Санкт-Петербургский политехнический университет Петра Великого

Физико-механический институт

Направление подготовки

«01.04.02 Прикладная математика и информатика»

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

Выполнила студентка гр. 5040102/20201 Харисова Т.А.

Преподаватель Баженов А.Н.

Санкт-Петербург

2023

Оглавление

Список иллюстраций

Список таблиц

Постановка задачи

Даны две интервальные выборки. Требуется определить для каждой выборки меру Жаккара, меру Оскорбина, моду и частоту, а также оценить меру совместности этих выборок.

Теория

Построение интервальных выборок из исходных точечных значений происходит с помощью метода “обынтерваливания”:

$$x = \dot{x} + [-\varepsilon, \varepsilon] \tag{1}$$

Здесь $\dot{x}$ – исходное точечное значение, ε – погрешность, x – итоговый интервал.

Размер выборки обозначен n, $X = \{x_i\}_{i=1}^{n}$ – интервальная выборка.

Внутренняя оценка интервальной выборки:

$$\underline{I} = \max_{1\le k\le n} \underline{x}_k \qquad \overline{I} = \min_{1\le k\le n} \overline{x}_k \qquad I = [\,\underline{I}, \overline{I}\,] \tag{2}$$

Внешняя оценка интервальной выборки:

$$\underline{J} = \min_{1\le k\le n} \underline{x}_k \qquad \overline{J} = \max_{1\le k\le n} \overline{x}_k \qquad J = [\,\underline{J}, \overline{J}\,] \tag{3}$$

Мера Жаккара:

$$JK = \frac{\min \overline{x}_k - \max \underline{x}_k}{\max \overline{x}_k - \min \underline{x}_k} \tag{4}$$

Алгоритм нахождения моды и частоты выборки:

1. Границы интервалов выборки сортируются по возрастанию значений: из интервальной выборки $[\underline{x}_i, \overline{x}_i\,]$, $i = 1..n$ получается набор значений $y_1, y_2, \dots, y_{2n}$, $y_j \le y_{j+1}$, $j = 1..2n-1$.
2. Формируется набор элементарных подынтервалов $z_j = [y_j, y_{j+1}]$, $j = 1..2n-1$.
3. Для каждого z_j подсчитывается число интервалов исходной выборки, включающих интервал z_j. Обозначим это число μ_j.
4. Частота выборки определяется как $\mu = \max\limits_{1\le j\le 2n-1} \mu_j$.
5. Мода выборки $mode\ X = \bigcup z_k$, где z_k – элементарные подынтервалы с $\mu_k = \mu$.

Если исходная выборка несовместна, ее можно сделать совместной методом центра неопределенностей:

$$x_i - k \operatorname{rad} x_i \le \beta \le x_i + k \operatorname{rad} x_i \tag{5}$$

где k – оптимальный корректирующий множитель (мера Оскорбина), $\operatorname{rad} x_i$ - радиус интервала

$$\operatorname{rad} x_i = \frac{1}{2}\left(\overline{x}_i - \underline{x}_i\right) \tag{6}$$

Для анализа совместности двух выборок X_1 и X_2 строится объединенная выборка

$$X = X_1 \cup R X_2 \tag{7}$$

Мера совместности R оценивается твином:

$$R \in [R_{in}, R_{out}] \qquad R_{in} = \frac{I_2}{I_1} \qquad R_{out} = \frac{J_2}{J_1} \tag{8}$$

Оценка R_{in} может быть уточнена: внутри интервала R_{in} находится значение R_{opt}, такое, что мера Жаккара объединенной выборки (7) при подстановке $R = R_{opt}$ принимает максимальное значение.

Реализация

Работа выполнена с помощью языка программирования Python в среде разработки Visual Studio Code. Ссылка на исходный код работы: Lab_1(github.com)

Для построения выборок использовались данные из файлов “+0_5V_1.txt” и ”-0_5V_1.txt”.

Результаты для скорректированных данных

Исходные данные подвергаются предварительной коррекции: вместо $\dot{x}_i$ рассматриваются $\dot{x}_i - \delta_i, i = 1..n,$ где δ_i – некоторая погрешность. Для “обынтерваливания” исходных данных берется $\varepsilon = 1/2^{14}$. Число интервалов в выборке $n = 1024$.

