

Санкт-Петербургский политехнический университет  
Петра Великого

Физико-механический институт  
Кафедра «Прикладная математика»

ЛАБОРАТОРНАЯ РАБОТА: 9  
ПО ДИСЦИПЛИНЕ  
«МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Выполнил студент  
Келарев Михаил Алексеевич  
группы 5030102/90101

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

Санкт-Петербург  
2022

# Содержание

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Постановка задачи</b>                   | <b>2</b>  |
| <b>2</b> | <b>Теория</b>                              | <b>3</b>  |
| 2.1      | Представление данных . . . . .             | 3         |
| 2.2      | Линейная регрессия . . . . .               | 3         |
| 2.2.1    | Описание модели . . . . .                  | 3         |
| 2.2.2    | Метод наименьших модулей . . . . .         | 3         |
| 2.3      | Предварительная обработка данных . . . . . | 4         |
| 2.4      | Коэффициент Жаккара . . . . .              | 4         |
| 2.5      | Процедура оптимизации . . . . .            | 4         |
| <b>3</b> | <b>Реализация</b>                          | <b>5</b>  |
| <b>4</b> | <b>Результат</b>                           | <b>6</b>  |
| <b>5</b> | <b>Приложение</b>                          | <b>10</b> |
| <b>6</b> | <b>Список литературы</b>                   | <b>11</b> |

## Список иллюстраций

|   |   |   |
|---|---|---|
| 1 | Схема установки для исследования фотоэлектрических характеристик.     | 2 |
| 2 | Исходные данные из экспериментов . . . . .                            | 6 |
| 3 | Интервальное представление исходных данных . . . . .                  | 6 |
| 4 | Линейная модель дрейфа данных . . . . .                               | 7 |
| 5 | Гистограммы значений множителей коррекции $w$ . . . . .               | 7 |
| 6 | Скорректированные модели данных . . . . .                             | 8 |
| 7 | Гистограммы скорректированных данных . . . . .                        | 8 |
| 8 | Значение коэффициента Жаккара от калибровочного множителя от $R_{21}$ | 9 |
| 9 | Гистограмма объединённых данных при оптимальном значении $R_{21}$ .   | 9 |

## Список таблиц

# 1 Постановка задачи

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

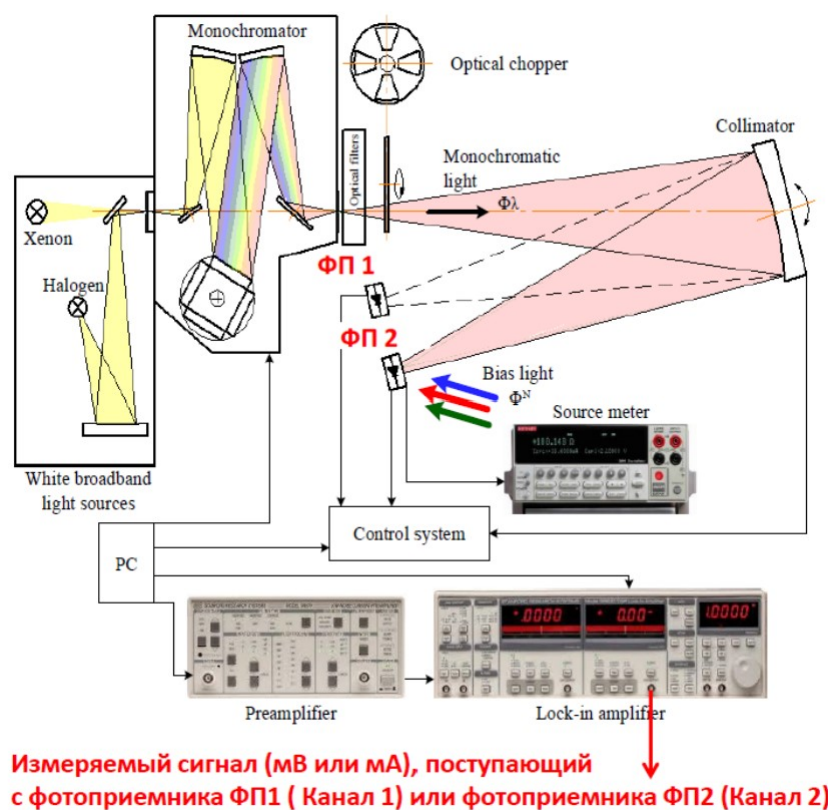


Рис. 1: Схема установки для исследования фотоэлектрических характеристик.

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

$$QE_2 = \frac{I_2}{I_1} * QE_1$$

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.** Требуется определить коэффициент калибровки

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

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

## 2 Теория

### 2.1 Представление данных

В первую очередь представим данные таким образом, чтобы применить понятия статистики данных с интервальной неопределенностью.

