

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ
ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ
ВЫСШАЯ ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ФИЗИКИ

Анализ данных с интервальной неопределённостью
Отчёт по лабораторной работе №1

Выполнил:

Студент: Ли Жуйци

Группа: 5040102/10201

Принял:

к. ф.-м. н., доцент

Баженов Александр Николаевич

2022 г.

Содержание

1. Постановка задача	3
2. Теория	4
2.1. Представление данных	4
2.2. Простая линейная регрессия	4
2.2.1. Описание модели	4
2.2.2. Метод наименьших модулей	4
2.3. Предварительная обработка данных	5
2.4. Коэффициент Жаккара	5
2.5. Процедура оптимизации	6
3. Реализация	7
4. Результаты	8
5. Обсуждение	11
6. Список литературы	12

Список иллюстраций

1. Схема установки для исследования фотоэлектрических характеристик. . .	3
2. Выборки полученные в ходе эксперимента	8
3. Интервальное представление данных с первой и второй выборки	8
4. Линейные модели данных	8
5. Гистограммы значений множителей коррекции w	9
6. I_1^c и I_2^c	9
7. Гистограммы I_1^c и I_2^c	9
8. Значение коэффициента Жаккара от калибровочного множителя	10
9. Гистограмма объединённых данных при оптимальном значении R_{21} . . .	10

1. Постановка задачи

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

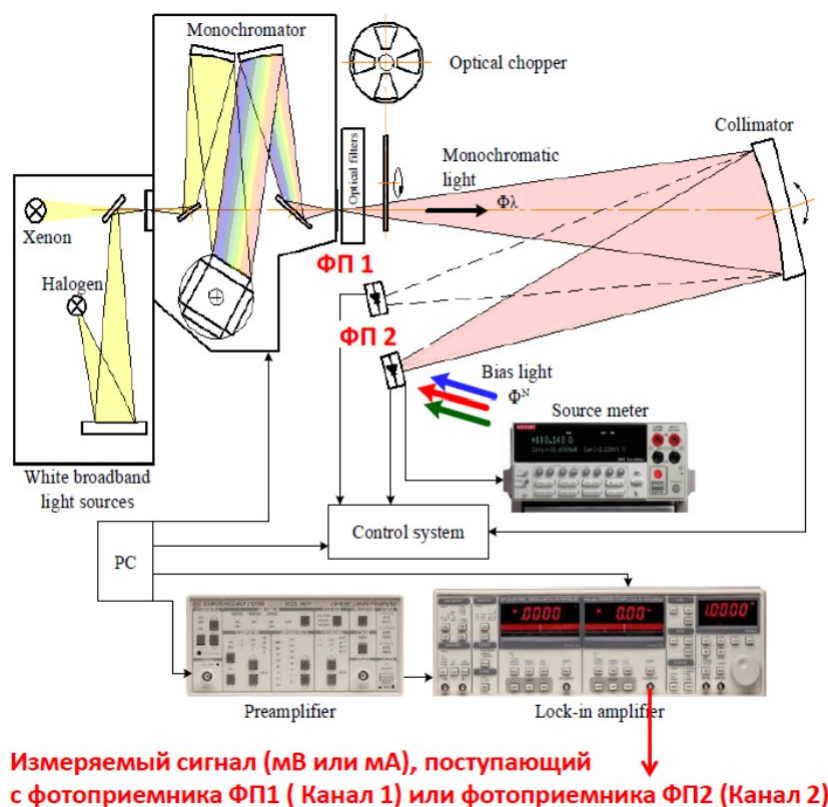


Рис. 1. Схема установки для исследования фотоэлектрических характеристик.

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

$$QE_2 = \frac{I_2}{I_1} * QE_1 \quad (1)$$

QE_2, QE_1 - эталонная эффективности эталонного и исследуемого датчиков, I - измеренные токи. Данные с датчиков находятся в файлах Канал2_700nm_2nm.csv и Канал1_700nm_2nm.csv.