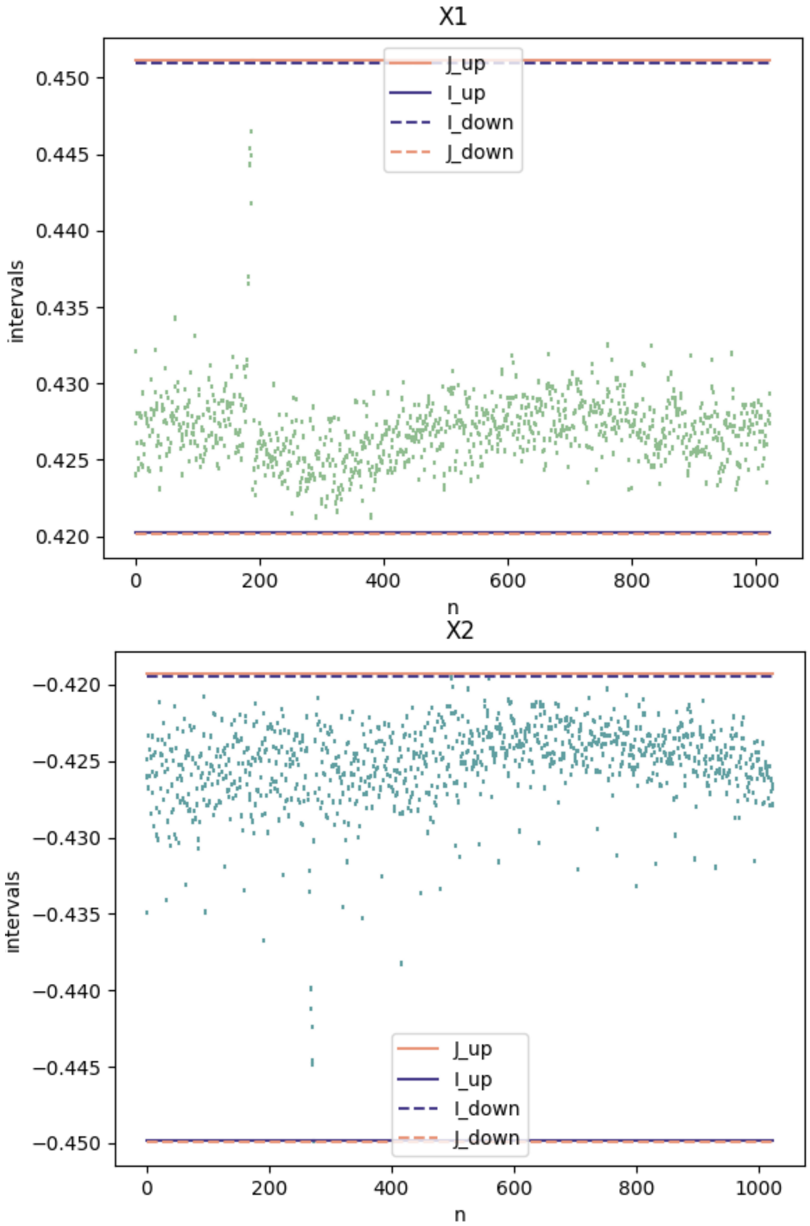

Рисунок 1. Интервальные выборки X_1 и X_2 и их оценки

Таблица 1. Меры интервальных выборок

	X_1	X_2
$[\underline{I}, \overline{I}]$	$[0.45099, 0.42023]$	$[-0.41944, -0.44983]$
$[\underline{J}, \overline{J}]$	$[0.42011, 0.45111]$	$[-0.44995, -0.41932]$
JK	-0.99226	-0.99216
k	503.97184	497.90976
μ	33	31
$mode$	$[[0.42676, 0.42681], [0.42731, 0.42736]]$	$[[-0.42492, -0.42487], [-0.42425, -0.42420]]$