Один из распространенных способов получения интервальных результатов в первичных измерениях - это "обинтерваливание" точечных значений, когда к точечному базовому значению  $x_0$ , которое считывается по показаниям измерительного прибора, прибавляется *интервал погрешности*  $\epsilon$ :

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

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  $x = (x_1, x_2, \dots, x_n)$ .

### 2.2 Линейная регрессия

#### 2.2.1 Описание модели

Линейная регрессия - регрессионная модель зависимости одной переменной от другой с линейной функцией зависимости:  $y_i = X_i b_i + \epsilon_i$  где  $X$  - заданные значения,  $y$  - параметры отклика,  $\epsilon$  - случайная ошибка модели. В случае, если у нас  $y_i$  зависит от одного параметра  $x_i$ , то модель выглядит следующим образом:

$$y_i = b_0 + b_1 * x_i + \epsilon_i$$

В данной модели мы пренебрегаем погрешностью 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 модулей. Этот метод основывается на минимизации нормы разности последовательности:

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

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

$$\sum_{i=1}^n |w_i| \rightarrow \min$$

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

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

$$1 \leq w_i, i = 1..n$$

## 2.3 Предварительная обработка данных

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

$$Lin(n) = A + B * n, n = 1, 2, \dots N$$

Поставив и решив задачу линейного программирования, найдем коэффициенты А, В 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:

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

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

$$I^c(n) = I^f(n) - B * n, n = 1, 2, \dots N$$

## 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 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  $[-1; 1]$ .

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

## 2.5 Процедура оптимизации

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

$$JK(I_1^c(n) * R \cup I_2^c(n)) \Rightarrow \max$$

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 выполнена на языке Python версии 3.10.  
Использовались дополнительные библиотеки:

