Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
**Высшая школа прикладной математики и вычислительной физики**

# Отчёт по лабораторной работе №9

по дисциплине
«Математическая статистика»

Выполнил студент
группы 5030102/90101 Лаэтин Андрей Алексеевич

Проверил
Доцент, к.ф.-м.н. Баженов Александр Николаевич

Санкт-Петербург
2022

# СОДЕРЖАНИЕ

**1 Постановка задачи** . . . . . . . . . . 3
1.1 Исходные данные . . . . . . . . . . 3
1.2 Задача . . . . . . . . . . 4

**2 Теория** . . . . . . . . . . 4
2.1 Представление данных . . . . . . . . . . 4
2.2 Линейная регрессия . . . . . . . . . . 4
2.2.1 Описание модели . . . . . . . . . . 4
2.2.2 Метод наименьших модулей . . . . . . . . . . 5
2.3 Предварительная обработка данных . . . . . . . . . . 5
2.4 Коэффициент Жаккара . . . . . . . . . . 6
2.5 Процедура оптимизации . . . . . . . . . . 6

**3 Реализация** . . . . . . . . . . 6

**4 Результаты** . . . . . . . . . . 7

**5 Приложение** . . . . . . . . . . 10

# 1 Постановка задачи

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

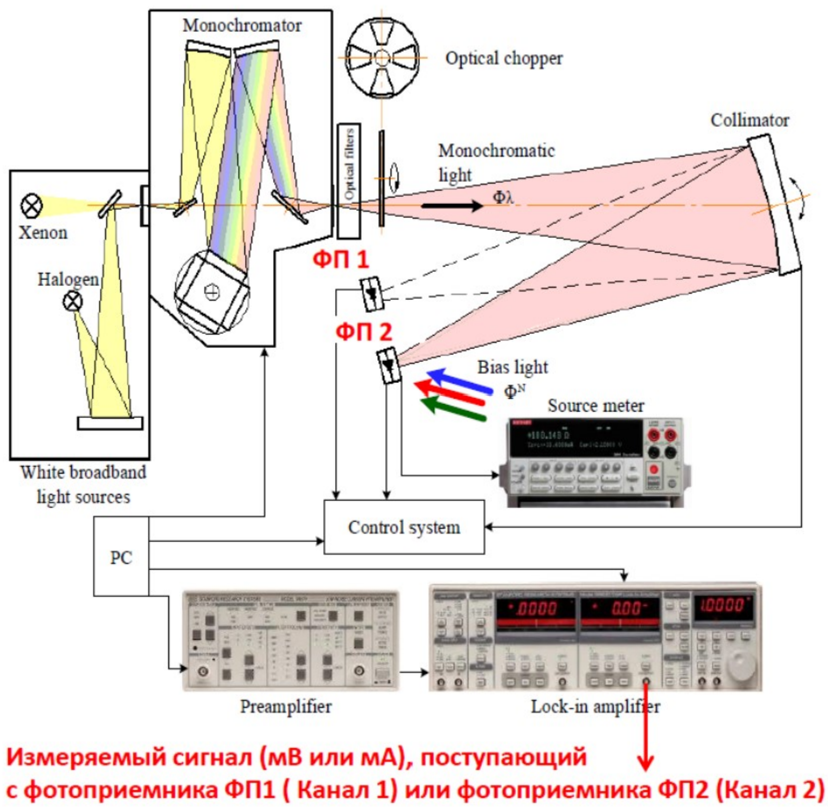



Рис. 1: Схема установки для исследования фотоэлектрических характеристик.

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

$$QE_{\text{ФП2}} = \frac{I_{\text{ФП2}}}{I_{\text{ФП1}}} * QE_{\text{ФП1}} \tag{1}$$

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

## 1.1 Исходные данные

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

### 1.2 Задача

Треубется определить коэффициент калибровки

$$R_{21} = \frac{I_2}{I_1} \tag{2}$$

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

# 2 Теория

## 2.1 Представление данных

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* $\varepsilon$:

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

Интервал погрешности зададим как

$$\varepsilon = [-\varepsilon; \varepsilon]$$

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

## 2.2 Линейная регрессия

### 2.2.1 Описание модели

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

$$y_i = X_i b_i + \varepsilon_i$$

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

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

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

### 2.2.2 Метод наименьших модулей

Для наиболее точного приближения входных с фотоприемников данных $y_i$ линейной регрессией $f(x_i)$ используется метод наименьших квадратов. 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

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

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

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

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

## 2.3 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. В последствии множитель коррекции данных необходимо применить к погрешностям выборки, чтобы получить данные, которые согласовывались с линейной моделью дрейфа:

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

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

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

### 2.4 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 оъединения - взятие минимума и максимума в интервальной арифметике соответственно. Можно заметить, что в силу возможности минимума по включению быть неправильным инервалом, коэффициент Жаккара может достишать значения только в интервале [-1; 1].

