

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Высшая школа прикладной математики и вычислительной физики

Отчёт по лабораторной работе №3

по дисциплине

«Математическая статистика»

Выполнила студентка
группы 5030102/90201

Кожевникова Диана Геннадьевна

Проверил
Доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург
2022

СОДЕРЖАНИЕ

Список иллюстраций	3
1 Постановка задачи	4
2 Теория	5
2.1 Представление данных	5
2.2 Линейная регрессия	5
2.2.1 Модель линейной регрессии	5
2.2.2 Метод наименьших модулей	5
2.3 Предварительная обработка данных	6
2.4 Коэффициент Жаккара	6
2.5 Процедура оптимизации	7
3 Результаты	7
4 Обсуждение	10
4.1 Модель дрейфа	10
4.2 Гистограммы скорректированных данных и объединенной выборки	10
4.3 Коэффициент Жаккара. Оптимальное значение коэффициента калибровки.	10
5 Реализация	10
6 Приложение	10
6.1 Список литературы	10
6.2 Ссылка на репозиторий	11

Список иллюстраций

1	Схема установки для исследования фотоэлектрических характеристик.	4
2	Исходные данные.	7
3	Исходные данные в интервальном представлении.	7
4	Линейная модель дрейфа данных.	8
5	Гистограммы значений множителей коррекции w	8
6	Модели данных после коррекции.	8
7	Гистограммы скорректированных данных.	8
8	Коэффициент Жаккара от калибровочного множителя.	9
9	Гистограмма объединенной выборки.	9

1 Постановка задачи

Исследование из области солнечной энергетики [1]. На рис. 1 приведена схема установки для исследования фотоэлектрических характеристик.

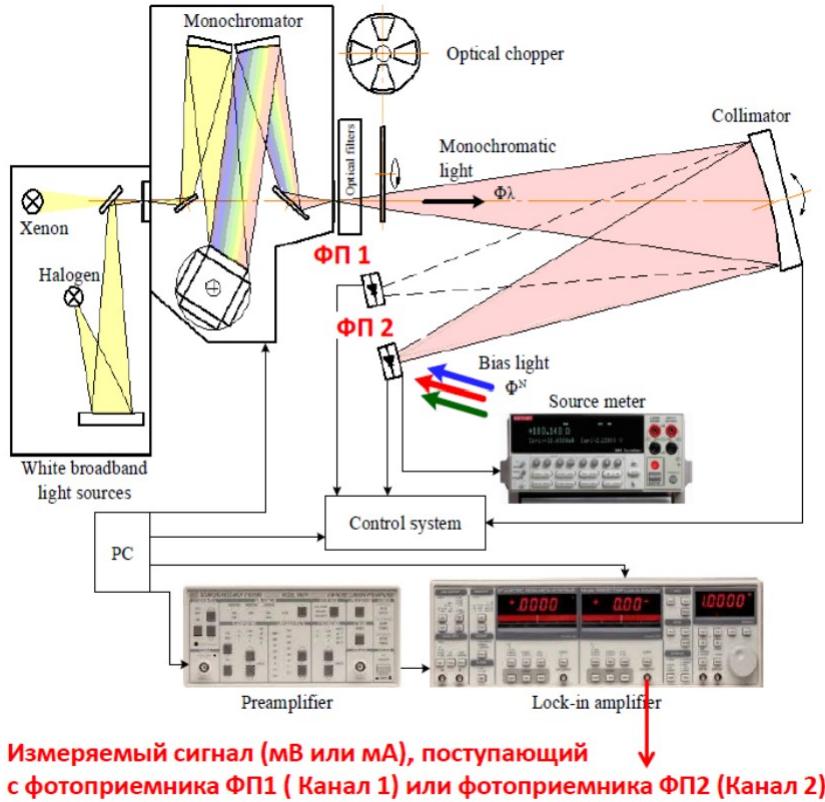


Рис. 1: Схема установки для исследования фотоэлектрических характеристик.

Калибровка датчика ФП1 производится по эталону ФП2. Зависимость между квантовыми эффективностями датчиков предполагается одинаковой для каждой пары измерений

$$QE_{\Phi\text{P}2} = \frac{I_{\Phi\text{P}2}}{I_{\Phi\text{P}1}} \cdot QE_{\Phi\text{P}1} \quad (1)$$

QE - квантовые эффективности, I - измеренные токи.

Исходные данные. Дано две выборки данных с интервальнойной неопределенностью. Одна из них относится к эталонному датчику ФП2, другая - к исследуемому датчику ФП1.

