

Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Кафедра «Прикладная математика»

**Отчёт по лабораторной работе №1
по дисциплине «Анализ данных с интервальной
неопределённостью»**

Выполнил студент:
Мишутин Дмитрий Валерьевич
группа: 5040102/20201

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2024 г.

Содержание

1	Постановка задачи	2
2	Теория	2
2.1	Индекс Жаккара	2
2.2	Частота моды	2
2.3	Оптимальный корректирующий множитель в методе центра неопределённости	2
2.4	Мультипликативная мера	3
2.5	Нахождение оптимального значения R	3
3	Реализация	4
4	Результаты	4
5	Выводы	9
6	Литература	9
7	Приложения	9

Список иллюстраций

1	Исходная интервальная выборка X_1	4
2	Исходная интервальная выборка X_2	5
3	Частота пересечений подынтервалов с интервалами выборки X_1	5
4	Частота пересечений подынтервалов с интервалами выборки X_2	6
5	Зависимость JS от R	7
6	Зависимость k_0 от R	7
7	Зависимость частоты моды от R	8
8	Зависимость T от R	8

1 Постановка задачи

Имеется две вещественные выборки. Необходимо на их основе построить две интервальные выборки X_1, X_2 . Рассматриваются 4 меры совместности интервальных данных: индекс Жаккара, частота моды, оптимальный корректирующий множитель в методе центра неопределённости, мультипликативная мера. Каждую выборку необходимо оценить в отдельности с использованием каждой меры совместности. После, для каждой меры найти такие значения R , которые были бы оптимальным для выборки $X = X_1 \cup RX_2$.

2 Теория

2.1 Индекс Жаккара

Индекс Жаккара (далее обозначим через JC) определяет степень совместности двух интервалов x, y .

$$JC(x, y) = \frac{wid(x \wedge y)}{wid(x \vee y)} \quad (1)$$

Здесь \wedge, \vee представляют собой операции взятия минимума и максимума по включению в полной арифметике Каухера. Формула 1 легко может быть обобщена на случай интервальной выборки $X = \{x_i\}_{i=1}^n$.

$$JC(X) = \frac{wid(\wedge_{i=1,n} x_i)}{wid(\vee_{i=1,n} x_i)} \quad (2)$$

2.2 Частота моды

Модой интервальной выборки называют совокупность интервалов пересечения наибольших совместных подвыборок рассматриваемой выборки. Наибольшая длина совместных подвыборок данной выборки называется частотой моды. Исследование частоты моды (обозначим далее через $тах\mu$) имеет смысл только для несовместных выборок.

2.3 Оптимальный корректирующий множитель в методе центра неопределённости

Для обеспечения совместности выборки интервальных измерений применяется метод “центра неопределённости”. Если выборка измерений несов-

местна, то путем одновременного увеличения величины неопределенности всех измерений в выражении можно всегда добиться того, чтобы выборка стала совместной.

$$\overline{x'_i} = \overline{x_i} + k * (\overline{x_i} - \underline{x_i}) \quad (3)$$

$$\underline{x'_i} = \underline{x_i} - k * (\overline{x_i} - \underline{x_i}) \quad (4)$$

Оптимальным (k_0) называется такое значение k , при котором непустое пересечение интервальной выборки является точкой.

2.4 Мультипликативная мера

Мультипликативная мера T учитывает степень совместности по нескольким функционалам качества одновременно, а ее значения находятся в интервале $[0, 1]$:

$$T = \prod_{i=1}^k T_i \quad (5)$$

В качестве множителей берем перечисленные выше меры, нормированные в $[0, 1]$:

$$T_1 = \frac{1}{2} * (1 + JC) \quad (6)$$

$$T_2 = \frac{\max \mu}{n} \quad (7)$$

$$T_3 = \frac{1}{1 + k_0} \quad (8)$$

2.5 Нахождение оптимального значения R

Для поиска оптимальных значений R разумно найти первое приближение. Возьмем за такое приближение внешнюю оценку R_{out} .

$$\underline{R_{out}} = \frac{\min_{i=1,n} \underline{x_{1i}}}{\max_{i=1,n} \underline{x_{2i}}} \quad (9)$$

$$\overline{R_{out}} = \frac{\max_{i=1,n} \overline{x_{1i}}}{\min_{i=1,n} \overline{x_{2i}}} \quad (10)$$

Будем исследовать поведение R в области, заданной R_{out} .

