Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Кафедра "Прикладная математика"

Отчет по лабораторной работе №9
по дисциплине
"Математическая статистика"

Выполнил студент:
Кротиков Сергей Ильич
группа: 5030102/90101
Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург, 2022

СОДЕРЖАНИЕ

СПИСОК ИЛЛЮСТРАЦИЙ

1 Постановка задачи

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

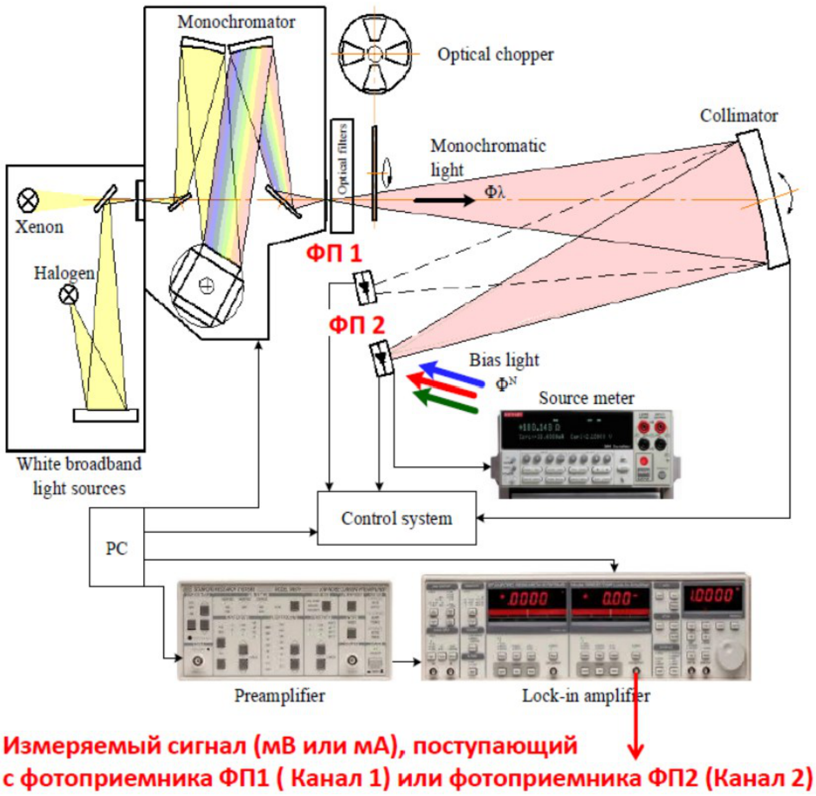

Рис. 1: Схема установки для исследования фотоэлектрических характеристик.

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

$$QE_{ФП2} = \frac{I_{ФП2}}{I_{ФП1}} * QE_{ФП1} \qquad (1)$$

QE - квантовые эффективности эталонного и исследуемого датчиков, I - измеренные токи.

Исходные данные. Имеется 2 выборки данных с интервальной неопределенностью. Одна из них относится к эталонному датчику ФП2, другая - к исследуемому датчику ФП1.

Задача. Треубется определить коэффициент калибровки

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

при помощи линейной регрессии на множестве интервальных данных и коэффициента Жаккара.

2 Теория

В первую очередь прдставим данные таким образом, чтобы применить понятия статистики данных с интервальной неопределенностью.
Один из распространённых способов получения интервальных результатов в первичных измерениях - это "обинтерваливание"точечных значений, когда к точечному базовому зачению x_0, которое считывается по показаниям измерительного прибора, прибавляется *интервал погрешности* ϵ:

$$\mathbf{x} = \dot{x} + \epsilon \tag{3}$$

Интервал погрешности зададим как

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

В конкретных измерениях примем $\epsilon = 10^{-4}$ мВ.
Согласно терминологии интервального анализа, рассматриваемая выборка - это вектор интервалов, или интервальный вектор $x = (x_1, x_2, ..., x_n)$.

2.1 Линейная регрессия

2.1.1 Описание модели

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:

$$y_i = b_0 + b_1 * x_i + \epsilon_i \tag{4}$$

В данной можели мы пренебрегаем прогрешностью и считаем, что она получается при измерении y_i.

