

Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

Отчет по лабораторной №2 по дисциплине «Стохастические модели и анализ данных»

Выполнил студент:
Коваленко Надежда
группа:
3640102/90201

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

Санкт-Петербург, 2020 г.

Оглавление

Постановка задачи	3
Теория	3
Критерий Фишера	3
Правило Стерджесса	3
Реализация.....	3
Практические результаты	4
Выводы.....	6
Литература	6

Постановка задачи

Есть данные, полученные регистратором. По этим данным нужно провести дисперсионный анализ для значений сигнала с применением критерия Фишера.

Теория

Критерий Фишера

Дисперсию совокупности можно оценить двумя способами [1].

Во-первых, дисперсия, вычисленная для каждой группы — это оценка дисперсии совокупности. Поэтому дисперсию совокупности можно оценить на основании групповых дисперсий. Такая оценка не будет зависеть от различий групповых средних.

Во-вторых, разброс выборочных средних тоже позволяет оценить дисперсию совокупности. Понятно, что такая оценка дисперсии зависит от различий выборочных средних. В качестве оценки дисперсии совокупности возьмем среднее выборочных дисперсий. Эта оценка называется внутригрупповой дисперсией. Обозначим ее $s_{ВНУ}^2$.

$$s_{ВНУ}^2 = \frac{1}{k} \sum_{i=1}^k s_i^2 = \frac{1}{k} \sum_{i=1}^k \frac{\sum_{j=1}^n (x_{ij} - \bar{X})^2}{k-1} \quad (1)$$

Где \bar{X} — среднее для части выборки, k — количество частей, на которое делим сходную выборку, n — количество элементов в подвыборке.

Также нужно вычислить межгрупповую дисперсию. Обозначается она $s_{МЕЖ}^2$.

Вычисление межгрупповой дисперсии происходит в несколько этапов:

1. Вычисление среднего значения для всех выбранных подвыборок $(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k)$.
2. Вычисление среднего этих средних: $\bar{X} = \frac{1}{k} \sum_{i=1}^k \bar{X}_i$.
3. $s_{МЕЖ}^2 = k * \frac{\sum_{i=1}^k (\bar{X}_i - \bar{X})^2}{k-1} \quad (2)$

Тогда $F = \frac{s_{МЕЖ}^2}{s_{ВНУ}^2} \quad (3).$

Правило Стерджесса

Данное правило [2] используется для определения оптимального количества интервалов, на которые разбивается наблюдаемый диапазон изменения случайной величины при изучении ее распределения.

По этому правилу число интервалов считается по формуле:

$$n = 1 + \lfloor \log_2 N \rfloor$$

Где n — число интервалов, N — общее число наблюдений.

Реализация

Лабораторная работа выполнена на языке Python (версия 3.8) с использованием библиотек NumPy, matplotlib в среде разработки Visual Studio Code. Детали можно посмотреть на GitHub по ссылке в литературе [3].

Практические результаты

Рассмотрим некоторые сигналы, которые представлены в файле. Их там 800. В каждом сигнале по 1024 значения.

Возьмем один из сигналов и проанализируем его. Рассмотрим сигнал 432. Он представлен на рис.1.

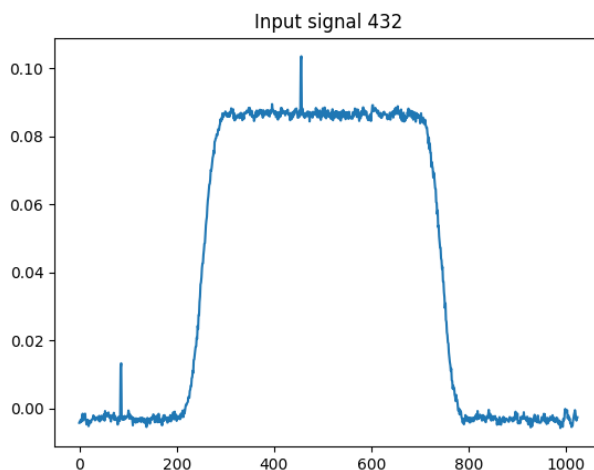


Рисунок 1. Исходный сигнал

Необходимо разделить сигнал на подобласти: сигнал, фон, переходные процессы. Сделано это было с помощью построения гистограммы, которая представлена на рис.2.

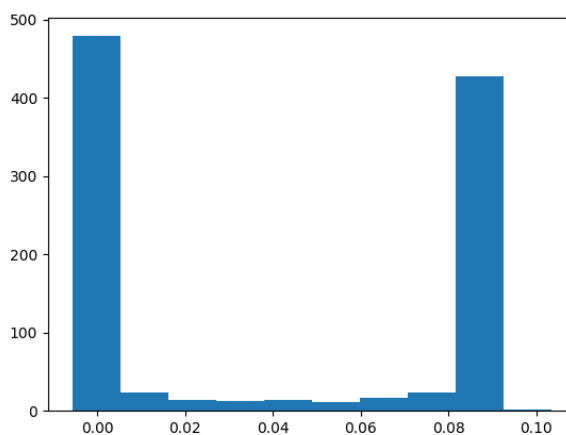


Рисунок 2. Гистограмма сигнала

Столбец гистограммы, в котором наибольшее число значений, отвечает за принадлежность к фону, следующий по величине столбец – за сигнал, а остальное – переходные процессы.