Требуется определить коэффициент калибровки

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

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 распространённых способов получения интервальных результатов в первичных измерениях - это "обинтервализация" точечных значений, когда к точечному базовому значению x_0 , которое считывается по показаниям измерительного прибора, прибавляется *интервал погрешности* ϵ :

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

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

2.2. Простая 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, ϵ - случайная ошибка модели. В случае, если у нас y_i зависит от одного параметра x_i , то модель выглядит следующим образом:

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

В данной модели мы считаем, что у заданных значений нет погрешности. Полагаем, 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:

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

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

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

$$b_0 + b_1 * x_i - w_i * \epsilon \leq y_i, i = 1..n \quad (7)$$

$$b_0 + b_1 * x_i + w_i * \epsilon \leq y_i, i = 1..n \quad (8)$$

$$1 \leq w_i, i = 1..n \quad (9)$$

2.3. Предварительная обработка данных

Из последующих результатов ясно, что для коэффициента калибровки необходима предварительная обработка данных. Для этого можем задаться линейной моделью дрейфа.

$$Lin_i(n) = A_i + B_i * n, n = 1, 2, ...N \quad (10)$$

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:

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

После построения линейной модели дрейфа необходимо построить "спрямленные" данные выборки: получить их можно путем вычитания из исходных данных линейную компоненту:

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

2.4. Коэффициент Жаккара

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

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

в качестве меры рассматривается ширина интервала, а вместо операций пересечения и объединения - операция взятия минимума и максимума по включению двух величин в интервальной арифметике (Каухера) соответственно. Можно заметить, что в минимума по включению быть неправильным интервалом, коэффициент Жаккара может нормирован в интервале $[-1; 1]$.

2.5. Процедура оптимизации

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

$$JK(I_1^c(n) * R \cup I_2^c(n)) \Rightarrow \max \quad (14)$$

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

Данная работа реализована на языке программирования Python 3.10 с использованием библиотек NumPy, Matplotlib и SciPy. Отчёт подготовлен в онлайн-редакторе LaTeX "Overleaf".

