Санкт-Петербургский политехнический университет имени Петра Великого
Физико-Механический институт

Лабораторная работа №4
по дисциплине
Математическая статистика

Выполнил студент группы 5030102/00101 Маковеев Лев
Преподаватель Баженов Александр Николаевич

Санкт-Петербург, 2023

Содержание

Список иллюстраций

1 Постановка задачи

1. Для выборки данных $\mathbf{X_1}$ с интервальной неопределенностью ϵ восстановить регрессионную зависимость с помощью метода варьирования неопределенности измерений с расширением интервалов, с помощью метода варьирования неопределенности измерений с расширением и сужением интервалов. Привести визуализацию регрессионных прямых и оптимизированных коэффициентов.

2. Привести анализ и визуализацию регрессионных остатков, получившихся в результате работы методов.

3. Получить и построить информационное множество задачи восстановления регрессионной зависимости. Построить коридор совместных зависимостей. Построить прогноз вне области данных.

2 Теория

2.1 Методы восстановления регрессионной зависимости

2.1.1 Варьирование неопределённости измерений

Имеется выборка данных $\mathbf{X_1}$ с интервальной неопределённостью. Для дальнейшей работы используется модель данных с уравновешенным интервалом погрешности:

$$\mathbf{x} = \overset{o}{x} + \epsilon, \quad \epsilon = [-10^{-3}, 10^{-3}]$$

Если величину коррекции каждого интервального наблюдения выборки выражать коэффициентом его уширения $\omega_i \geq 1$, а общее изменение выборки характеризовать суммой этих коэффициентов, то минимальная коррекция выборки в виде вектора коэффициентов $\omega = (\omega_1, ..., \omega_n)$, необходимая для совместности задачи построения зависимости $x = \beta_0 + \beta_1 \dot{i}$ может быть найдена решением задачи условной оптимизации

$$\min_{\omega,\beta} \sum_{i=1}^{n} \omega_i$$

при ограничениях

$$\begin{cases} mid\mathbf{x_i} - \omega_i\epsilon \leq \beta_0 + \beta_1 \dot{i} \leq mid\mathbf{x_i} + \omega_i\epsilon, \\ \omega_i \geq 1, \end{cases} \qquad i = 1, 2, ..., n.$$

Результирующие значения коэффициентов ω_i, строго превосходящие единицу, указывают на наблюдения, которые требуют уширения интервалов неопределённости для обеспечения совместности данных и модели.

2.1.2 Варьирование неопределённости измерений с расширением и сужением интервалов

Поставим задачу условной оптимизации следующим образом:

$$\min_{\omega,\beta} \sum_{i=1}^{n} \omega_i$$

при ограничениях

$$\begin{cases} mid\mathbf{x_i} - \omega_i\epsilon \le \beta_0 + \beta_1 \dot{i} \le mid\mathbf{x_i} + \omega_i\epsilon, \\ \omega_i \ge 0, \end{cases} \qquad i = 1, 2, ..., n.$$

Отличие двух постановок состоит в том, что интервалы измерений могут как расширяться в случае $\omega_i > 1$, так и сужаться при $0 \le \omega_i < 1$.

2.2 Информационное множество задачи

Один из главных вопросов при построении регрессии – оценивание её параметров. В зависимости от прикладных целей характер и назначение искомых оценок могут существенно разниться.

Внешняя интервальная оценка параметра определяется минимальным и максимальным значениями, которых может достигать значение параметра в информационном множестве.

В совокупности интервальные оценки параметров задают брус, описанный вокруг информационного множества и именуемый внешней интервальной оболочкой информационного множества

2.3 Построение прогноза внутри и вне области данных

Одним из способов использования регрессионной модели является предсказание значений выходной переменной для заданных значений входной. С помощью построенной выше модели можно получить прогнозные значения выходной переменной в точках эксперимента.

Ценность модели также заключается в возможности её употребления для предсказания выходной переменной в точках, где измерения не производились. Расширив область определения аргумента для построенной выше модели, можно получить оценки для значений выходной переменной (экстраполяция).