Задача. Необходимо определить коэффициент калибровки

$$R_{21} = \frac{I_2}{I_1} \quad (2)$$

методом линейной регрессии на множестве интервальных данных и коэффициента Жаккара.

2 Теория

2.1 Представление данных

Необходимо представить полученные данные так, чтобы применить к ним понятие статистики данных с интервальнойной неопределенностью.

Один из способов получения интервальных результатов в первичных измерениях – прибавка к точечному базовому значению x_0 , полученному из показаний измерительного прибора, некоторого интервала погрешности ϵ .

$$\mathbf{x} = \dot{x} + \epsilon \quad (3)$$

Пусть интервал погрешности:

$$\epsilon = [-\epsilon; \epsilon]$$

В данном случае принимаем ϵ равным 10^{-4} мВ.

Согласно терминологии интервального анализа, рассматриваемая выборка есть вектор интервалов, или интервальный вектор $x = (x_1, x_2, \dots, x_n)$.

2.2 Линейная регрессия

2.2.1 Модель линейной регрессии

Линейная регрессия – регрессионная модель зависимости одной переменной от другой с линейной функцией зависимости:

$$y_i = X_i b_i + \epsilon_i$$

где X – заданные значения, y – параметры отклика, ϵ – случайная ошибка модели. В случае, если у нас y_i зависит от одного параметра x_i , то модель выглядит следующим образом:

$$y_i = b_0 + b_1 \cdot x_i + \epsilon_i \quad (4)$$

В данной модели пренебрегаем погрешностью и считаем, что она получена при измерении y_i .

2.2.2 Метод наименьших модулей

Для наиболее точного приближения входных данных y_i линейной регрессией $f(x_i)$ используется метод наименьших модулей. Данный метод основан на минимизации нормы разности последовательности:

$$\|f(x_i) - y_i\|_{l^1} \rightarrow \min \quad (5)$$

Здесь ставится задача линейного программирования (ЗЛП), решение которой дает нам коэффициенты b_0 и b_1 , а также вектор множителей коррекции данных w . Получаем следующую

ЗЛП:

$$\sum_{i=1}^n |w_i| \rightarrow \min \quad (6)$$

$$b_0 + b_1 \cdot x_i - w_i \cdot \epsilon \leq y_i, i = \overline{1, n} \quad (7)$$

$$b_0 + b_1 \cdot x_i + w_i \cdot \epsilon \leq y_i, i = \overline{1, n} \quad (8)$$

$$1 \leq w_i, i = \overline{1, n} \quad (9)$$

2.3 Предварительная обработка данных

Для оценки постоянной, как можно будет увидеть далее, необходима предварительная обработка данных. Займемся линейной моделью дрейфа.

$$Lin(n) = A + B \cdot n, n = \overline{1, N} \quad (10)$$

Поставив и решив задачу линейного программирования, найдем коэффициенты A , B , вектор w множителей коррекции данных для каждого из фотоприемников ФП1 и ФП2.

Для данных от первого фотоприемника ФП1: $A = 4.365712$, $B = 4.3657 \cdot 10^{-3}$.

Для данных от второго фотоприемника ФП2: $A = 4.668742$, $B = 4.6687 \cdot 10^{-3}$.

В последствии множитель коррекции данных необходимо применить к погрешностям выборки для получения данные, согласующихся с линейной моделью дрейфа:

$$I^f(n) = \dot{x}(n) + \epsilon \cdot w(n), n = \overline{1, N} \quad (11)$$

По итогу необходимо построить "спрямленные" данные выборки: получить их можно путем вычитания из исходных данных линейной компоненты:

$$I^c(n) = I^f(n) - B \cdot n, n = \overline{1, N} \quad (12)$$

2.4 Коэффициент Жаккара

Коэффициент Жаккара – это мера сходства множеств. В интервальных данных рассматривается некоторая модификация этого коэффициента, где в качестве меры множества (в данном случае интервала) рассматривается его длина, а в качестве пересечения и объединения – взятие минимума и максимума по включению двух величин в интервальной арифметике Каухера соответственно. Можно заметить, что в силу возможности минимума по включению быть неправильным интервалом, коэффициент Жаккара может достигать значения только в интервале $[-1; 1]$.

$$JK(x) = \frac{(\bigcap_{i=1}^n x_i)}{(\bigcup_{i=1}^n x_i)} \quad (13)$$

2.5 Процедура оптимизации

Чтоб найти оптимальный параметр калибровки R_{21} необходимо поставить и решить задачу максимизации коэффициента Жаккара, зависящего от параметра калибровки:

$$JK(I_1^c(n) \cdot R \cup I_2^c(n)) = \rightarrow \max \quad (14)$$

где I_1^c и I_2^c - полученные спрямленные выборки, а R - параметр калибровки. Найденный таким образом R будет искомым оптимальным R_{21} в силу наибольшего совпадения, оцененного коэффициентом Жаккара.

