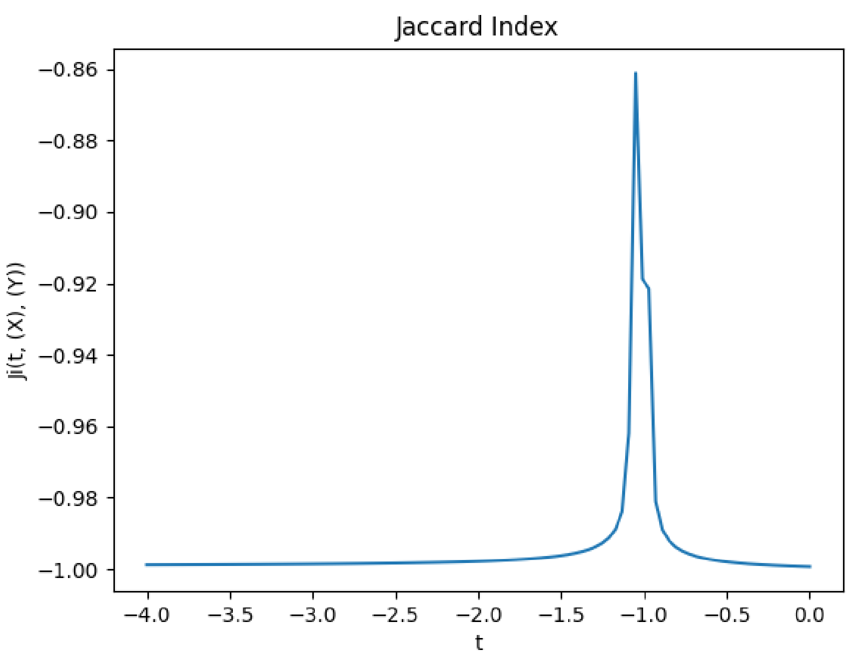

Рис. 2: Коэффициент Жаккара для t, функционал (6)

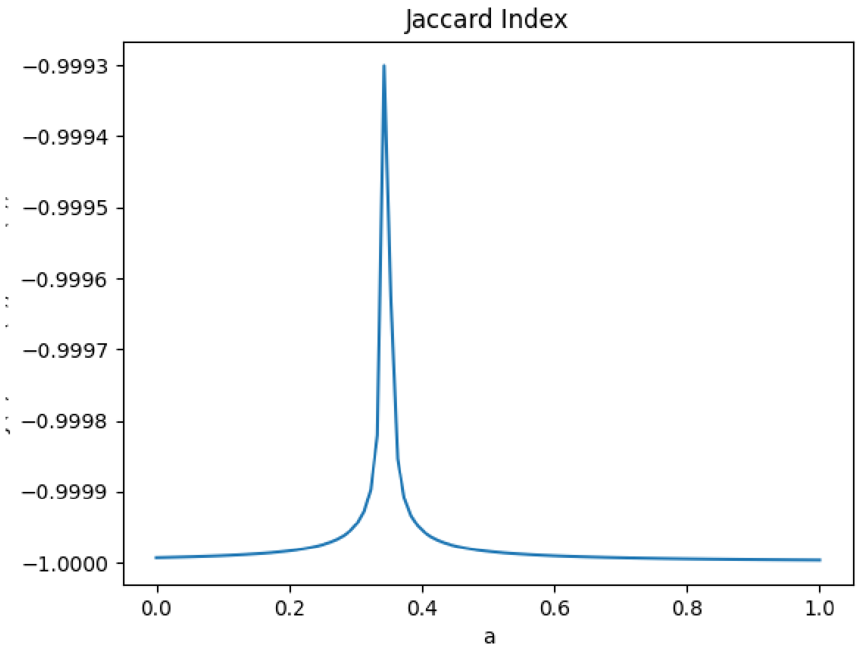

Рис. 3: Коэффициент Жаккара для a, функционал (7)