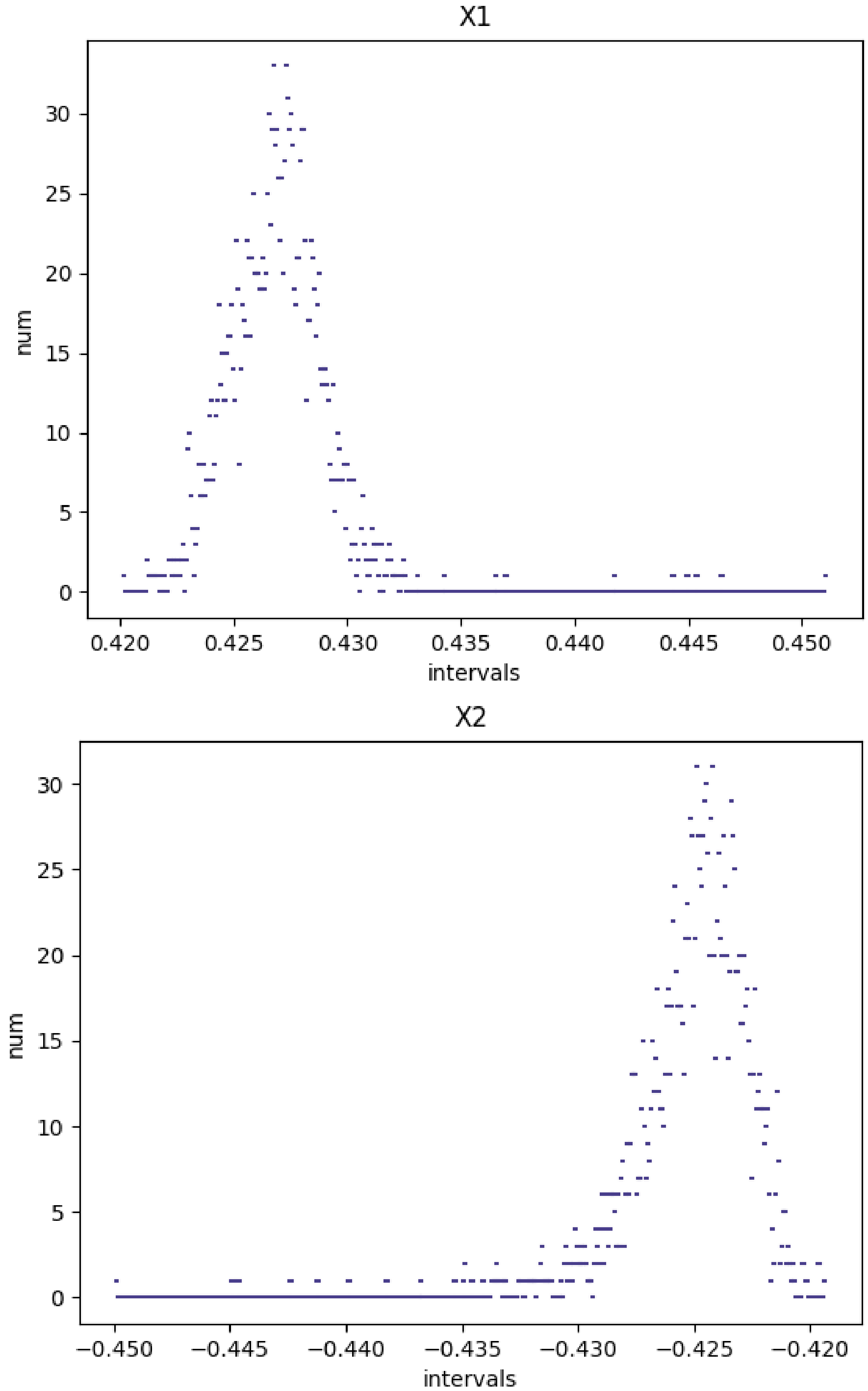

Рисунок 2. Частота выборок X_1 и X_2

Таблица 2. Оценки меры совместности выборок X_1 и X_2

R_{in}	$[-1.07043, -0.93004]$
R_{out}	$[-1.07103, -0.92952]$

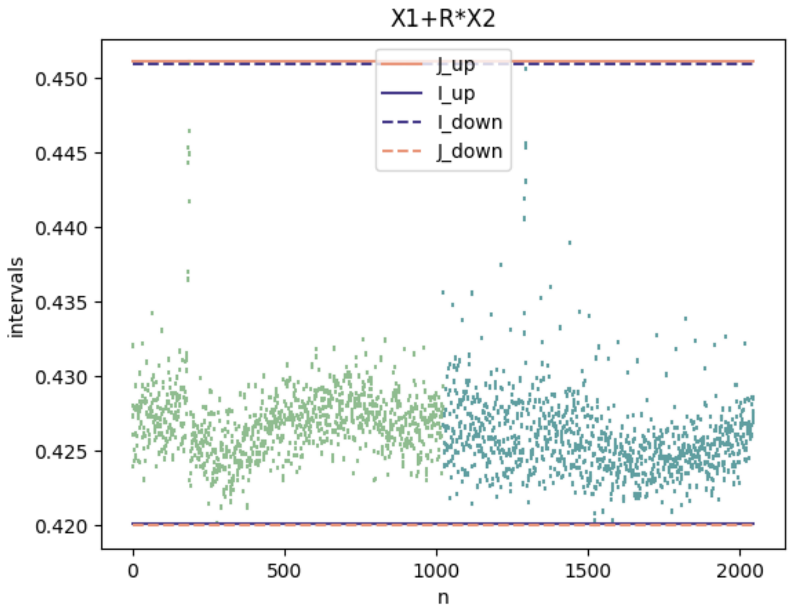

Рисунок 3. Объединенная выборка $X_1 \cup RX_2$ при оптимальном R

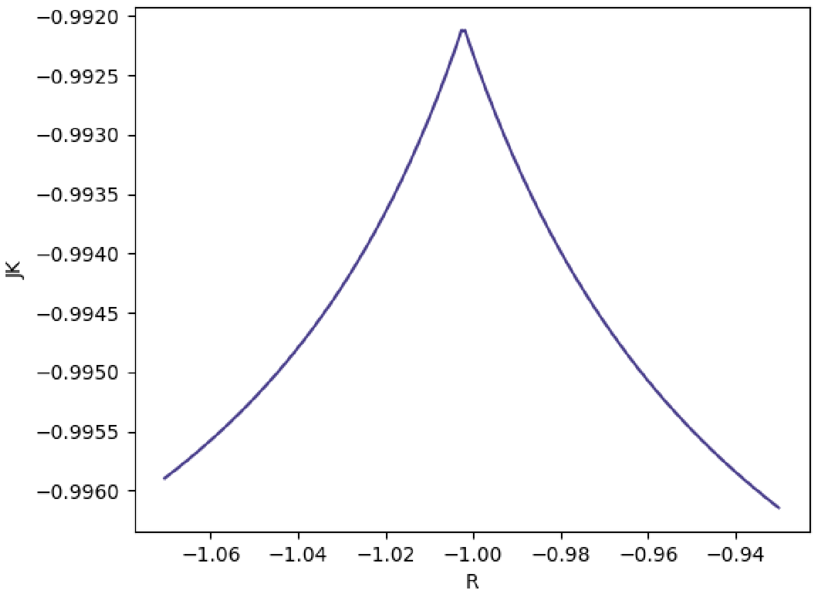

Рисунок 4. Зависимость меры Жаккара от коэффициента R

В результате получаем $R_{opt} = -1.00234$, при данных значениях мера Жаккара объединенной выборки максимальна и равна $JK = -0.99226$.