2.1.2 Метод наименьших модулей

Для наиболее точного приближения входных с фотоприемников данных y_i линейной регрессией $f(x_i)$ используется метод наименьших модулей. Этот метот основывается на минимизации нормы разности последовательности:

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

В данном случае ставится задача линейного программирования, решение которой дает нам коэффициенты b_0 и b_1, а также вектор множителей коррекции данных w. По итогу получается следующая задача линейного программирования

$$\sum_{i=1}^{n} |w_i| \to min \tag{6}$$

$$b_0 + b_1 * x_i - w_i * \epsilon \leq y_i, i = 1..n \tag{7}$$

$$b_0 + b_1 * x_i + w_i * \epsilon \leq y_i, i = 1..n \tag{8}$$

$$1 \leq w_i, i = 1..n \tag{9}$$

2.2 Предварительная обработка данных

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

$$Lin(n) = A + B * n, n = 1, 2, ...N \tag{10}$$

Поставив и решив задачу линейного программирования, найдем коэффициенты A, B и вектор w множителей коррекции данных для каждого из фотоприемников ФП1 и ФП2: для данных с первого фотоприемника A = 4.74835, B = $9.17308 * 10^{-6}$, а для данных со второго - A = 5.18171, B = $1.10476 * 10^{-5}$. В последствии множитель коррекции данных необходимо применить к погрешностям выборки, чтобы получить данные, которые согласовывались с линейной моделью дрейфа:

$$I^f(n) = \dot{x}(n) + \epsilon * w(n), n = 1, 2, ...N \tag{11}$$

В итоге необходимо построить "спрямленные"данные выборки: получить их можно путем вычитания из исходных данных линейную компоненту:

$$I^c(n) = I^f(n) - B * n, n = 1, 2, ...N \tag{12}$$

2.3 Коэффициент Жаккара

Коэффициент Жаккара - мера сходства множеств. В интервальных данных рассматривается некоторая модификация этого коэффициента: 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].

$$JK(x) = \frac{wid(\wedge x_i)}{wid(\vee x_i)} \tag{13}$$

2.4 Процедура оптимизации

Чтоб найти оптимальный параметр калибровки $R_2 1$ необходимо поставить и решить задачу максимизации коэффициента Жаккара, зависящего от параметра калибровки:

$$JK(I_1^c(n) * R \cup I_2^c(n)) = \rightarrow max \tag{14}$$

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

3 Реализация

Лабораторная работа была реализована при помощи языка программирования Python 3.9 с использованием библиотек NumPy, MatPlotLib и SciPy. Работа выполнена в Jupyter Notebook и GNU Octave. Отчет выполнен в редакторе LaTex TeXstudio.