Однако, прежде чем разделить сигнал на области однородности необходимо определить наличие выбросов и сгладить их. Это было выполнено с помощью медианного фильтра, то есть значение выброса становится равным среднему арифметическому его соседних элементов. Результат сглаживания представлен на рис. 3.

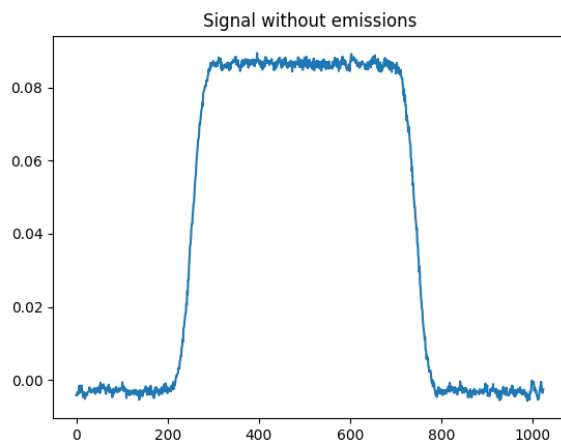


Рисунок 3. Сигнал без выбросов

С помощью гистограммы теперь можно разделить сигнал на разные области, которые представлены на рис. 4.

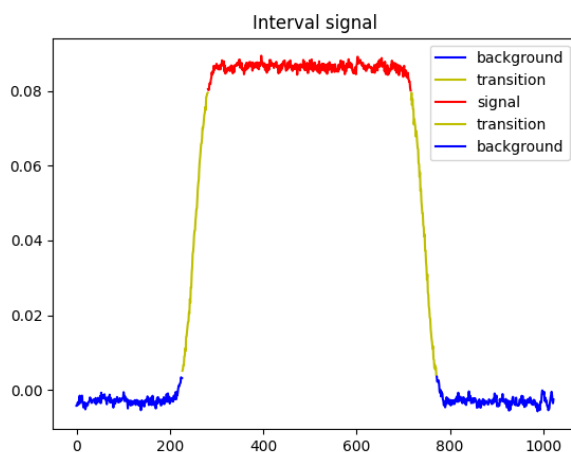


Рисунок 4. Деление преобразованного сигнала на участки однородности

Данный сигнал после разделения на области с выбросами представлен на рис.5.

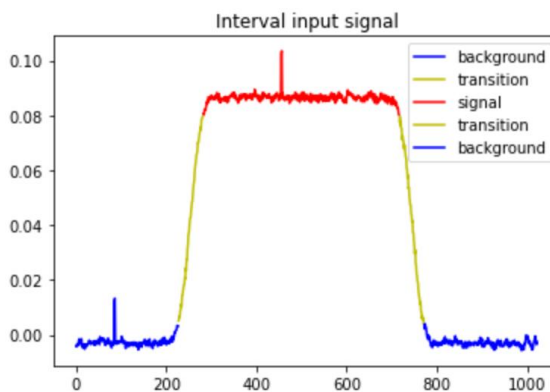


Рисунок 5. Деление исходного сигнала на участки однородности

Чтобы определить однородность каждой части сигнала, необходимо применить критерий Фишера. Результаты вычисления представлены в табл.1.

Номер промежутка	Тип промежутка	Число разбиений (k)	Значение межгрупповой дисперсии (2)	Значение внутри групповой дисперсии (1)	Значение критерия Фишера (3)
1	Фон	4	0.000006	0.000067	0.036318
2	Переходный процесс	5	0.003587	0.000054	66.503898
3	Сигнал	5	0.000001	0.000059	0.020053
4	Переходный процесс	5	0.003595	0.000059	61.032616
5	Фон	4	0.000001	0.000035	0.032911

Для данного сигнала критерий Фишера будет равен:

- для левого фона - 0.036318,
- для левого переходного процесса - 66.503898,
- для сигнала - 0.020053,
- для правого переходного процесса - 61.032616,
- для правого фона - 0.032911

Выводы

Так как критерий Фишера для фона и сигнала близок к 1, то эти части сигнала – однородные.

Литература

[1] Гланц, С. Медико-биологическая статистика. Пер. с англ. — М.: Практика, 1998. — 459 с.

[2] Правило Стерджесса

URL: https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%B0%D0%B2%D0%B8%D0%BB%D0%BE_%D0%A1%D1%82%D1%91%D1%80%D0%B4%D0%B6%D0%B5%D1%81%D0%B0

[3] Реализация применения критерия хи-квадрат.

URL: <https://github.com/NadezhdaKovalenko/StochasticModels>