Значение меры Оскорбина объединенной выборки (рис.5) достигает минимального значения $k = 503.97184$ при $R = -1.00234$

Значение частоты объединенной выборки (рис.6) достигает максимального значения $\mu = 64$ при $R = -1.00234$

Оценка меры совместности: $[-1.00234, [-1.07103, -0.92952]]$.

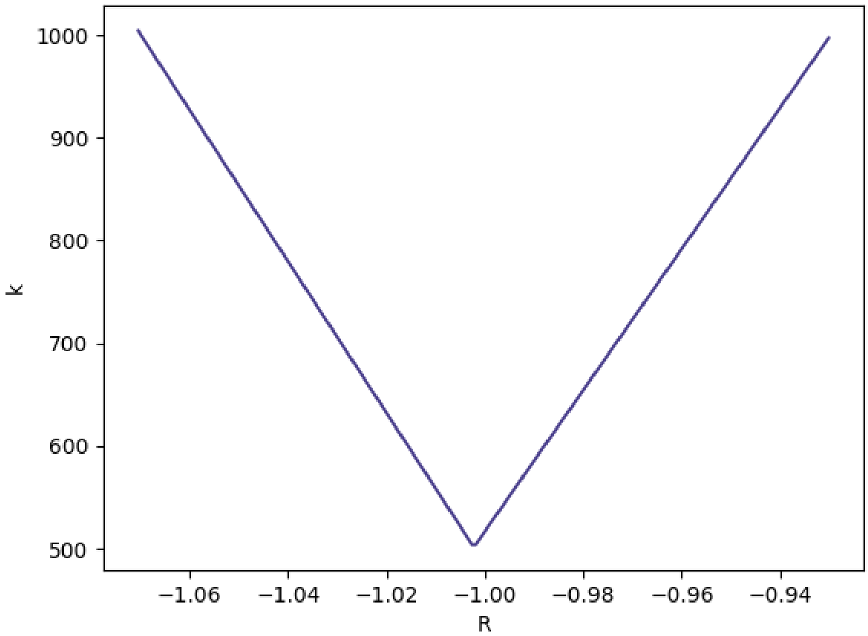

Рисунок 5. Зависимость меры Оскорбина от коэффициента R

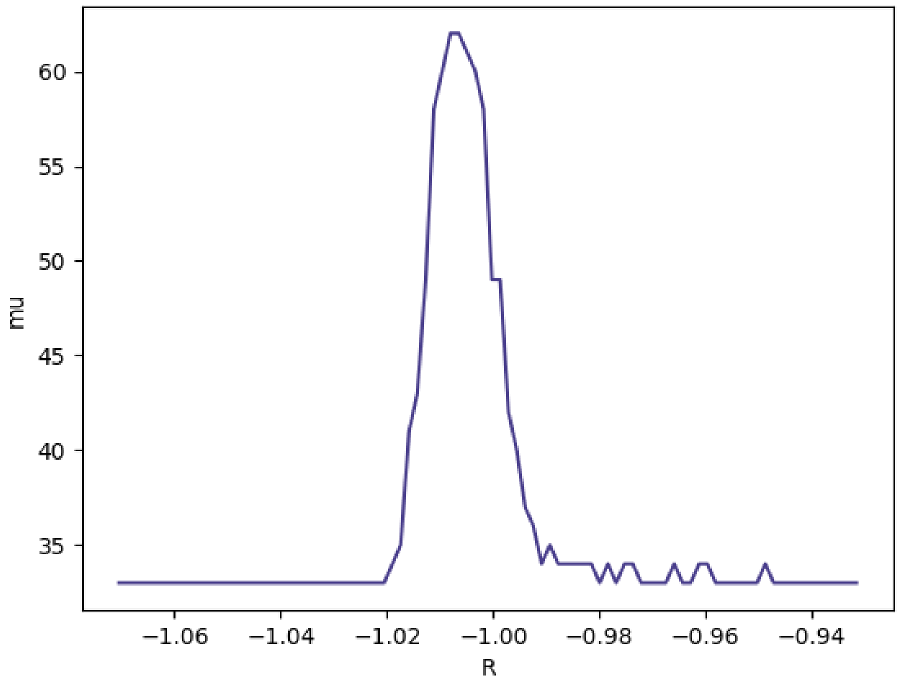

Рисунок 6. Зависимость частоты выборки от коэффициента R

Результаты для данных без коррекции

При отсутствии вычитания погрешности δ_i получаются следующие результаты.