4 Результаты

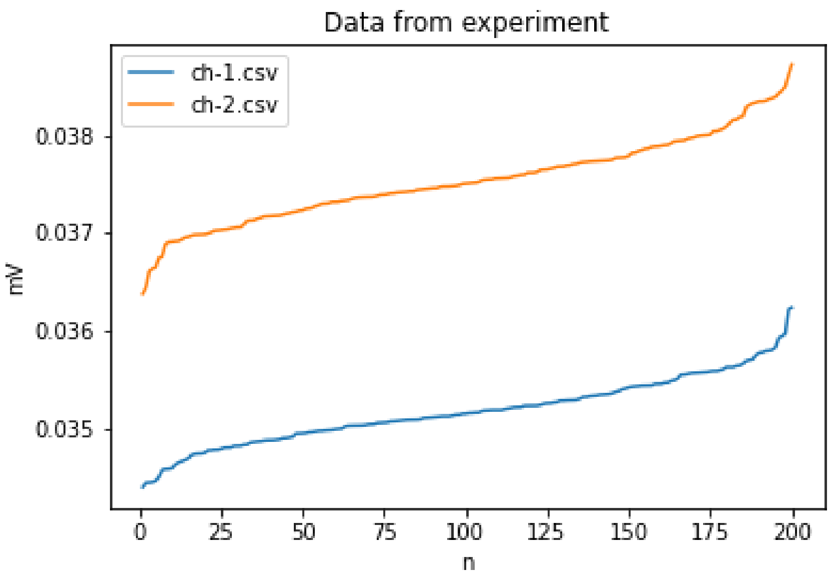

Рис. 2: Исходные данные из экспериментов

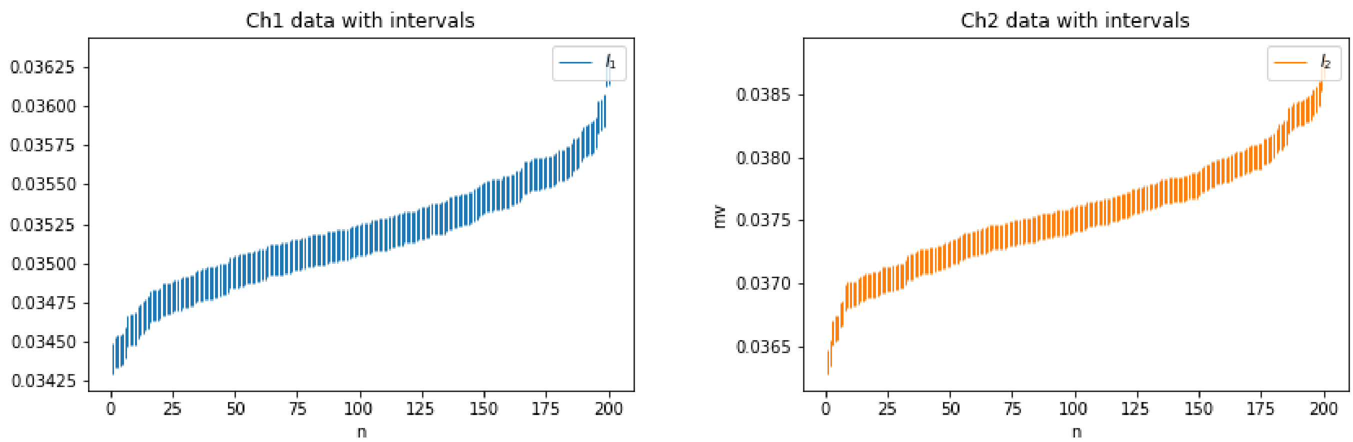

Рис. 3: Интервальное представление исходных данных

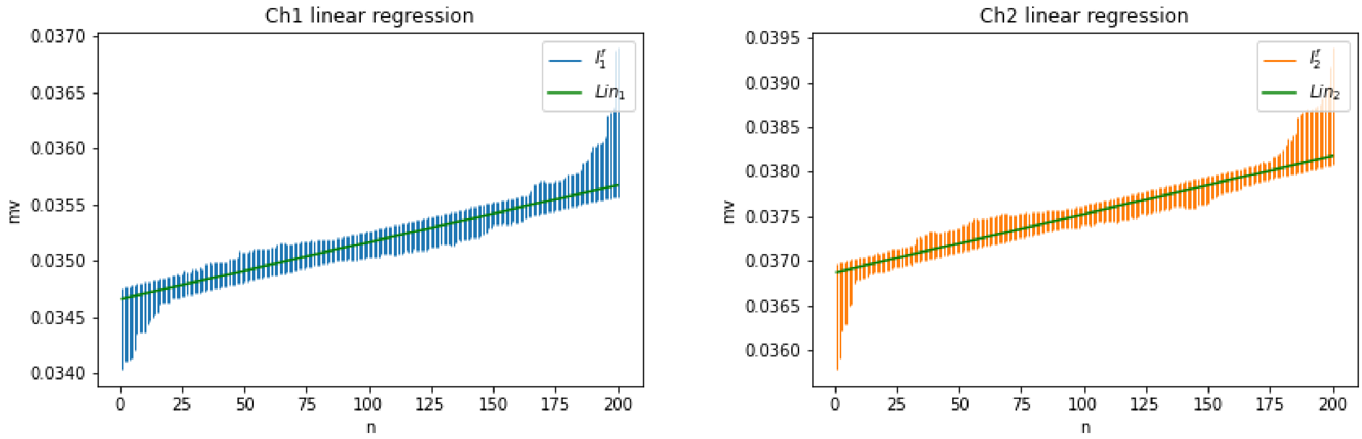

Рис. 4: Линейная модель дрейфа данных