4. Результаты

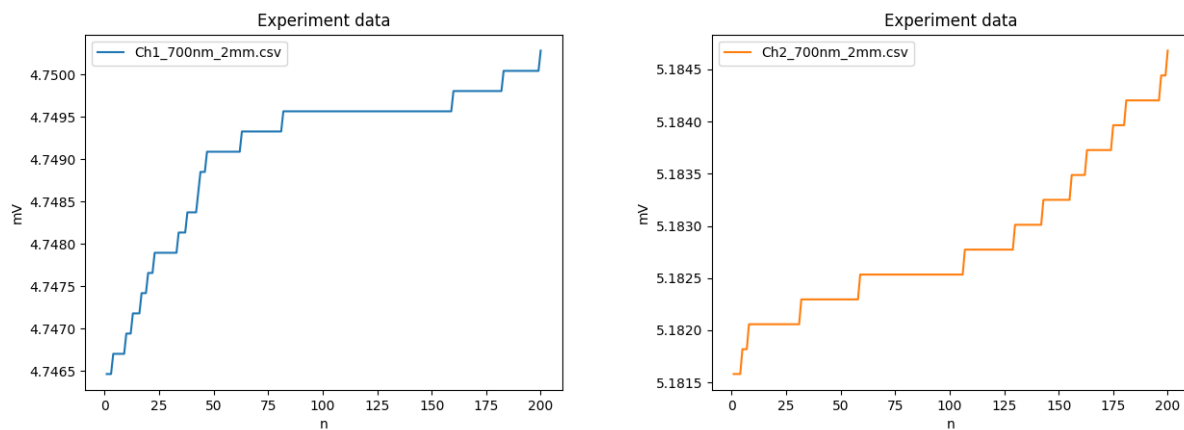


Рис. 2. Выборки полученные в ходе эксперимента

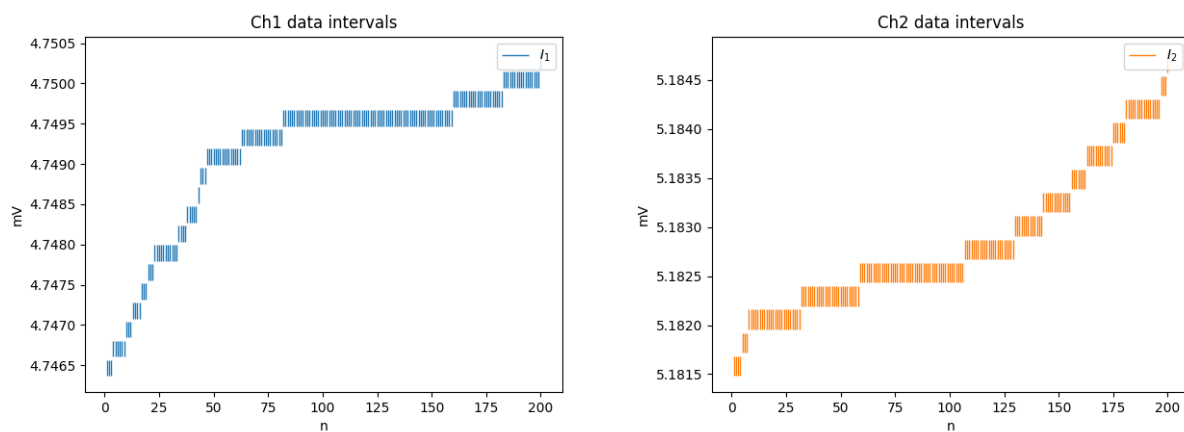


Рис. 3. Интервальное представление данных с первой и второй выборки

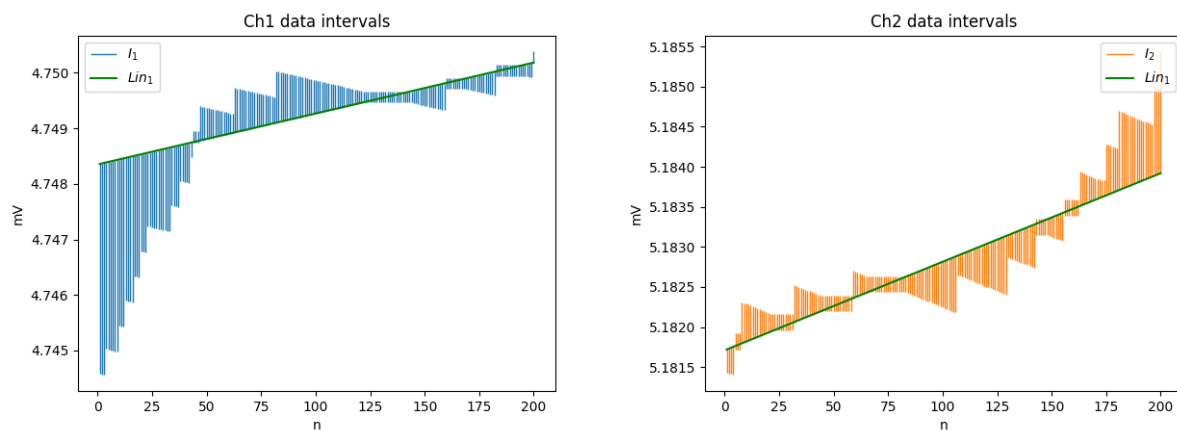


Рис. 4. Линейные модели данных

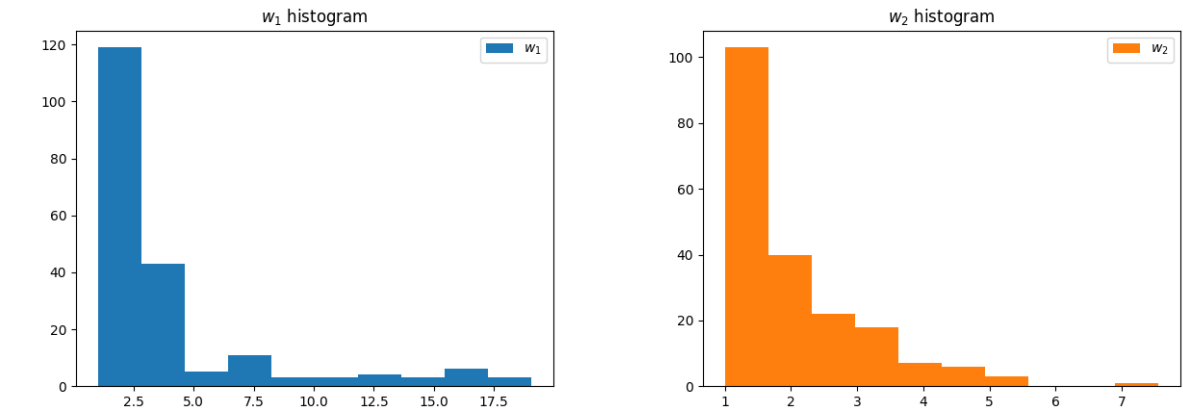


Рис. 5. Гистограммы значений множителей коррекции w

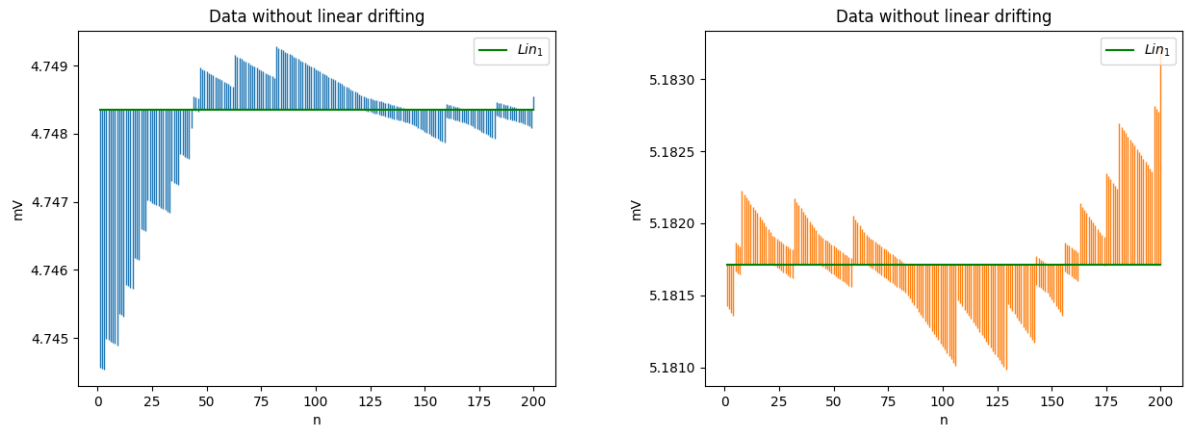


Рис. 6. I_1^c и I_2^c

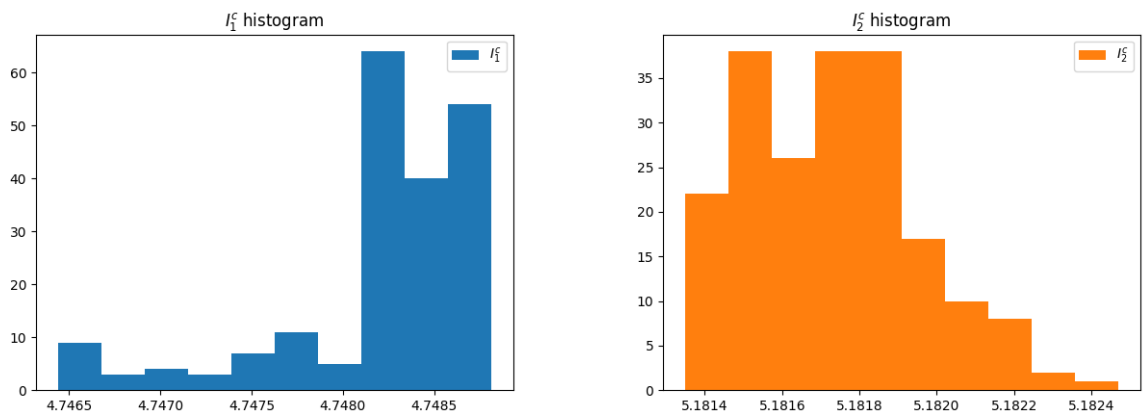


Рис. 7. Гистограммы I_1^c и I_2^c

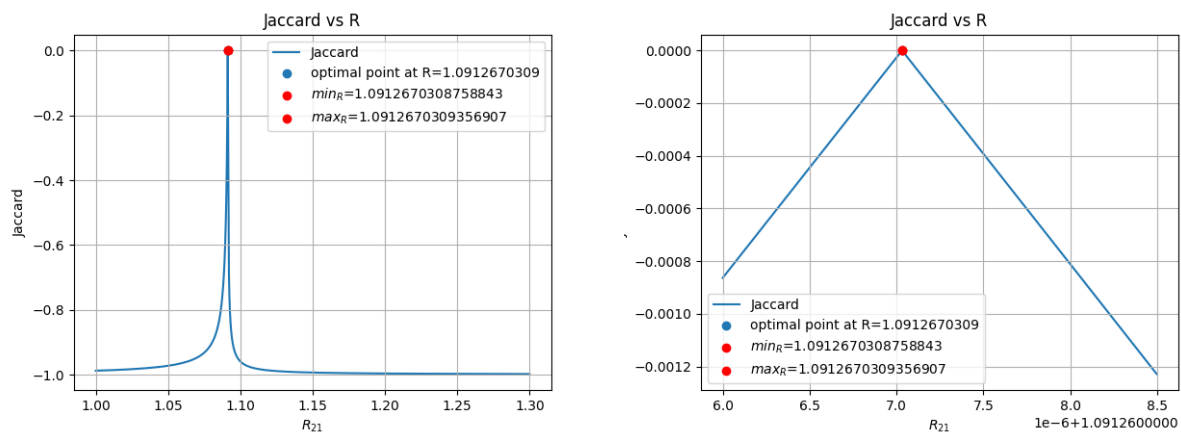


Рис. 8. Значение коэффициента Жаккара от калибровочного множителя

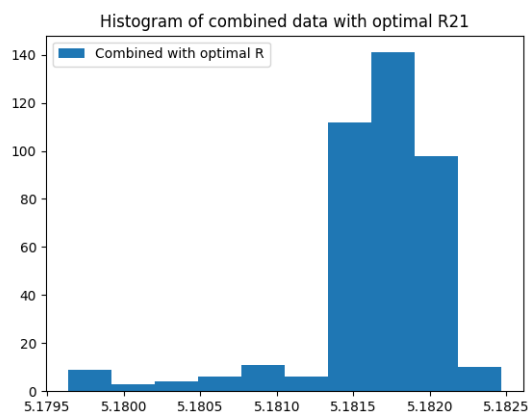


Рис. 9. Гистограмма объединённых данных при оптимальном значении R_{21}