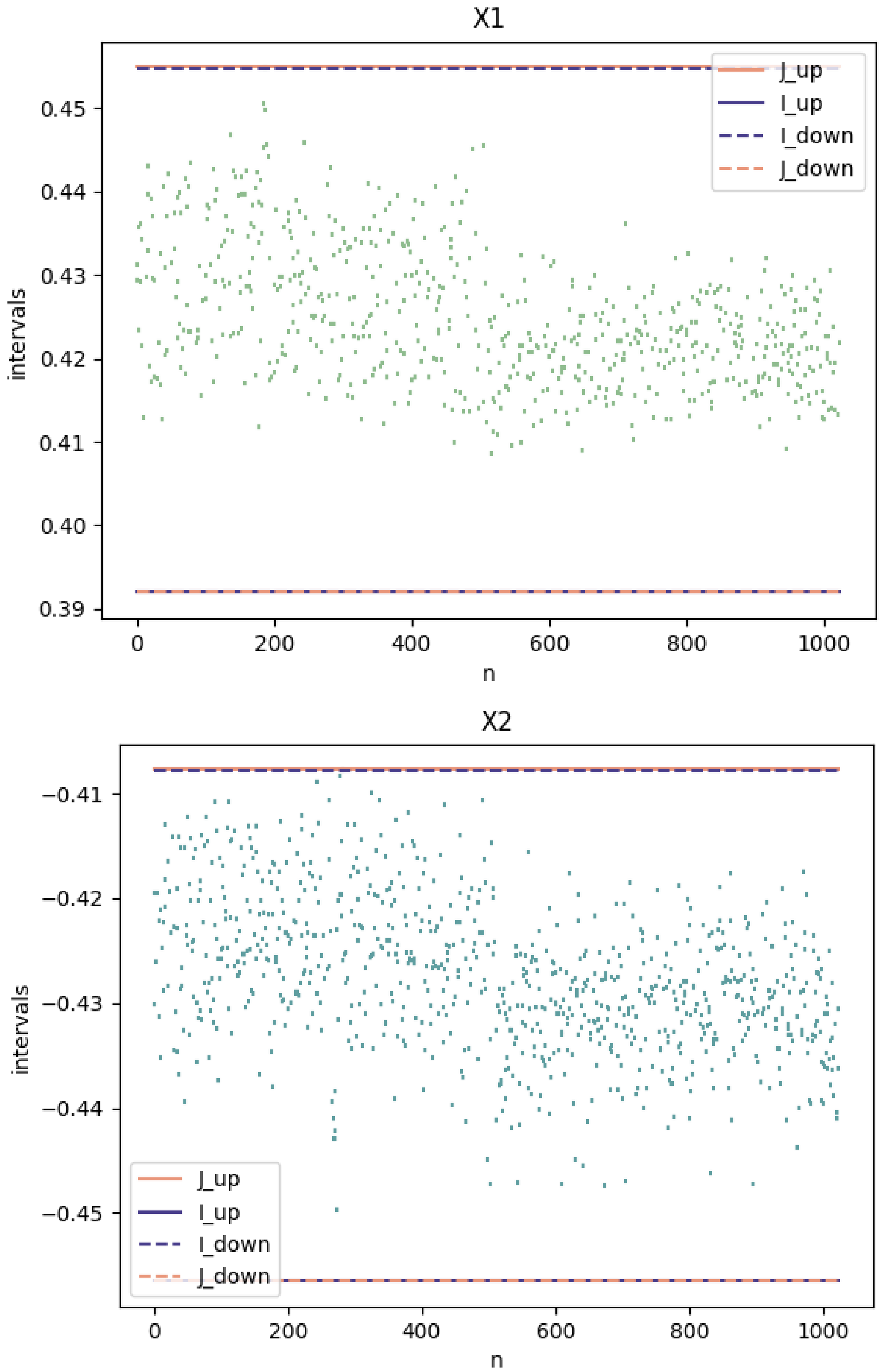

Рисунок 7. Интервальные выборки X_1 и X_2 (без коррекции) и их оценки

Таблица 3. Меры интервальных выборок (без коррекции)