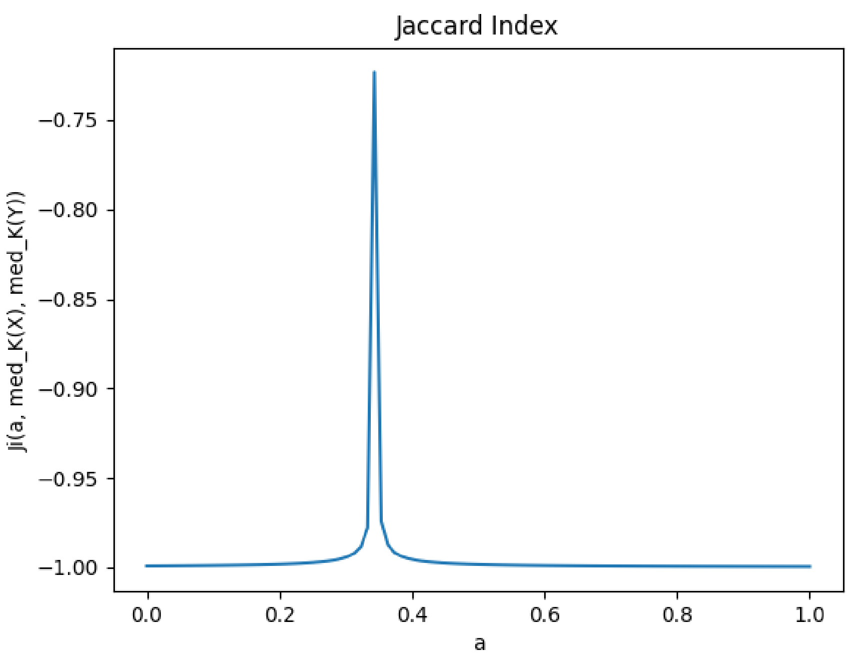

Рис. 4: Коэффициент Жаккара для a, функционал (8)

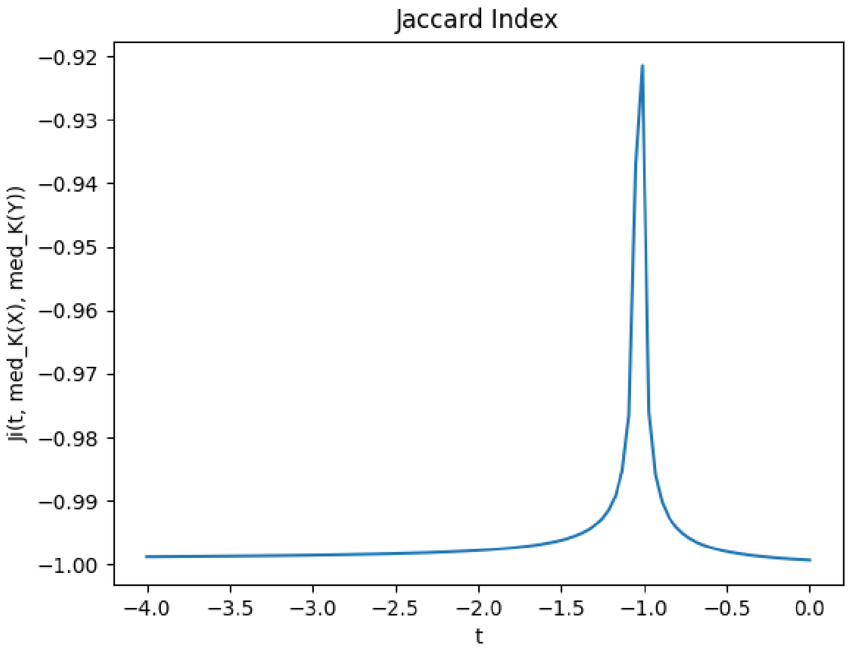

Рис. 5: Коэффициент Жаккара для t, функционал (8)

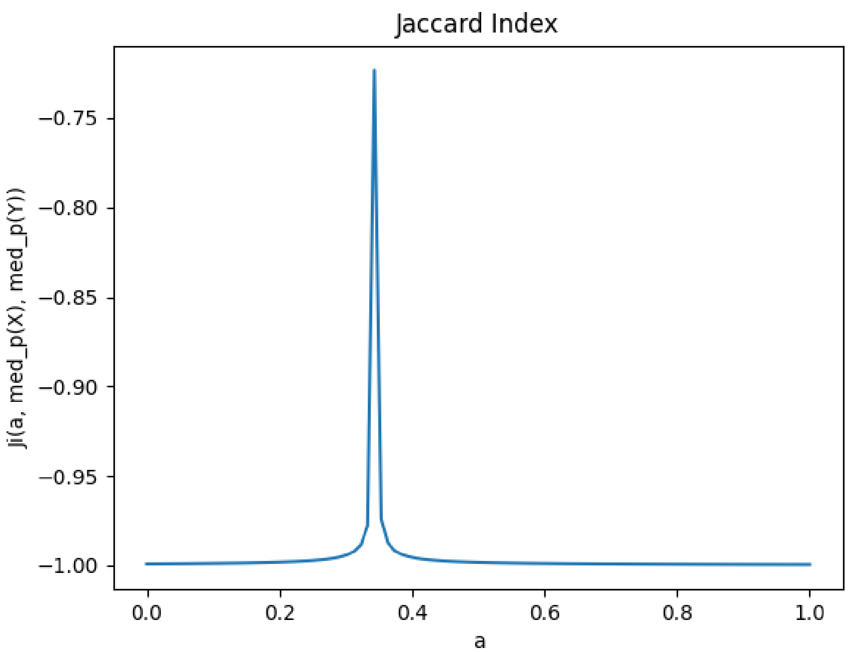

Рис. 6: Коэффициент Жаккара для a, функционал (9)