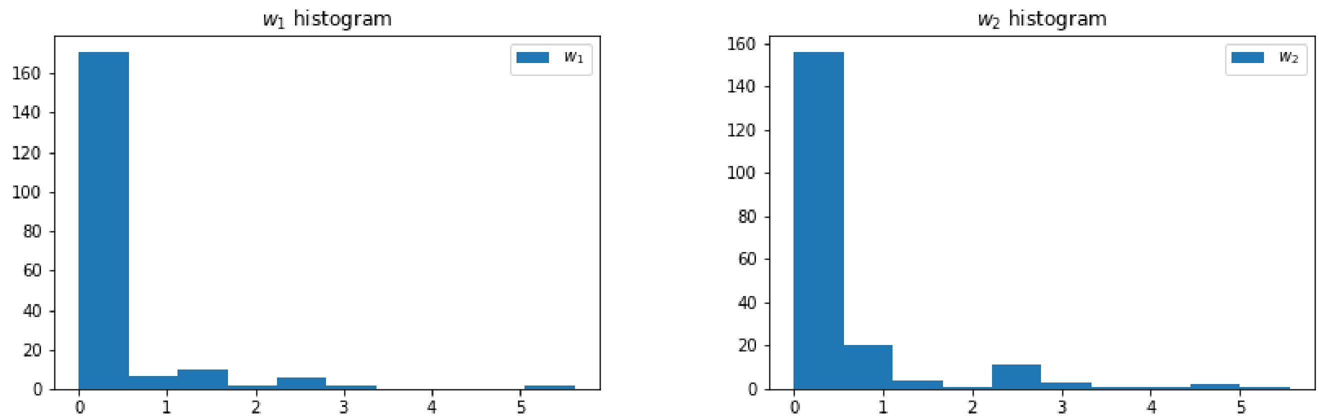

Рис. 5: Гистограммы значений множителей коррекции w

Результаты линейного приближения токов.

- Для первого фотоприемника:

$$A_1 = 0.0346557, \quad B_1 = 5.09353 \cdot 10^{-6}$$

- Для второго фотоприемника:

$$A_2 = 0.0368672, \quad B_2 = 6.54667 \cdot 10^{-6}$$

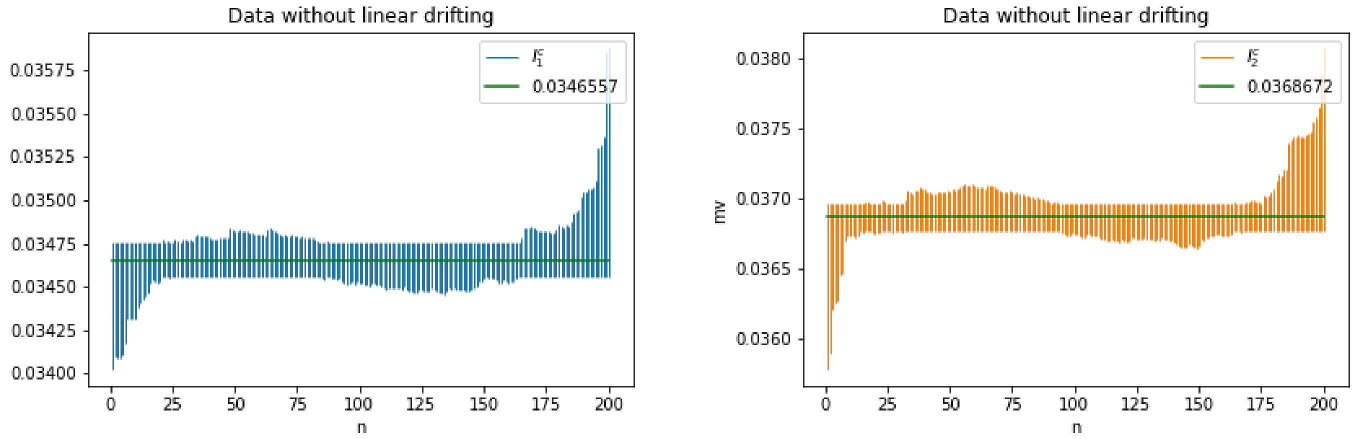

Рис. 6: Скорректированные модели данных

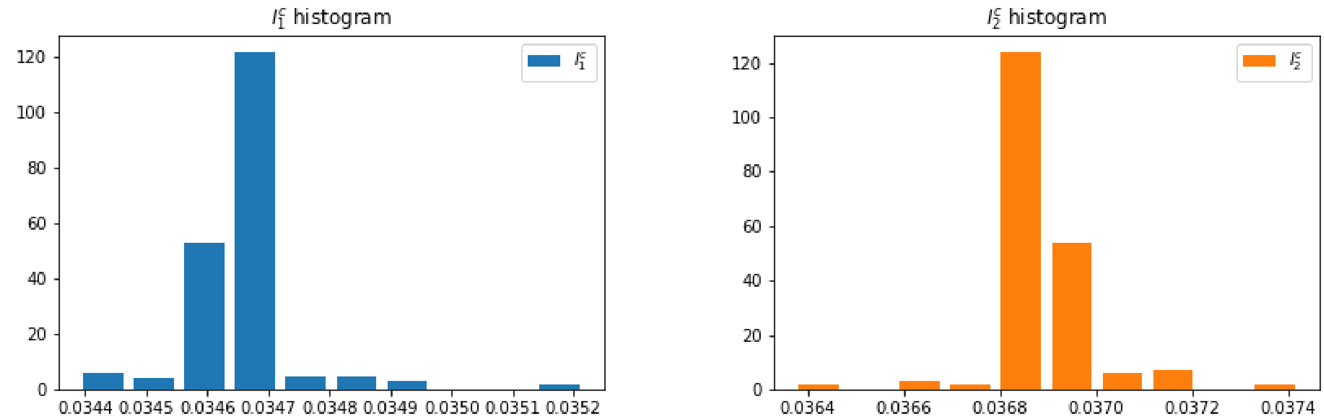

Рис. 7: Гистограммы скорректированных данных

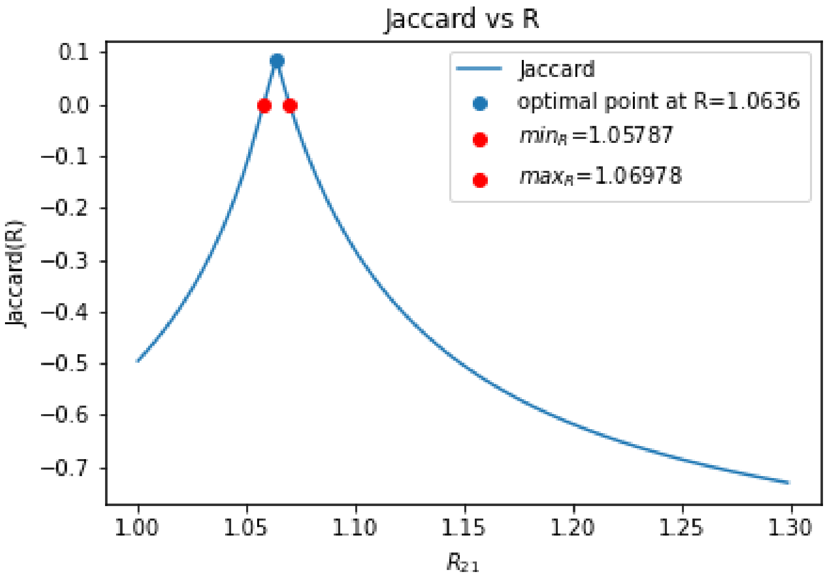

Рис. 8: Значение коэффициента Жаккара от калибровочного множителя от R_{21}

Результаты исследования:

$$R_{opt} = 1.063625, \quad jaccard(R) = 0.083612$$

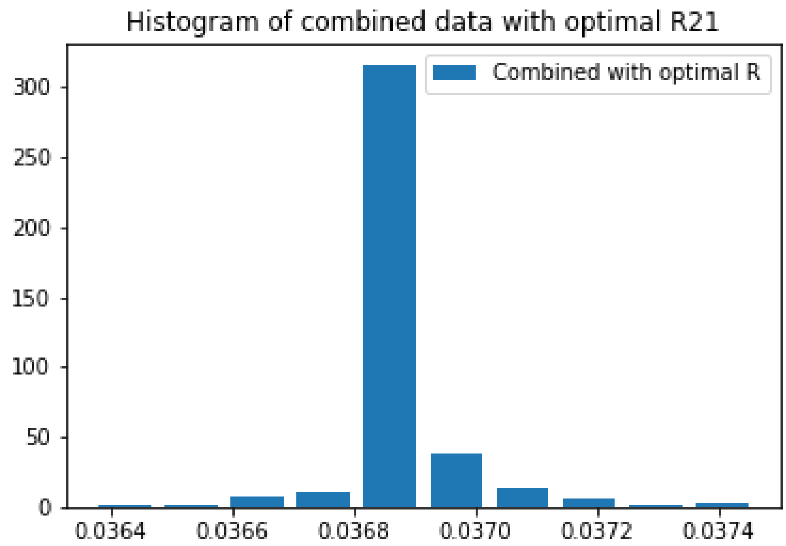

Рис. 9: Гистограмма объединённых данных при оптимальном значении R_{21}

5 Список литературы

1. М.З.Шварц. Данные технологических испытаний оборудования для калибровки фотоприемников солнечного излучения. 2022.

6 Приложение

`https://github.com/Krotikov/matStat` - GitHub репозиторий