	X_1	X_2
$[\underline{I}, \overline{I}]$	$[0.45471, 0.39209]$	$[-0.40784, -0.45642]$
$[\underline{J}, \overline{J}]$	$[0.39197, 0.45483]$	$[-0.45654, -0.40772]$
JK	-0.99618	-0.99508
k	1025.96608	795.93472
μ	17	15
$mode$	$[0.42255, 0.42261]$	$[-0.43103, -0.43097]$

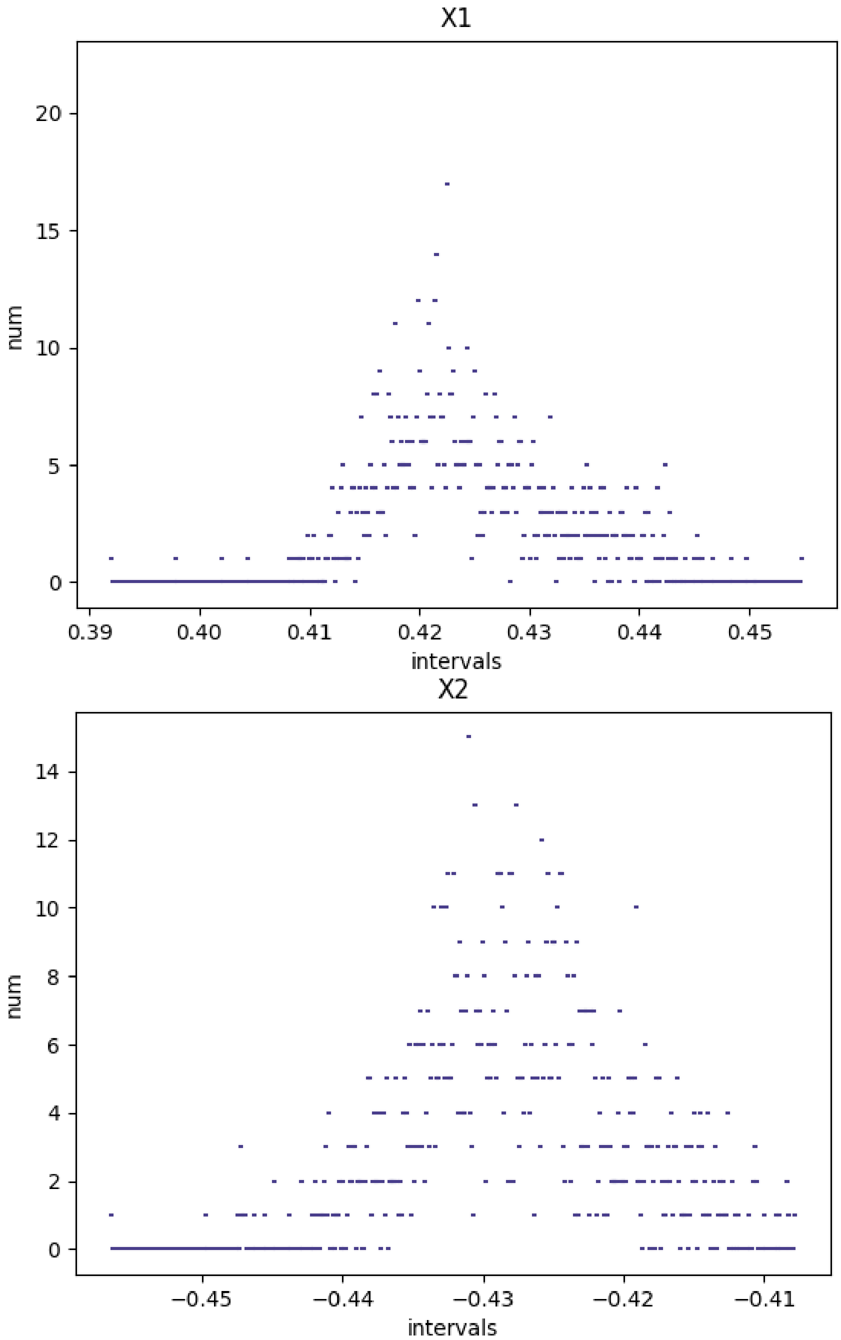

Рисунок 8. Частота выборок X_1 и X_2 (без коррекции)

Таблица 4. Оценки меры совместности выборок X_1 и X_2 (без коррекции)