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

### 2.5 Процедура оптимизации

Чтоб найти оптимальный параметр калиброфки $R_21$ необходимо поставить и решить задачу максимизации коэффициента Жаккара, зависящего от парамертра калибровки:

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

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 коэффицентом Жаккара.

## 3 Реализация

Лабораторная работа была реализована при помощи языка программирования Python 3.9 с использованием библиотек NumPy, MatPlotLib и SciPy.

# 4 Результаты

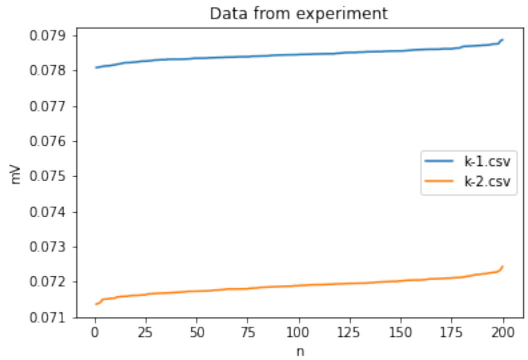



Рис. 2: Исходные данные из экспериментов

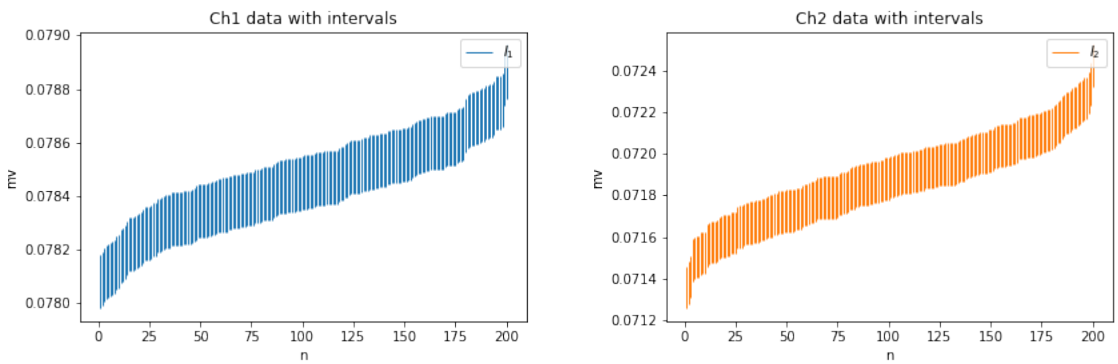



Рис. 3: Интервальное представление исходных данных

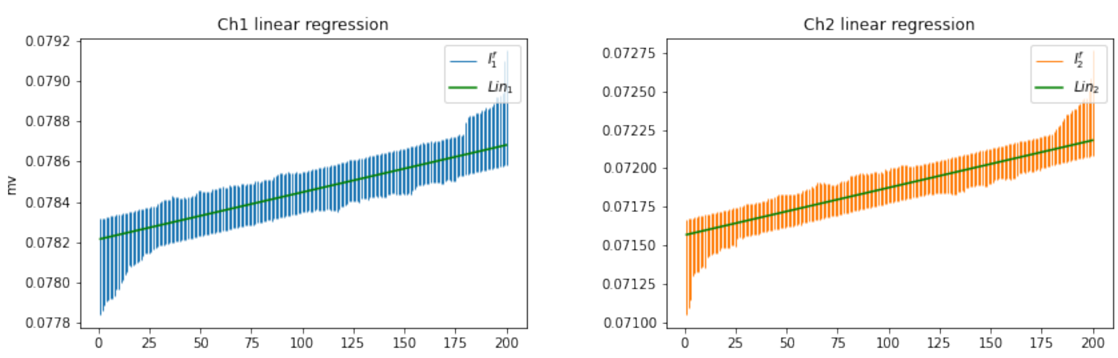



Рис. 4: Линейная модель дрейфа данных

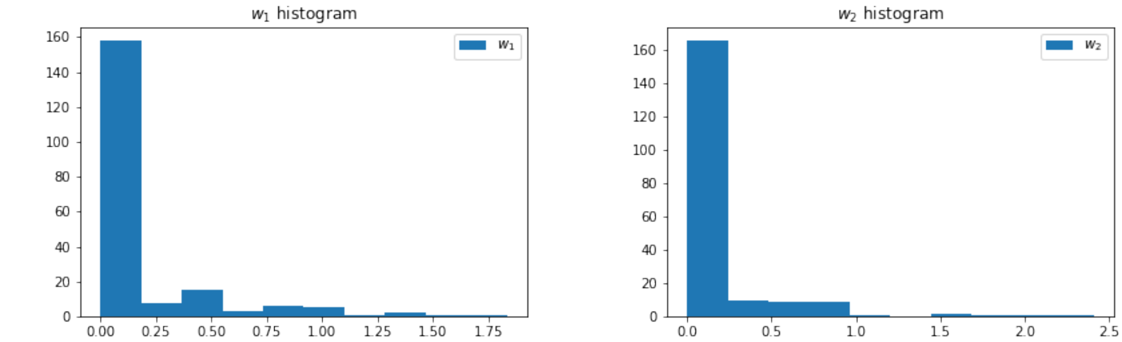



Рис. 5: Гистограммы значений множителей коррекции w

Результаты линейного приближения токов.

