Санкт-Петербургский политехнический университет имени Петра Великого
Физико-Механический институт

Лабораторная работа №3

по дисциплине

Математическая статистика

Выполнил студент группы 5030102/00101 Маковеев Лев
Преподаватель Баженов Александр Николаевич

Санкт-Петербург, 2023

Содержание

Список иллюстраций

1 Постановка задачи

1. Для выборки данных $\mathbf{X_1}$ с интервальной неопределенностью ϵ привести диаграмму рассеяния с учетом неопределенности. Вычислить базовые оценки исходной выборки.

2. Оценить выборку с помощью набора мер совместности:

   (a) Размер максимальной клики μ_j

   (b) Величина коэффициента вариабельности по Оскорбину k_O

   (c) Мера совместности Жаккара J_i

2 Теория

2.1 Предварительные оценки

Имеется выборка данных $\mathbf{X_1}$ с интервальной неопределённостью. Для дальнейшей работы используется модель данных с уравновешенным интервалом погрешности:

$$\mathbf{x} = \overset{o}{x} + \epsilon, \quad \epsilon = [-10^{-3}, 10^{-3}]$$

$\overline{x}$ - множество верхних интервалов выборки, $\underline{x}$ - множество нижних интервалов выборки
Внешние оценки выборки:

$$\underline{\mathbf{J}} = \underset{1 \le k \le n}{min}\, \underline{x_k} \quad \overline{\mathbf{J}} = \underset{1 \le k \le n}{max}\, \overline{x_k}$$

2.2 Меры совместности выборок

2.2.1 Вычисление моды выборки и максимальной клики

Имеется выборка данных $\mathbf{X}$ с интервальной неопределённостью. Алгоритм для нахождения моды $mode\mathbf{X}$ и максимальной клики μ интервальной выборки:

1. $\mathbf{I} = \bigcap_{i=1}^{n} \mathbf{x_i}$

2. (a) Если $\mathbf{I} \neq \varnothing$
   $mode\mathbf{X} = \mathbf{I}$
   $\mu = n$

   (b) Если $\mathbf{I} = \varnothing$
   помещаем все концы интервалов $\overline{\mathbf{x_1}}, \underline{\mathbf{x_1}}, ..., \overline{\mathbf{x_n}}, \underline{\mathbf{x_n}}$ рассматриваемой выборки в один массив $Y = (y_1, y_2, ..., y_{2n})$. Упорядочиваем элементы в Y по возрастанию значений.
   Порождаем интервалы $z_i = [y_i, y_{i+1}], i = 1, 2, ..., 2n - 1$ (назовём их элементарными подинтервалами измерений); для каждого z_i подсчитываем число μ_i интервалов из выборки $\mathbf{X}$, включающих интервал z_i.
   Вычисляем $\mu = \underset{1 \le i \le 2n-1}{max} \mu_i$; выбираем номера k интервалов z_k, для которых μ_k равно максимальному, т. е. $\mu_k = \mu$, и формируем из таких k множество $K = \{k\} \subset \{1, 2, ..., 2n - 1\}$;
   $mode\mathbf{X} = \bigcup_{k \in K} z_k$

2.2.2 Оптимизация по Оскорбину

Поставим задачу линейного программирования в простейшем виде:

$$\min_{\omega,\beta} \omega$$

при ограничениях

$$\begin{cases} mid\mathbf{x_i} - \omega\epsilon \le \beta \le mid\mathbf{x_i} + \omega\epsilon, \\ \omega \ge 1, \end{cases} \qquad i = 1, 2, ..., n.$$

Решением данной задачи будут число ω на которое следует увеличить интервал неопределённости и число k, являющееся оценкой постоянной.

2.2.3 Индекс Жаккара

В качестве числовой характеристики степени совпадения двух интервалов $\mathbf{x}, \mathbf{y}$ рассмотрим величину

$$Ji(\mathbf{x}, \mathbf{y}) = \frac{wid(\mathbf{x} \wedge \mathbf{y})}{wid(\mathbf{x} \vee \mathbf{y})}$$

Рассмотренная мера обобщает обычное понятие меры совместности на различные типы взаимной совместности интервалов. Если пересечение интервалов $\mathbf{x}, \mathbf{y}$ пусто, т. е. $\mathbf{x} \cap \mathbf{y} = \varnothing$, то $\mathbf{x} \wedge \mathbf{y}$ — неправильный интервал и числитель формулы имеет отрицательное значение.