3 Результаты

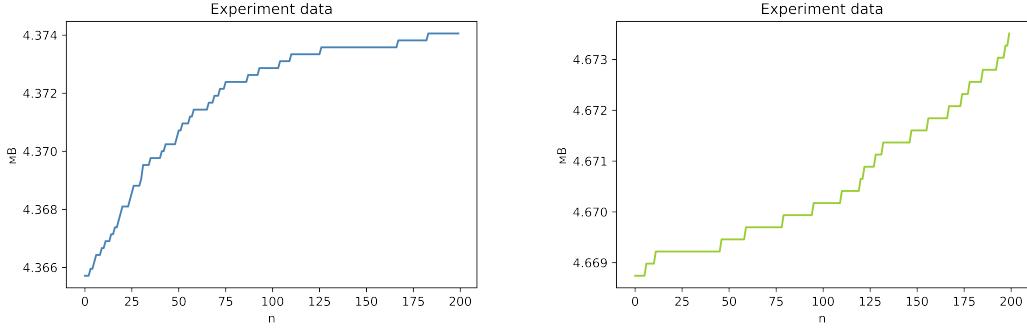


Рис. 2: Исходные данные.

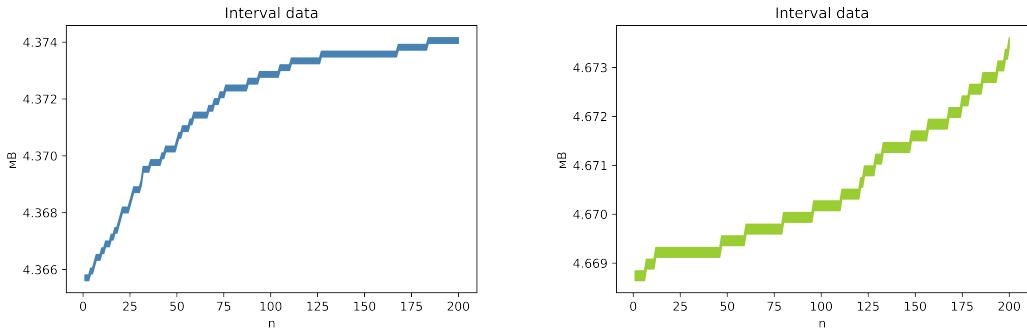


Рис. 3: Исходные данные в интервальном представлении.

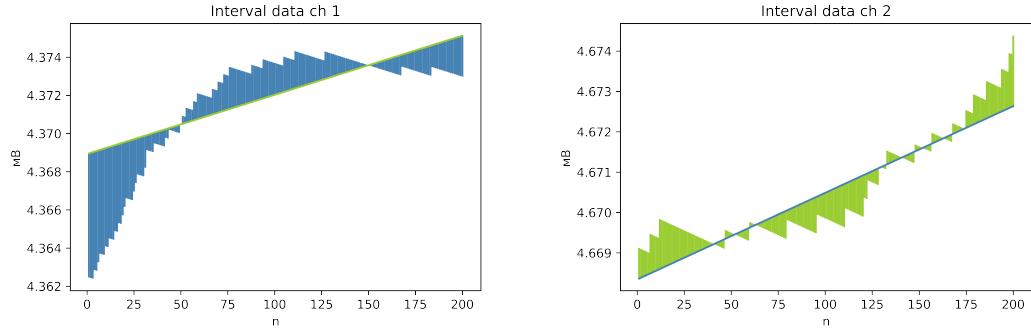


Рис. 4: Линейная модель дрейфа данных.

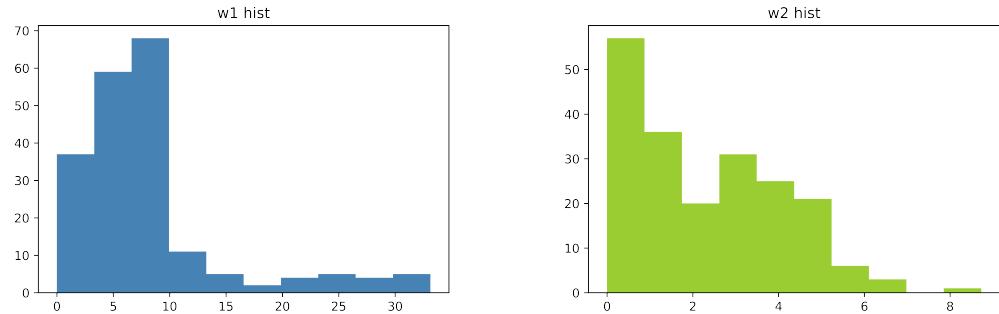


Рис. 5: Гистограммы значений множителей коррекции w .