R_{in}	$[-1.16407, -0.89692]$
R_{out}	$[-1.16473, -0.89642]$

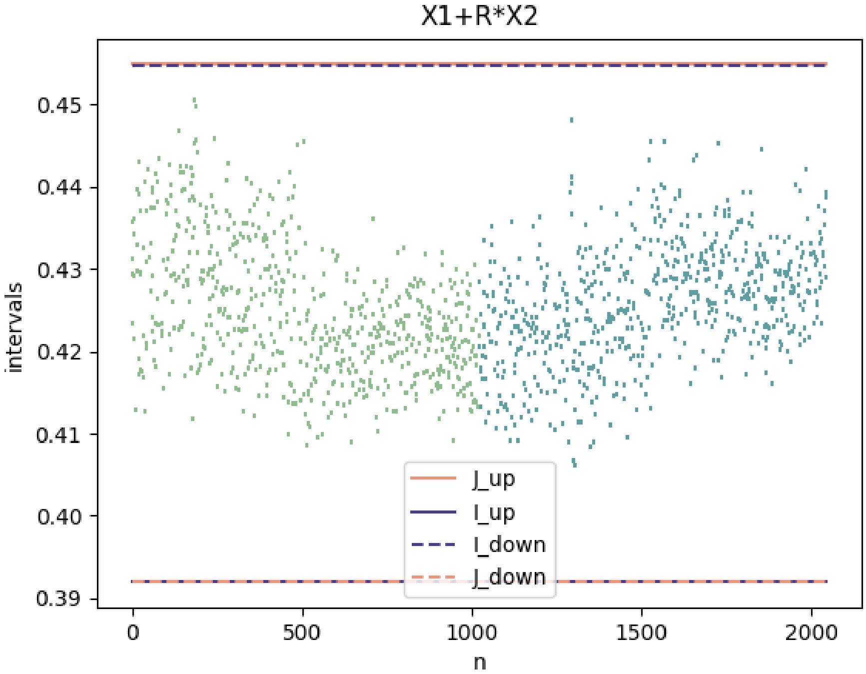

Рисунок 9. Объединенная выборка $X_1 \cup RX_2$ при оптимальном R (без коррекции)

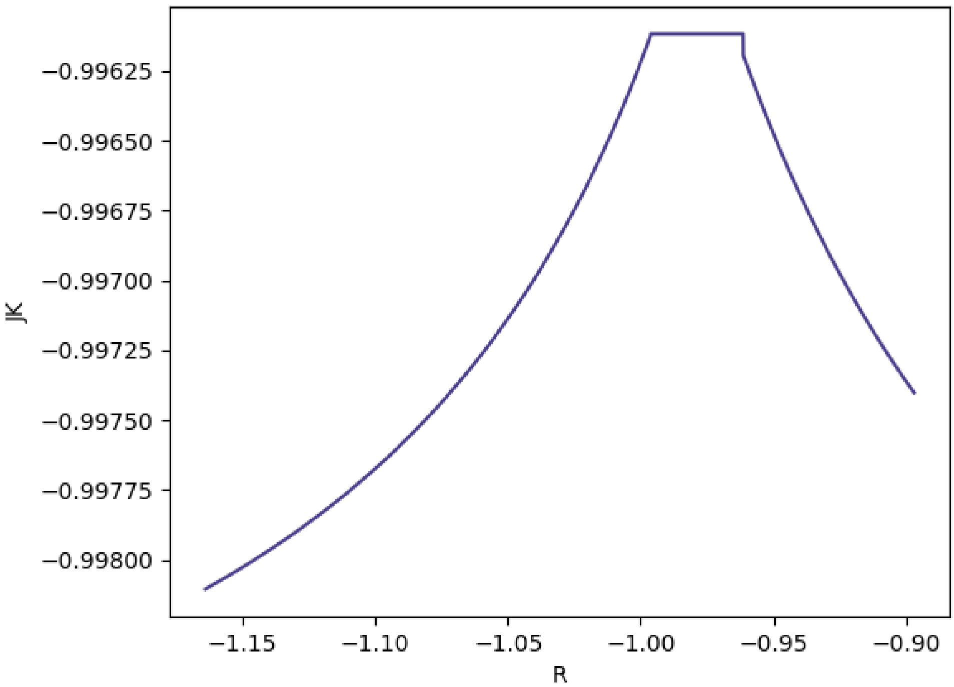

Рисунок 10. Зависимость меры Жаккара от коэффициента R (без коррекции)