3 Реализация

Из языка Python 3.12.2 были использованы следующие модули:

- “numpy” — генерация множества чисел;
- “matplotlib.pyplot” — построение и отображение графиков;
- “typing”, “annotations” — строгая типизация с аннотациями;
- “os” — взаимодействие с ОС.

4 Результаты

Данные были взяты из файлов $+0_5V/+0_5V_13.txt$ и $-0_5V/-0_5V_13.txt$.
С коррекцией при помощи вспомогательных данных из файла $ZeroLine/ZeroLine_13.txt$.
Размер выборок: 1024. Интервальная выборка на основе изначальных строится по формулам:

$$x = [x_0 - \varepsilon, x_0 + \varepsilon], \varepsilon = 2^{-14} \quad (11)$$

где x_0 - точечное значение.

Сначала посмотрим на исходные интервальные выборки X_1, X_2 .

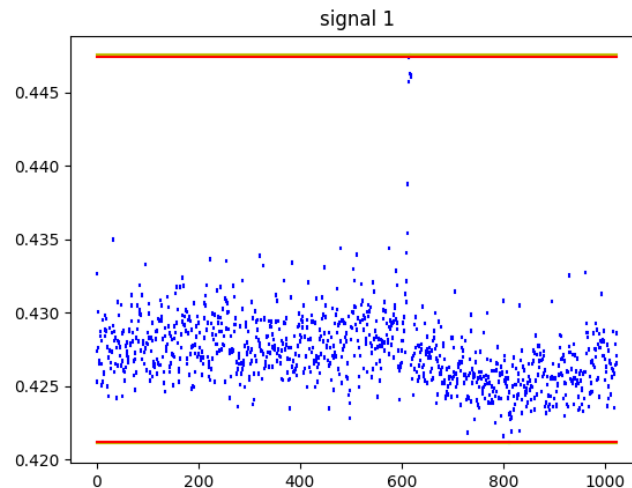


Рис. 1: Исходная интервальная выборка X_1

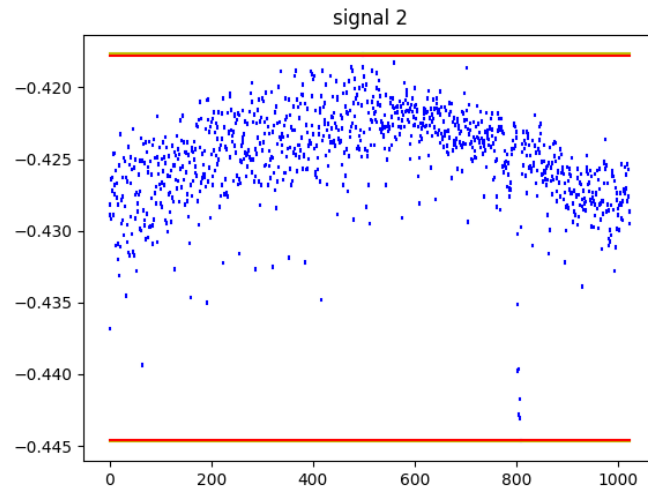


Рис. 2: Исходная интервальная выборка X_2

Также построим графики частоты пересечений подынтервалов для исходных выборок.