5. Обсуждение

Множители коррекции w .

Рассмотрим Рис.5, видно, что половина данных не требует коррекции. Это означает, что линейная модель дрейфа данных является разумным приближением.

Гистограмма I_1^f, I_2^f и Совмещённой выборки с оптимальным коэффициентом калибровки.

Сравнивая гистограмму объединённых данных при оптимальном значении R_{21} (Рис. 9) с гистограммами скорректированных данных (Рис. 7), видно, что гистограмма объединённых данных повторяет форму гистограммы входных данных с ФП1, однако пик гистограммы смещен в сторону значения пика на гистограмме ФП2.

Коэффициент Жаккара.

Рассмотрим рис.8. Оптимальное значение параметра калибровки R_{21} можно принять равным 1.091267. Однако видно, что коэффициент Жаккара при оптимальном значении превышает 0, а интервал, при котором $JK \geq 0$, ($10^{-9} - 10^{-10}$). Это связано с наличием различных погрешностей, которые на практике невозможно устранить, но несмотря на их наличие, поведение функции Жаккара позволило найти оптимальный калибровочный коэффициент. Однако знак коэффициента Жаккара может свидетельствовать о том, является ли минимум по включению правильным интервалом, что в свою очередь говорит о совместности двух выборок. Таким образом можно сказать что область где $JK_{R_{21}} \geq 0$ является оценкой искомой величины R_{21} .

6. Список литературы

1. М.З.Шварц. Данные технологических испытаний оборудования для калибровки фотоприемников солнечного излучения. 2022.
2. <https://github.com/Li-Rui-QI/interval-uncertainty.git>