- Для первого фотоприёмника:

$$A_1 = 0.0782143,\ B_1 = 2.33735 \cdot 10^{-6}$$

- Для второго фотоприёмника:

$$A_2 = 0.0715656,\ B_2 = 3.08633 \cdot 10^{-6}$$

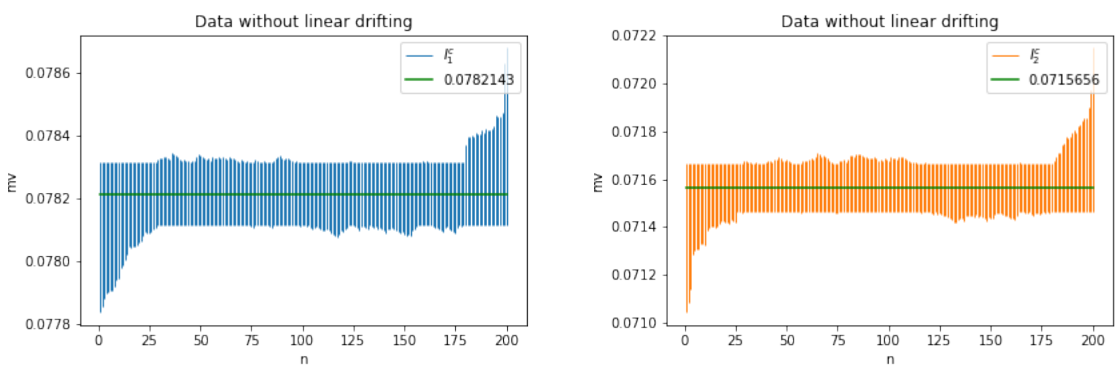
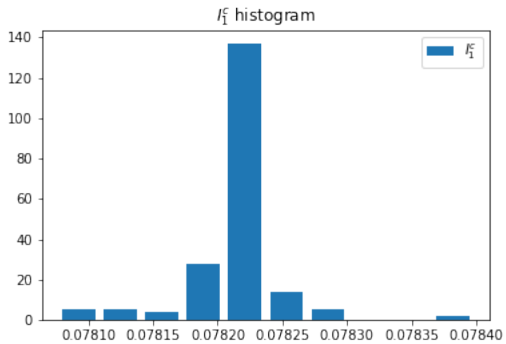



Рис. 6: Скорректированные модели данных



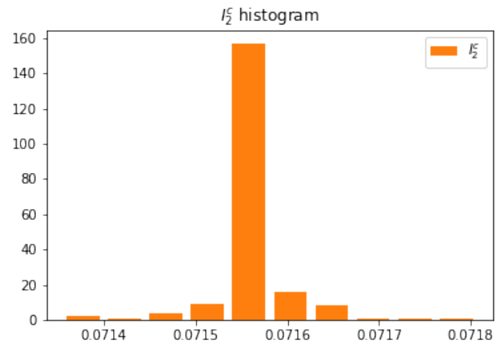




Рис. 7: Гистограммы скорректированных данных

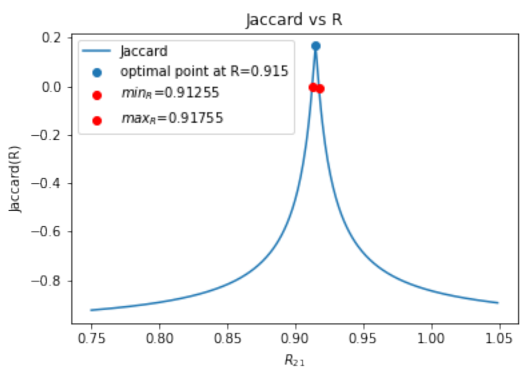



Рис. 8: Значение коэффициента Жаккара от калибровочного множителя от R21

Результаты исследования:

$$Ropt = 0.915, jaccard(R) = 0.165814$$

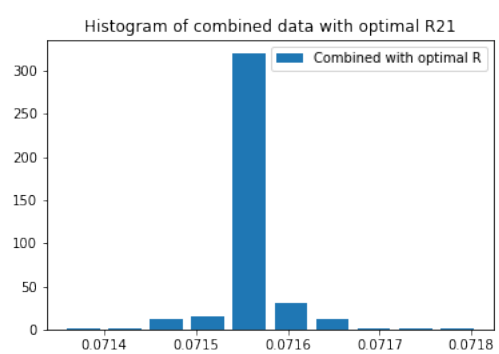



Рис. 9: Гистограмма объединённых данных при оптимальном значении R21

# 5 Приложение

Код программы GitHub URL:
https://github.com/A21l63/math-prob-stat