3 Результаты

В качестве выборки брались данные из файла «Канал 1_600nm_0.03.csv»

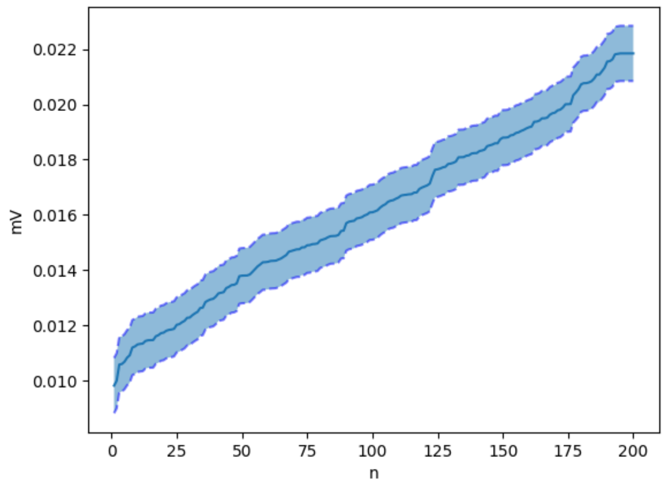

Рис. 1: Данные выборки $\mathbf{X_1}$ с интервальной неопределённостью

3.1 Восстановление регрессионной зависимости

3.1.1 Варьирование неопределённости измерений

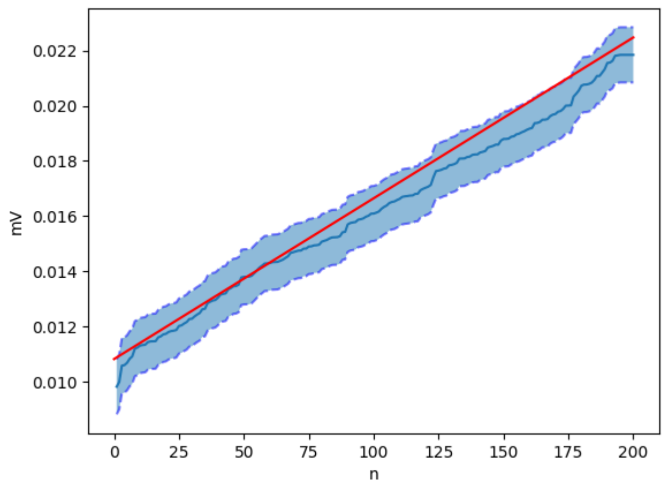

Рис. 2: Диаграмма рассеяния выборки $\mathbf{X_1}$ и регрессионная прямая по модели 2.1.1

На Рис. 2 красным цветом приведена регрессионная прямая.
Вычисления с использованием программы 2.1.1 дают следующие результаты для регрессионных коэффициентов

$$\beta_0 = 1.08 \cdot 10^{-2}$$

$$\beta_1 = 5.83 \cdot 10^{-5}$$

Все компоненты вектора ω оказались равны 1, то есть, расширения интервалов измерений не понадобилось. Таким, образом

$$\min_{\omega,\beta} \sum_{i=1}^{n} \omega_i = 200$$

3.1.2 Варьирование неопределённости измерений с расширением и сужением интервалов

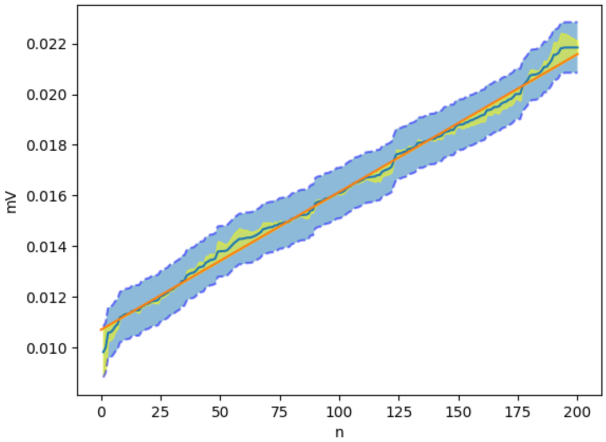

Рис. 3: Диаграмма рассеяния выборки $\mathbf{X_1}$ и регрессионная прямая по модели 2.1.2

На Рис. 3 оранжевым цветом приведена регрессионная прямая.
Вычисления с использованием программы 2.1.2 дают следующие результаты для регрессионных коэффициентов

$$\beta_0 = 1.07 \cdot 10^{-2}$$

$$\beta_1 = 5.44 \cdot 10^{-5}$$

Жёлтым цветом показаны скорректированные интервалы выборки $\mathbf{X_1}$.