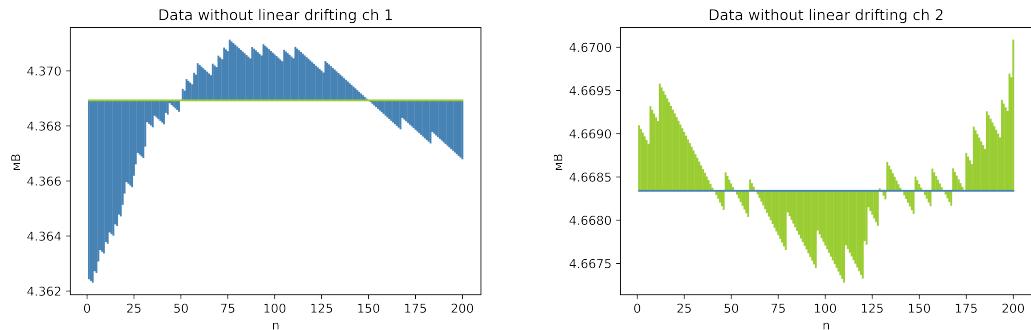


Рис. 6: Модели данных после коррекции.

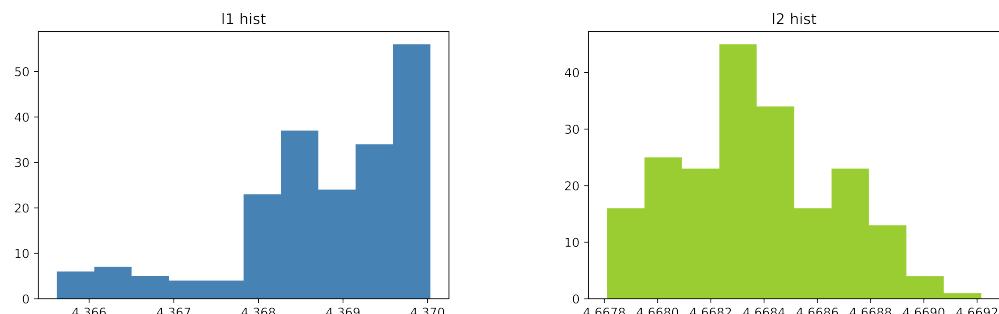


Рис. 7: Гистограммы скорректированных данных.

