Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЁТ ПО КУРСОВОЙ РАБОТЕ ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Выполнил
студент группы 3630102/80201

Игнатьев Даниил Дмитриевич

Проверил
к. ф.-м. н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2021

СОДЕРЖАНИЕ

СПИСОК ИЛЛЮСТРАЦИЙ

1 Постановка задачи

Пусть имеется тороидальная камера с магнитными катушками (ТОКАМАК) и детектор-счётчик. В каждый момент времени специальным программным обеспечением получаются сигналы, преобразованные в данные, о текущем местоположении плазмы и количестве выделившихся частиц. В нашей работе необходимо:

1. Считать и обработать полученные данные с прибора
2. Построить графики зависимостей положенения плазмы от времени и количества выделенных нейтронов от времени
3. Найти функцию корреляции между движением плазмы и количеством вылетающих нейтронов

2 Теория

Корреляция (correlation), и ее частный случай для центрированных сигналов – ковариация, является методом анализа сигналов. Приведем один из вариантов использования метода. Допустим, что имеется сигнал s(t), в котором может быть (а может и не быть) некоторая последовательность x(t) конечной длины T, временное положение которой нас интересует. Для поиска этой последовательности в скользящем по сигналу s(t) временном окне длиной T вычисляются скалярные произведения сигналов s(t) и x(t). Тем самым мы "прикладываем"искомый сигнал x(t) к сигналу s(t), скользя по его аргументу, и по величине скалярного произведения оцениваем степень сходства сигналов в точках сравнения.

Корреляционный анализ дает возможность установить в сигналах (или в рядах цифровых данных сигналов) наличие определенной связи изменения значений сигналов по независимой переменной, то есть, когда большие значения одного сигнала (относительно средних значений сигнала) связаны с большими значениями другого сигнала (положительная корреляция), или, наоборот, малые значения одного сигнала связаны с большими значениями другого (отрицательная корреляция), или данные двух сигналов никак не связаны (нулевая корреляция).

В функциональном пространстве сигналов эта степень связи может выражаться в нормированных единицах коэффициента корреляции, т.е. в косинусе угла между векторами сигналов, и, соответственно, будет принимать значения от 1 (полное совпадение сигналов) до -1 (полная противоположность) и не зависит от значения (масштаба) единиц измерений.

В варианте автокорреляции (autocorrelation) по аналогичной методике производится определение скалярного произведения сигнала s(t) с собственной копией, скользящей по аргументу. Автокорреляция позволяет оценить среднестатистическую зависимость текущих отсчетов сигнала от своих предыдущих и последующих значений (так называемый радиус корреляции значений сигнала), а также выявить в сигнале наличие периодически повторяющихся элементов.

Особое значение методы корреляции имеют при анализе случайных процессов для выявления неслучайных составляющих и оценки неслучайных параметров этих процессов.

2.1 Описание корреляции

Рассмотрим, как можно сравнить две последовательности данных, состоящие из значений, одновременно выбираемых из двух соответсвующих сигналов. Если два сигнала похоже меняются при переходе от точки к точке, то меру их корреляции можно вычислить, взяв сумму произведений соответсвующих пар точек. Данное предложение становится более аргументированным, если рассмотреть две независимые и случайные последовательности данных. В этом случае сумма произведений стремится к исчезающе малому случайному числу по мере увеличения пар точек. Это объясняется тем, что все числа, положительные и отрицательные, равновероятны, так что пары произведений компенсируются при сложении. В то же время, если сумма конечна, это указывает на наличие корреляции. Отрицательная сумма указывает на отрицательную корреляцию, т.е. увеличение одной переменной связано с уменьшением другой. Таким образом, взаимную корреляцию $r_{12}(n)$ двух последовательностей данных $x_1(n)$ и $x_2(n)$, содержащих по N элементов, можно записать как