В предельном случае несовпадающих вещественных вырожденных интервалов $\mathbf{x} = x$ и $\mathbf{y} = y$, $x \neq y$, имеем

$$Ji(\mathbf{x}, \mathbf{y}) = -1$$

В целом получаем

$$-1 \le Ji(\mathbf{x}, \mathbf{y}) \le 1$$

Мера совместности, введённая для двух интервалов допускает естественное обобщение на случай интервальной выборки. Определим меру $Ji(\mathbf{x}, \mathbf{y})$ для выборки как

$$Ji(\mathbf{x}, \mathbf{y}) = \frac{wid(\bigwedge_\mathbf{i} \mathbf{x_i})}{wid(\bigvee_\mathbf{i} \mathbf{x_i})}$$

В связи несовместностью выборки будем использовать следующую меру, которая имеет место и в случае несовместных выборок.

$$\rho(mode\mathbf{X}) = \frac{wid(mode\mathbf{X})}{wid(\bigwedge_\mathbf{i} \mathbf{x_i})}$$

Назовём данную конструкцию относительная ширина моды. В отличие от минимума по включению, мода выборки всегда является правильным интервалом. В целом получаем

$$0 \le \rho(mode\mathbf{X}) \le 1$$

3 Результаты

В качестве выборки брались данные из файла «Канал 1_600nm_0.03.csv»

3.1 Предварительные оценки

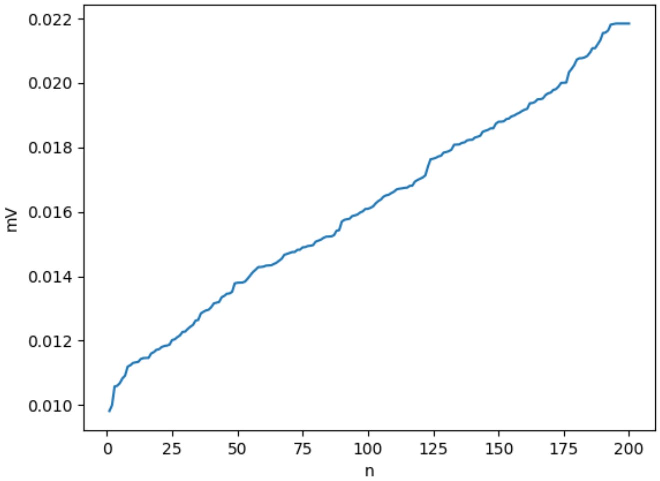

Рис. 1: Данные выборки $\mathbf{X_1}$

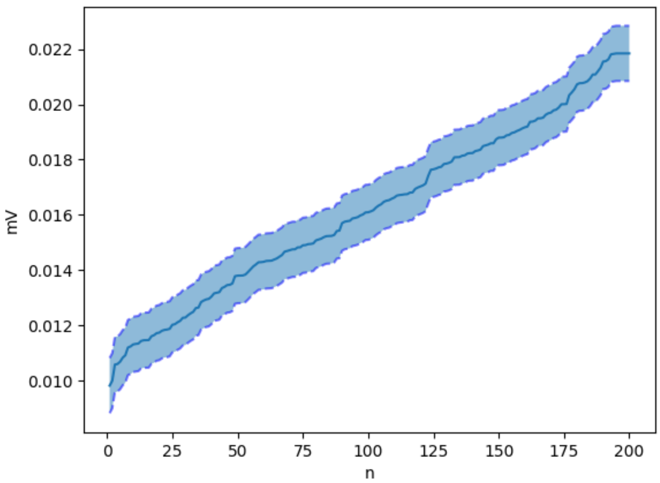

Рис. 2: Данные выборки $\mathbf{X_1}$ с уравновешенным интервалом погрешности

Вычисление внешних оценок дают следующие результаты:

$$\underline{\mathbf{J}} = 0.009 \quad \overline{\mathbf{J}} = 0.023$$

Верхние и нижние вершины оценок $\mathbf{J}$ совпадают с границами отображения на рис. 2.

3.2 Меры совместности выборок

3.2.1 Вычисление моды выборки и максимальной клики

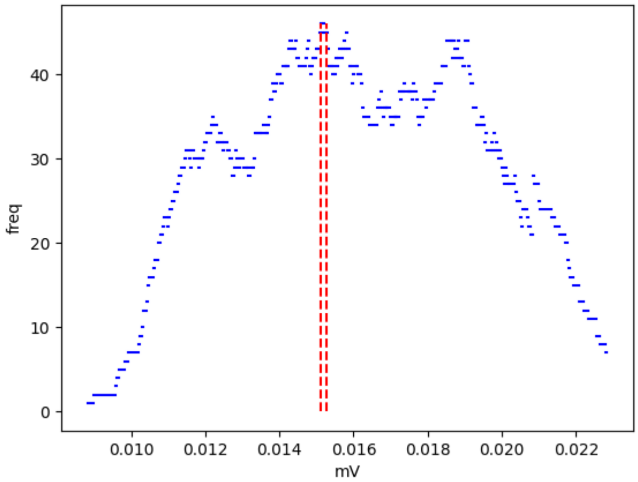

Рис. 3: График частот при вычислении моды выборки$\mathbf{X_1}$

По результатам вычисления моды выборки находим размер максимальной клики:

$$max\ \mu_i(\mathbf{X_1}) = 46$$

Индексы таких элементов образуют множество K, а из них образуется мода

$$mode\ \mathbf{X_1} = [0.015125, 0.015132] \cup [0.015168, 0.015203] \cup [0.015264, 0.015282]$$

На рис. 4 показаны элементы выборки $\mathbf{X_1}$, в которые входит мода.

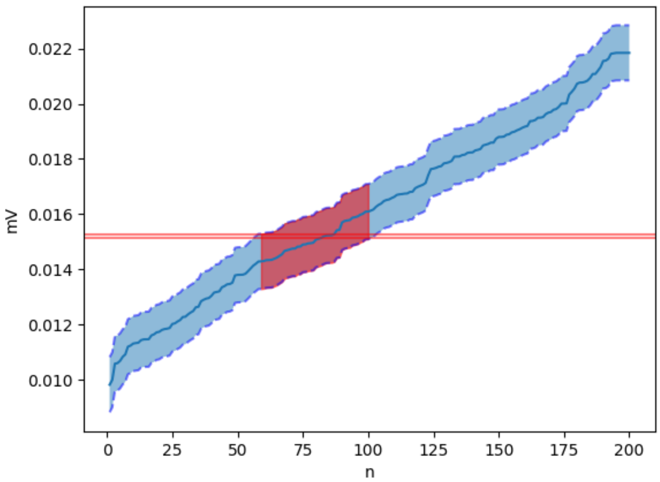

Рис. 4: Элементы выборки $\mathbf{X_1}$, в которые входит мода

3.2.2 Оптимизация по Оскорбину

Вычисления дают следующие результаты:

$$oskorbin\_center\_k = 0.158$$
$$\omega = 6.01$$

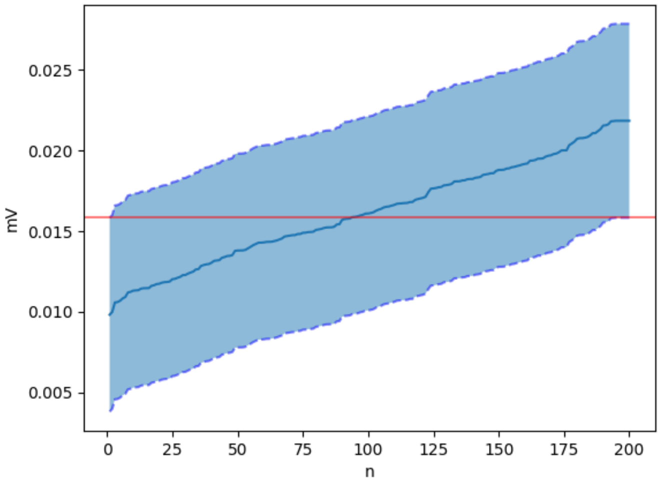

Рис. 5: Диаграмма рассеяния выборки $\mathbf{X_1}$ с увеличенным в ω раз интервалом неопределённости.

3.2.3 Индекс Жаккара и Относительная ширина моды

Вычисления дают следующие результаты:

$$Ji(\mathbf{X_1}) = -0.715$$

$$\rho(mode\mathbf{X_1}) = 0.00466$$

4 Обсуждение

4.1 Мода выборки и максимальная клика

По вычисленным значениям максимальной клики видно, что она более чем в 4 раза меньше размера выборки, что говорит о большой степени несовместности данной выборки.

4.2 Оптимизация по Оскорбину

Полученная оценка постоянной достаточно близка к полученной ранее моде. Величина однородного расширения интервалов очень велика, что соответствует весьма большой степени несовместности выборки.

4.3 Индекс Жаккара

Отрицательность меры Жаккара соответствует несовместности выборки, а её модуль — высокой степени этой несовместности.

5 Реализация

Лабораторная работа выполнена с использованием языка программирования Python 3.10 в среде разработки PyCharm Community с использованием библиотек matplotlib, numpy, scipy.

6 Дополнительно

Код лабораторной работы:
https://github.com/S0krat/MakoveevLev_mathstat2023/tree/main/Lab3