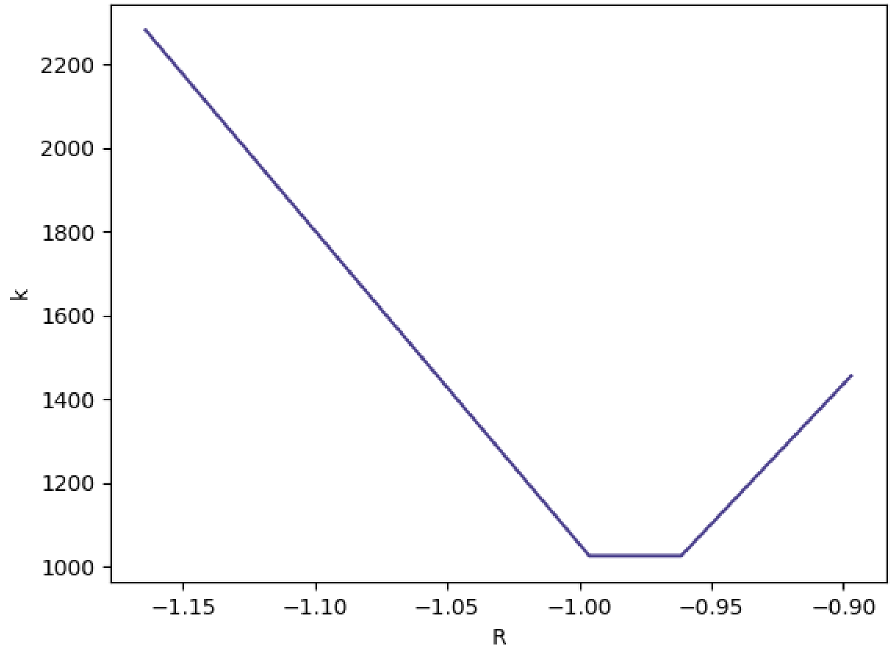

Рисунок 11.Зависимость меры Оскорбина от коэффициента R (без коррекции)

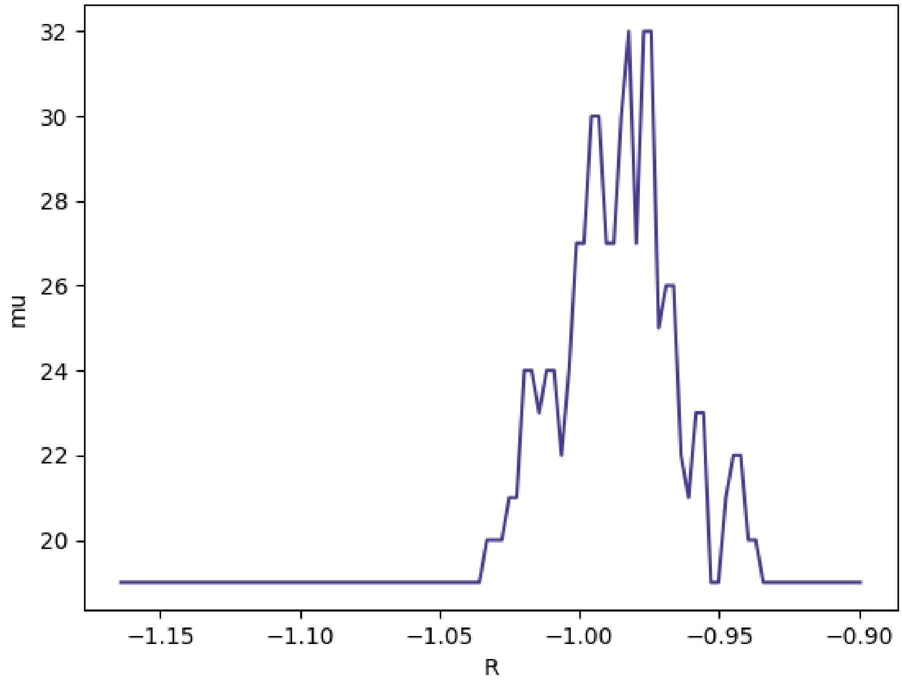

Рисунок 12. Зависимость частоты выборки от коэффициента R (без коррекции)

В результате получаем $R_{opt} = [-0.99617, -0.96144]$, при данных значениях мера Жаккара объединенной выборки максимальна и равна $JK = -0.99618$.

Значение меры Оскорбина объединенной выборки (рис.11) достигает минимального значения $k = 1025.96608$ при $R = [-0.99617, -0.96144]$.

Значение частоты объединенной выборки достигает максимального значения $\mu = 32$ при $R = -0.98234$ и на промежутке $[-0.97700, -0.97432\ ]$.

Оценка меры совместности: $[[-0.99617, -0.96144], [-1.16473, -0.89642]]$.

Анализ результатов

По графику на рис.1 и из данных в табл.1 видим, что обе исходные выборки несовместны: их внутренние оценки образуют неправильный интервал, мера Жаккара отрицательна, мера Оскорбина ненулевая, а частота выборки меньше числа элементов в ней. По графику частоты на рис.2 можно заметить постепенное увеличение частоты при приближении к значениям моды.

По графику на рис.4 явно определяется оптимальное значение меры совместности, достигаемое при максимальном значении меры Жаккара. По графикам на рис.5 и рис.6 можно заметить, что при оптимальном значении меры совместности мера Оскорбина и частота выборки также достигают своих экстремумов. Итоговая выборка остается несовместной, что можно понять по отрицательному значению меры Жаккара. Мера Жаккара объединенной выборки равна значению меры Жаккара первой выборки, аналогично с мерой Оскорбина. Частота объединенной выборки равна сумме частот исходных выборок.

Для выборок, построенных по данным без коррекции, наблюдаются похожие результаты. Выборки оказываются менее совместны, чем те, что построены по данным с коррекцией. Оптимальные значения меры совместности образуют интервал.