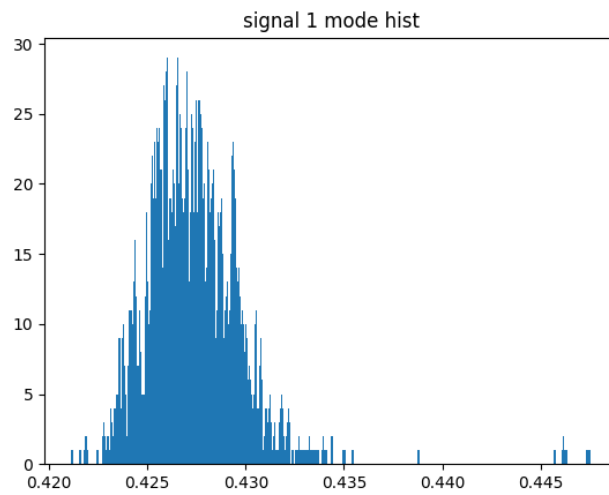


Рис. 3: Частота пересечений подынтервалов с интервалами выборки X_1

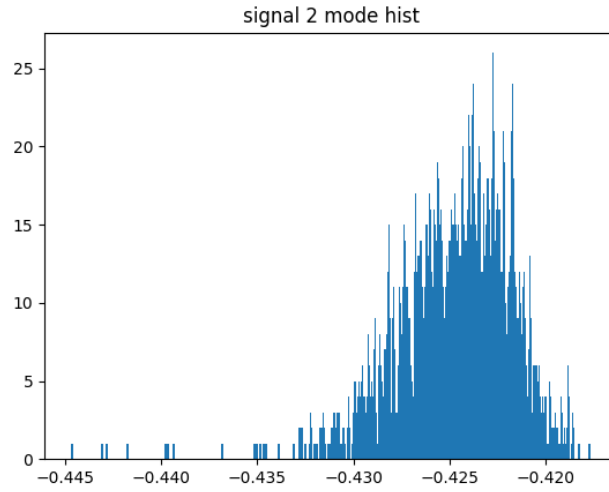


Рис. 4: Частота пересечений подынтервалов с интервалами выборки X_2

Проанализируем выборки.

Индекс Жаккара: $JC(X_1) = -0.9909$, $JC(X_2) = -0.9911$.

Частота моды: $max\mu(X_1) = 29$, $max\mu(X_2) = 26$.

Оптимальный корректирующий множитель: $k_0(X_1) = 109.375$, $k_0(X_2) = 111.666$.

Мультипликативная мера: $T(X_1) = 1.1623e^{-6}$, $T(X_2) = 1.0001e^{-6}$.

Верхняя и нижняя границы $\underline{R} = -1.0082$, $\bar{R} = -1.0065$.

Найдем оптимальные R (для наглядности на графиках изображены более широкие интервалы значений R).

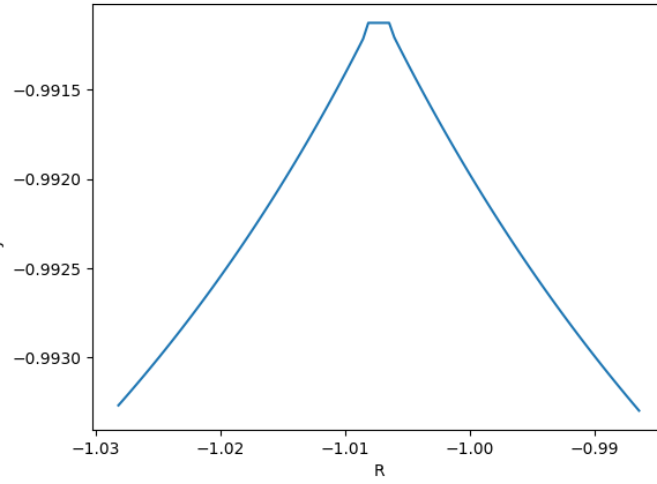


Рис. 5: Зависимость JC от R

Оптимальное значение R относительно JC равно $R_{opt} = [-1.0069, -1.0065]$ при $J(X) = -0.9911$.

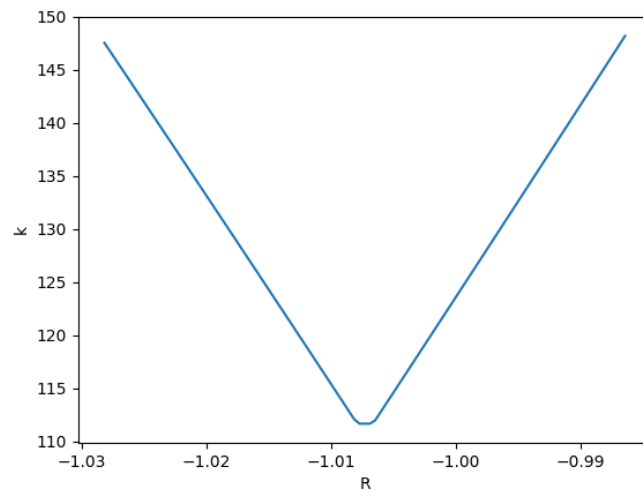


Рис. 6: Зависимость k_0 от R

Оптимальное значение R относительно k_0 равно $R_{opt} = [-1.0069, -1.0067]$ при $k_0(X) = 111.666$.