1. `scipy`
2. `numpy`
3. `matplotlib`

## 4 Результат

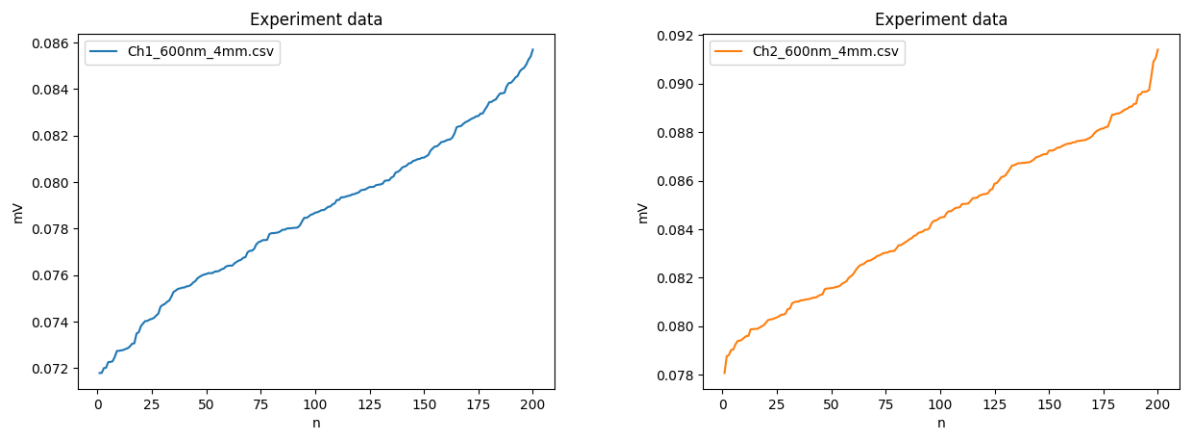


Рис. 2: Исходные данные из экспериментов

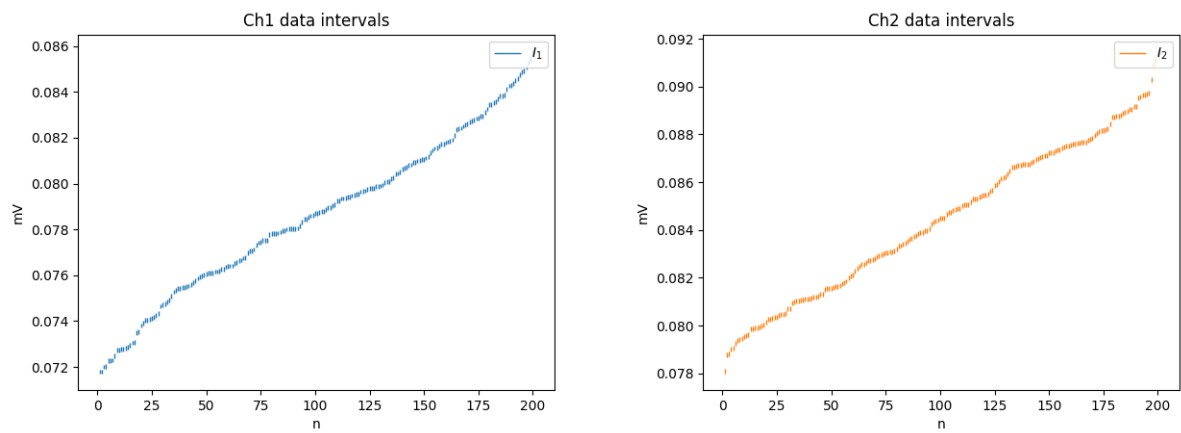


Рис. 3: Интервальное представление исходных данных

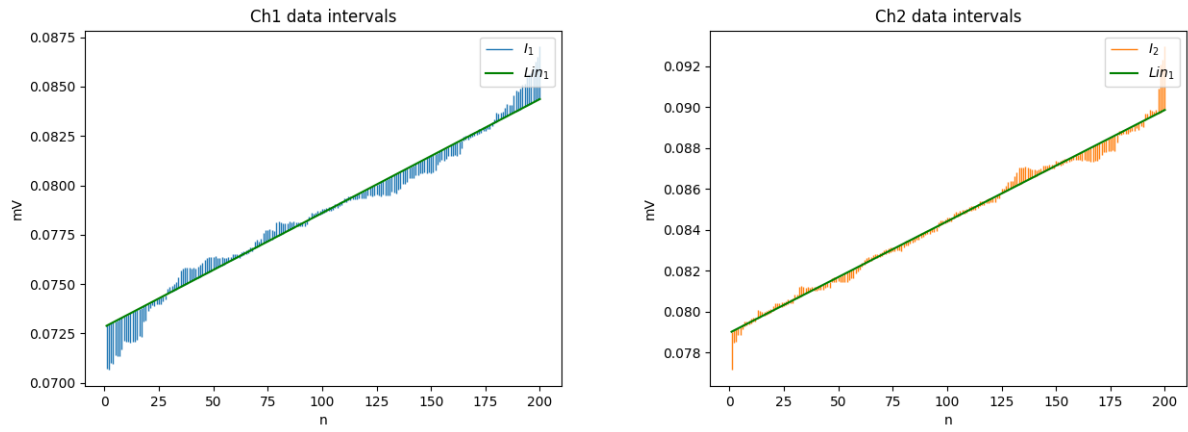


Рис. 4: Линейная модель дрейфа данных

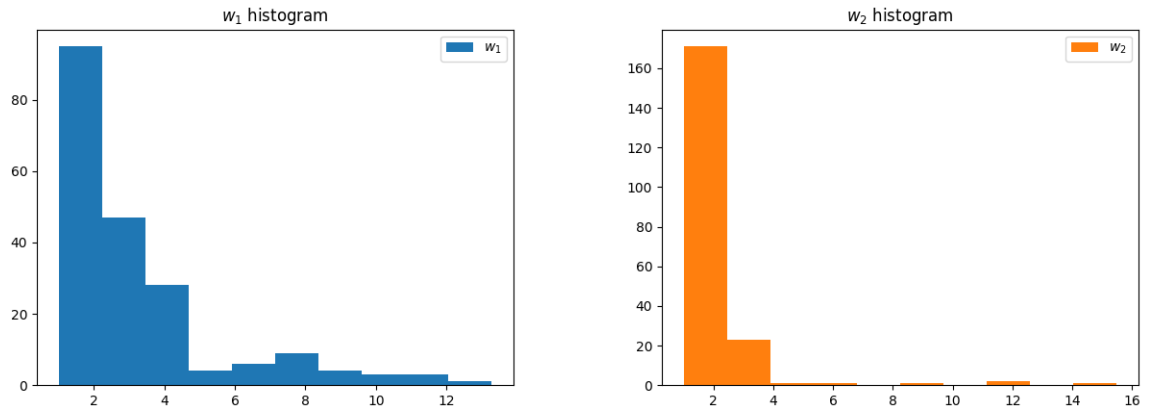


Рис. 5: Гистограммы значений множителей коррекции  $w$

Результаты линейного приближения токов:

- Первый фотоприемник

$$A_1 = 0.0728307$$

$$B_1 = 5.76887e - 05$$

- Второй фотоприемник

$$A_2 = 0.0789563$$

$$B_2 = 5.44961e - 05$$



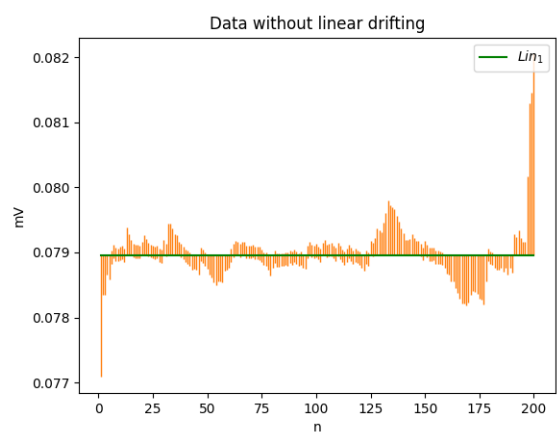
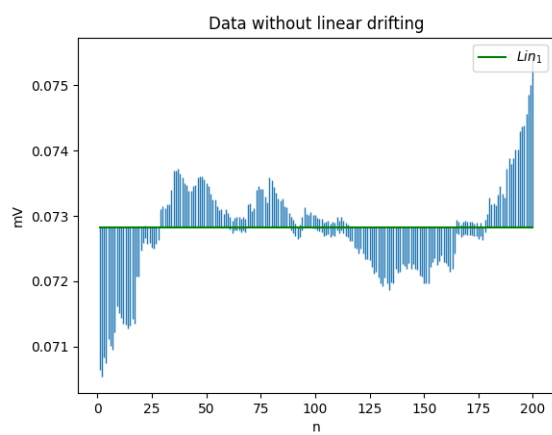


Рис. 6: Скорректированные модели данных

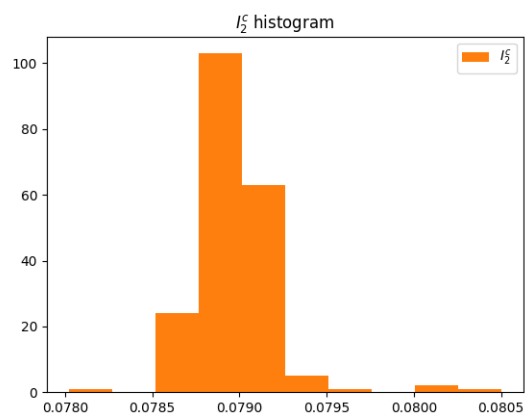
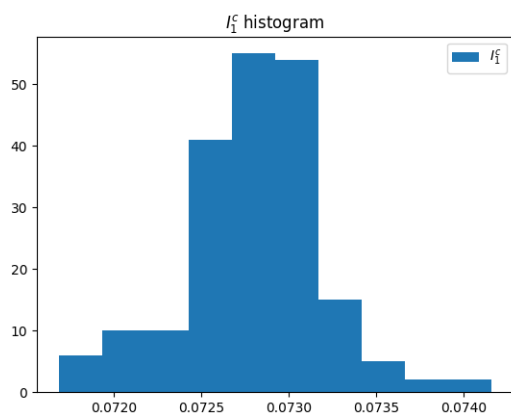


Рис. 7: Гистограммы скорректированных данных

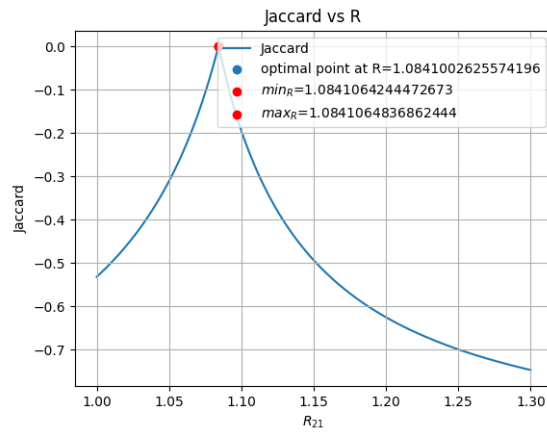


Рис. 8: Значение коэффициента Жаккара от калибровочного множителя от  $R_{21}$

$$JK(x) = -0.11943227$$

$$R_{opt} = 1.0841$$

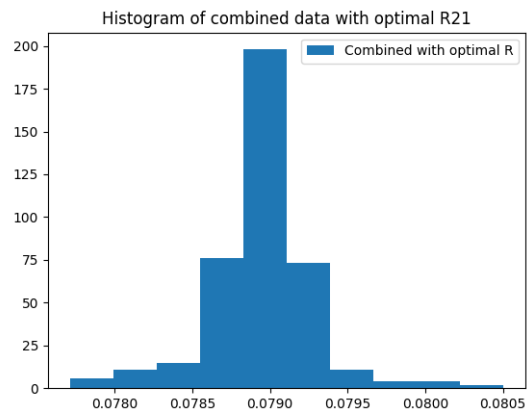


Рис. 9: Гистограмма объединённых данных при оптимальном значении  $R_{21}$

## 5 Приложение

Код программы GitHub URL:

<https://github.com/Fourroubles/Math-Statistic>

## 6 Список литературы

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