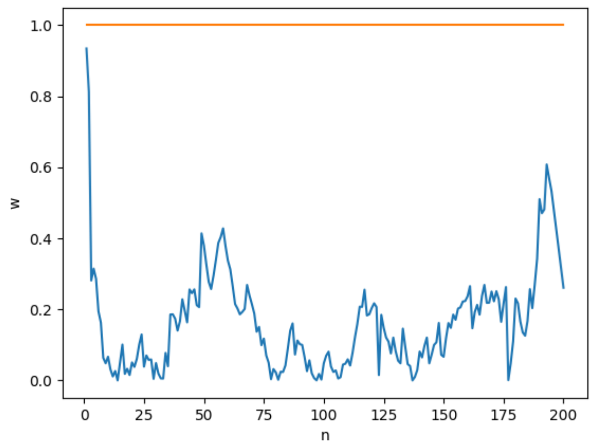

Рис. 4: Вектор ω

Как видно из Рис.4, все интервалы в диапазоне замеров сузились, и получается:

$$\min_{\omega,\beta} \sum_{i=1}^{n} \omega_i = 32.8 < 200$$

3.2 Анализ регрессионных остатков

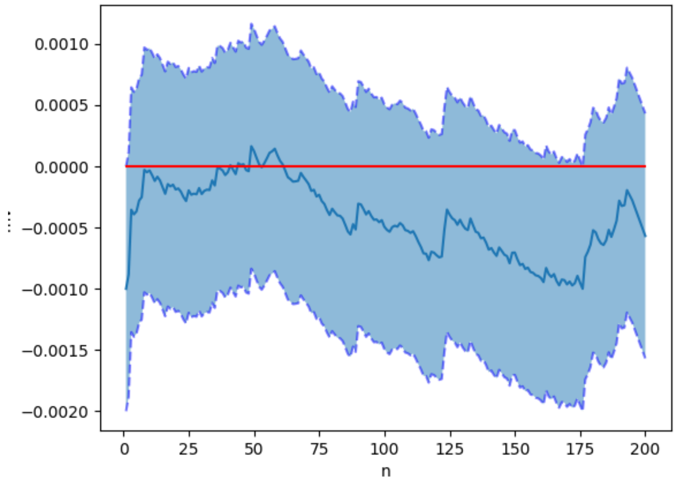

Рис. 5: Диаграмма рассеяния регрессионных остатков по модели 2.1.1

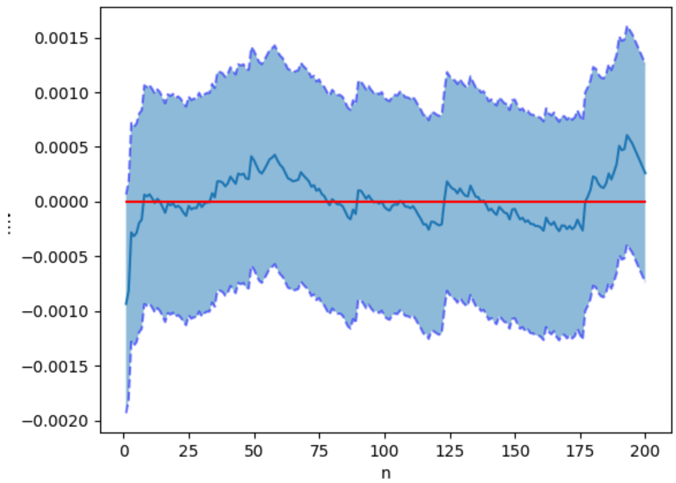

Рис. 6: Диаграмма рассеяния регрессионных остатков по модели 2.1.2

3.3 Информационное множество задачи

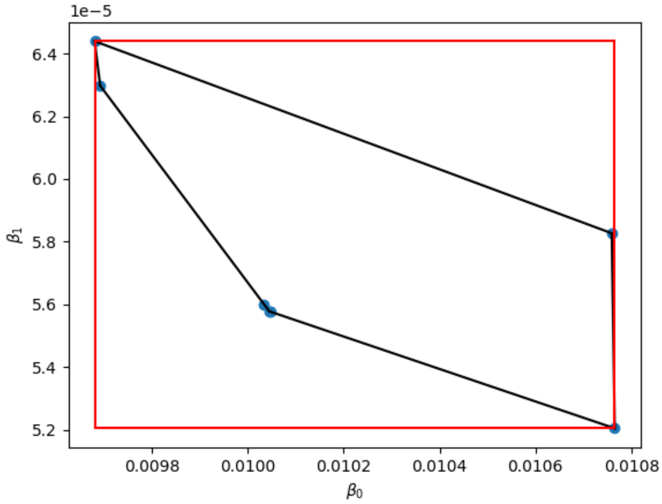

Рис. 7: Информационное множество по модели 2.1.2, интервальная оболочка — красный брус