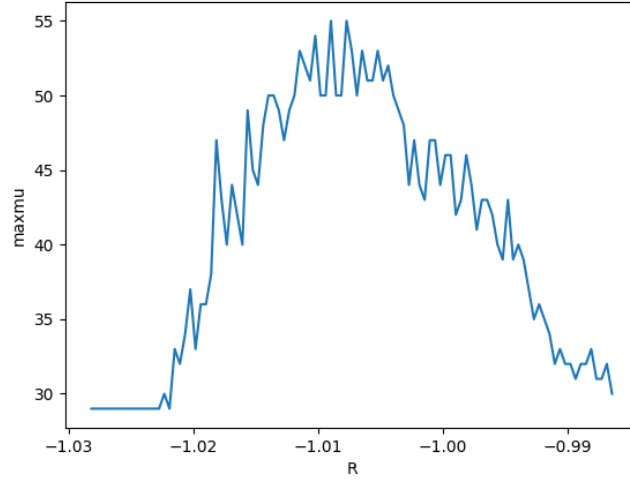


Рис. 7: Зависимость частоты моды от R

Оптимальное значение R относительно $\max\mu$ равно $R_{opt} = [-1.009, -1.0088], [-1.0078, -1.0075]$ при $\max\mu(X) = 55$.

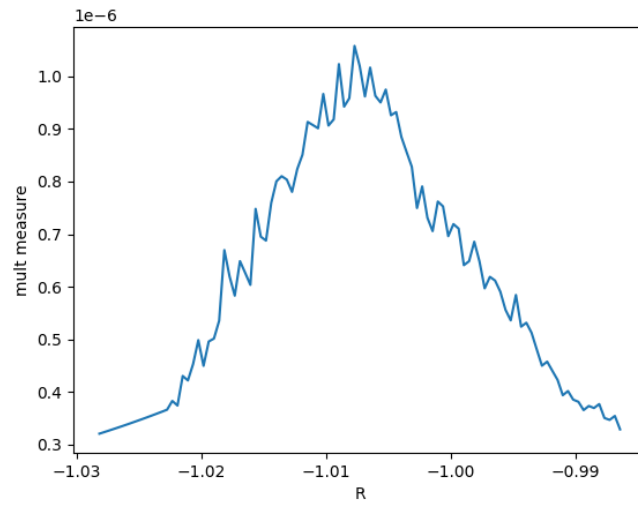


Рис. 8: Зависимость T от R

Оптимальное значение R относительно T равно $R_{opt} = -1.0077$ при $T(X) = 1.0578e^{-6}$.

Итоговая оценка совместности: $[-1.0077, [-1.0082, -1.0065]]$.

5 Выводы

Из полученных результатов можно заметить следующее. Из теории следует, что после совмещения двух выборок описанным образом нельзя получить более совместную выборку, чем худшая из двух изначальных. Это означает, что

$$J(X) \leq \min(JC(X_1), JC(X_2)) \quad (12)$$

$$k_0(X) \geq \max(k_0(X_1), k_0(X_2)) \quad (13)$$

$$\max \mu(X) \leq \max \mu(X_1) + \max \mu(X_2) \quad (14)$$

На практике во всех трех случаях удалось достичь равенства. Также можно заметить, что данное правило не распространяется на мультипликативную меру. Более того, только для мультипликативной меры R_{opt} задается точным числом, а не интервалом/мульти-интервалом.

6 Литература

- [Баженов А.Н. «Интервальный анализ. Основы теории и учебные примеры: учебное пособие»;](#)
- [Баженов А.Н. «Естественнонаучные и технические применения интервального анализа: учебное пособие»;](#)
- [Баженов А.Н. Репозиторий “Students” на GitHub;](#)

7 Приложения

Исходники лабораторной работы выложены на [GitHub](#).