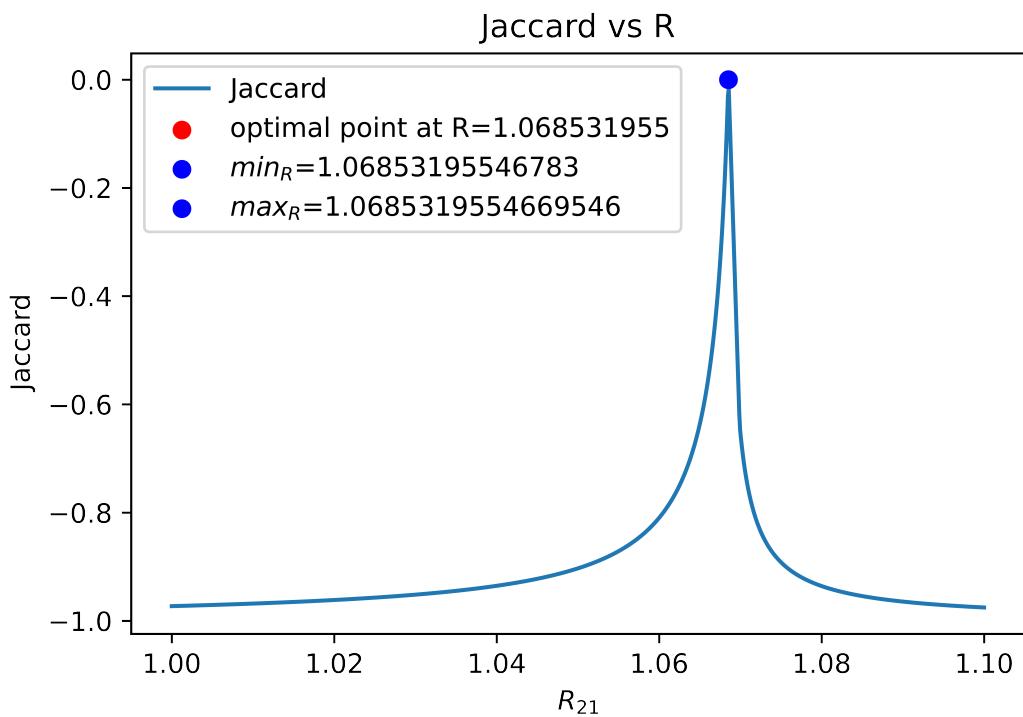


Рис. 8: Коэффициент Жаккара от калибровочного множителя.

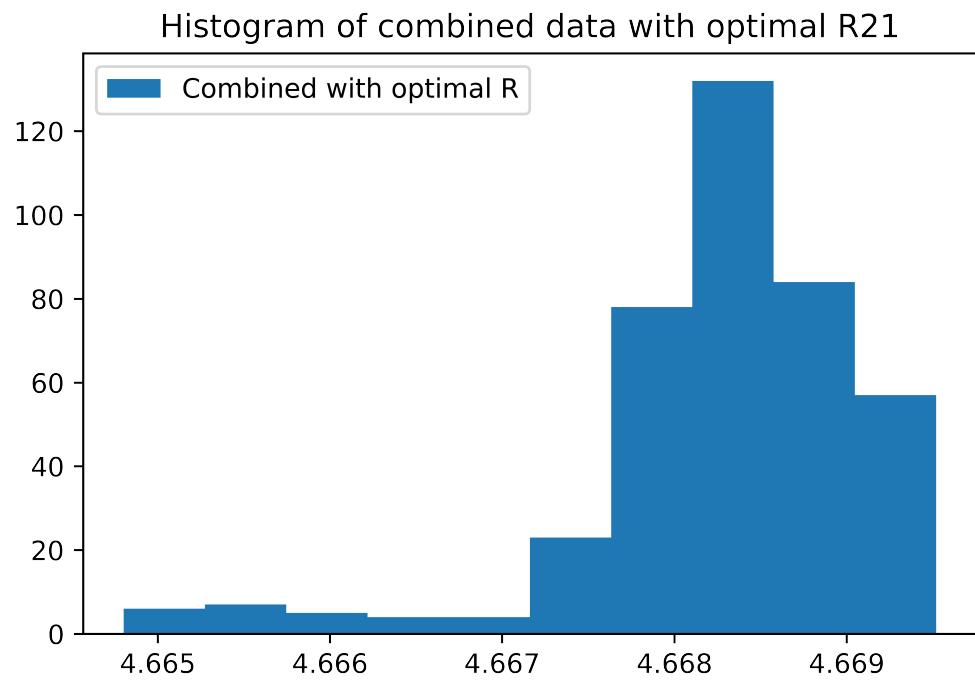


Рис. 9: Гистограмма объединенной выборки.

4 Обсуждение

4.1 Модель дрейфа

Из гистограмм w_1 , w_2 видно, что примерно половина данных требуют коррекции, величины коэффициентов коррекции достаточно малы, что позволяет сделать вывод о том, что линейная модель дрейфа достаточно хорошо описывает данные.

4.2 Гистограммы скорректированных данных и объединенной выборки

Гистограмма скорректированной выборки имеет пик в районе 4.668-4.669, он же характерен для гистограммы совмещенной выборки. Также видно, что границы совмещенной выборки совпадают с границами для эталонной (второй) выборки.

4.3 Коэффициент Жаккара. Оптимальное значение коэффициента калибровки.

Полученное при помощи коэффициента Жаккара оптимальное значение коэффициента калибровки $R_{21} = 1.068531955$, видно, что интервал на котором данный коэффициент положительный очень мал (порядка 10^{-12}). Из этого можно сделать вывод, что данные имеют неточности, которые сложно устранимы.

5 Реализация

Лабораторная работа выполнена на языке Python версии 3.8.10 в среде разработки Jupyter Notebook . Использовались дополнительные библиотеки:

1. scipy
2. numpy
3. matplotlib
4. pandas

6 Приложение

6.1 Список литературы

1. А.Н. Баженов. Введение в анализ данных с интервальной неопределенностью.
2. М.З.Шварц. Данные технологических испытаний оборудования для калибровки фотоприемников солнечного излучения. 2022.

6.2 Ссылка на репозиторий

Код программы GitHub: <https://github.com/diakozhh/MathStat>