Внешние интервальные оценки параметров модели

$$mid\ \beta_0 = [9.7 \cdot 10^{-3}, 1.08 \cdot 10^{-2}],$$

$$mid\ \beta_1 = [5.2 \cdot 10^{-5}, 6.4 \cdot 10^{-5}].$$

3.4 Коридор совместных зависимостей

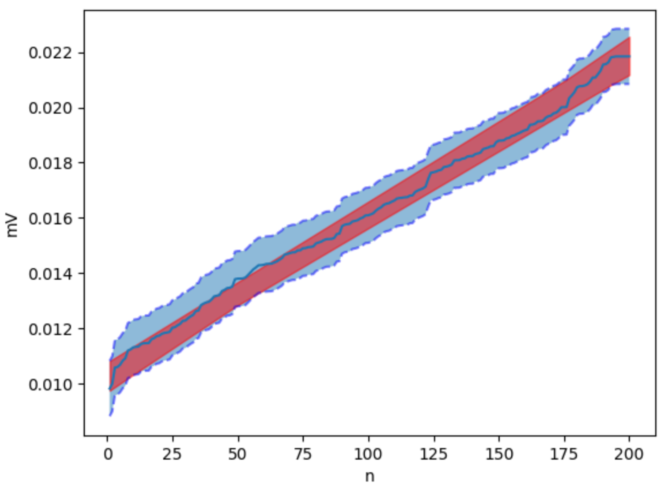

Рис. 8: Коридор совместных зависимостей

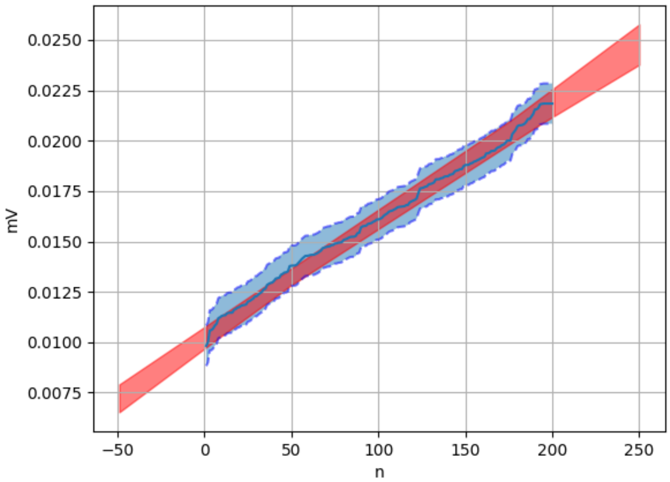

Рис. 9: Коридор совместных зависимостей. Построение прогноза

4 Обсуждение

4.1 Восстановление регрессионной зависимости

Постановка задачи с возможностью одновременного увеличения и уменьшения радиусов неопределённости измерений позволяет более гибко подходить к задаче оптимизации. Действительно, регрессионная прямая, полученная по модели 2.1.2 более точно и естественно аппроксимирует данные выборки.

4.2 Анализ регрессионных остатков

Из сравнения Рис. 5 и Рис. 6 видно, что интервальные выборки остатков получились с весьма разными свойствами. Формально диаграмма рассеяния на первом рисунке является более узкой, то есть внешняя оценка более компактная. В то же время вторая диаграмма рассеяния выглядит более естественно.

4.3 Коридор совместных зависимостей

Из рис. 9 видно, что величина неопределённости прогнозов растёт по мере удаления от области, в которой производились исходные измерения. Это обусловлено видом коридора зависимостей, расширяющимся за пределами области измерений, и согласуется со здравым смыслом.

5 Реализация

Лабораторная работа выполнена с использованием языка программирования Python 3.10 в среде разработки PyCharm Community с использованием библиотек matplotlib, numpy, scipy, pandas.

6 Дополнительно

Код лабораторной работы:
https://github.com/S0krat/MakoveevLev_mathstat2023/tree/main/Lab4