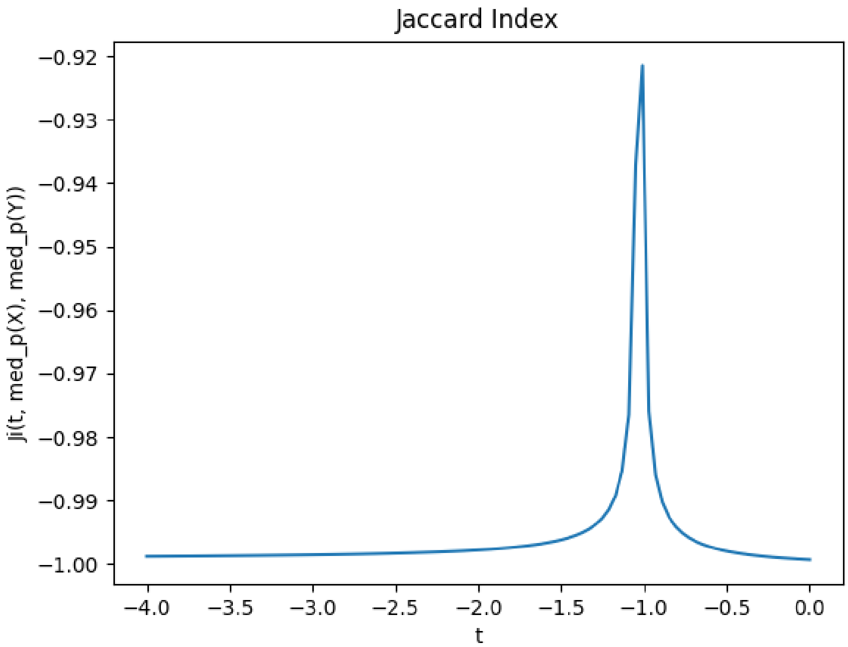

Рис. 7: Коэффициент Жаккара для t, функционал (9)

6 Итоги

В процессе выполнения лабораторной работы мы ознакомились с методами получения интервальных оценок для констант. Фокусировались на четырех различных функционалах ((6), (7), (8), (9)), основанных на вычислении коэффициента Жаккара, но применяемых к различным интервальным множествам.

Мы пришли к следующим выводам:

1. Важно тщательно подбирать функционал для поиска оценок. Разные функционалы могут приводить к несовместимости интервалов, как это наблюдалось при использовании функционала $Ji(const, \mathbf{X}, \mathbf{Y})$. В то же время, для функционалов $\mathrm{Ji}(const, \mathrm{med}_K \mathbf{X}, \mathrm{med}_K \mathbf{Y})$ и $\mathrm{Ji}(const, \mathrm{med}_P \mathbf{X}, \mathrm{med}_P \mathbf{Y})$ интервалы оказались совместимыми.

2. Можно отметить, что использование функционалов, зависящих от медиан (8), (9), приводит к коэффициенту Жаккара, близкому к 1 (значение 0.7746) для оценки t, что указывает на высокую степень совпадения выборок $t * \mathbf{X}$ и Y.

3. В ходе исследования мы также обнаружили, что функционал, зависящий от моды (7), требует значительно больше времени на вычисление оценок по сравнению с другими функционалами. Это свидетельствует о том, что для больших интервальных выборок использование коэффициента Жаккара, основанного на модах (7), нежелательно. Предпочтительнее использовать коэффициент Жаккара, основанный на одной из медиан (Крейновича или Пролубникова).

4. Кроме того, при нахождении медианы Пролубникова, если мы вводим порядок в алгебре IR в центральном смысле $((\overline{\mathbf{a}} + \underline{\mathbf{a}})/2 \leq (\overline{\mathbf{b}} + \underline{\mathbf{b}})/2)$, оценки, полученные с использованием функционала медианы Крейновича и медианы Пролубникова, совпадают.