$$r_{12} = \sum_{n=0}^{N-1} x_1(n)x_2(n)$$

Однако, такое определение взаимной корреляции даёт результат, который зависит от числа взятых точек. Чтобы это исправить, результат нормируется на число точек. Данную операцию можно рассматривать как усреднение суммы произведений. Получаем следующее улучшенное определение:

$$r_{12} = \frac{1}{N}\sum_{n=0}^{N-1} x_1(n)x_2(n)$$

Впрочем, чтобы данное определение можно было использовать, его также нужно модифицировать. В некоторых случаях корреляция, определенная указанным выше способом, может быть нулевой, хотя две последовательности коррелируют на 100%. Это может произойти, например, когда два сигнала идут не в фазе (как часто и бывает). Разность фаз может, например, объясняться тем, что x_1 – некий эталонный сигнал, а x_2 – запаздывающий выход схемы. Чтобы преодолеть подобный сдвиг фаз, необходимо сдвинуть (или задержать) один из сигналов относительно другого. Обычно, чтобы выровнять сигналы перед определением корреляции, x_2 смещается влево. Это эквивалентно замене $x_2(n)$ на $x_2(n+j)$, где j представляет величину задержки – число точек выборки, на которое x_2 смещается в лево. Альтернативной и эквивалентной процедурой является смещение x_1 вправо. В результате получаем такую формулу для взаимной корреляции:

$$r_{12}(j) = \frac{1}{N}\sum_{n=0}^{N-1} x_1(n)x_2(n+j) = r_{12}(-j) = \frac{1}{N}\sum_{n=0}^{N-1} x_2(n)x_1(n-j)$$

На практике, когда два сигнала коррелируют, их фазовая связь скорее всего неизвестна, так что корреляцию нужно находить для нескольких различных задержек, чтобы установить наибольшее значение корреляции, которое затем считается истинным.
Разумеется, также можно рассмотреть корреляцию в непрерывной области, и некоторые аналоговые схемы корреляции организованны именно так. В непрерывной области $n \to t$ и $j \to \tau$ и

$$r_{12}(\tau) = \lim_{t\to\infty} \frac{1}{T}\int_{-\infty}^{\infty} x_1(t)x_2(t+\tau)dt$$

На практике обрабатываться будут записи конечной длины:

$$r_{12}(\tau) = \frac{1}{T}\int_{0}^{T} x_1(t)x_2(t+\tau)dt$$

Есть и другая сложность, связанная с нахождением взаимной корреляции последовательностей данных конечной длины. Если при смещении x_2 влево сигналы уже не перекрываются, и данные в конце последовательностей не формируют парные произведения – возникает так называемый *краевой эффект*. Одно из возможных решений возникшей проблемы заключается в том, чтобы длину одной последовательности сделать в два раза больше длины, необходимой для нахождения корреляции. Для этого можно записать больше данных или, если одна из последовательностей периодична, повторить последовательность. Другое возможное решение – скорректировать все расчитанные значения.

3 Программная реализация

Курсовая работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm.

4 Ход работы

На вход подаётся .txt файл с данными сигнала.

Первый этап работы – обработка информации. Необходимо считать построчно файл, убирая лишние знаки, пробелы и табуляции.

Второй шаг – визуализировать обработанные данные. Необходимо построить графики положения плазмы и количества вылетивших нейтронов от момента времени.

Третий этап – построение функций корреляции. Для этого первым делом из векторов, с которыми мы будем работать удалим невалидные значения, заменив их на 0. Далее применим функцию сдвига, которая будет в зависимости от значения параметра сдвигать элементы вектора. Для построения функции корреляции в цикле будем вызывать функцию сдвига $\mathrm{shift}(\tau)$, далее перемножать компоненты получившихся векторов между друг другом и суммировать результаты. Полученные результаты будем сохранять.

Четвёртый шаг – нормировка данных. Воспользуемся формулой:

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

на выходе получим по оси ординат значения на промежутке [0, 1]

Визуализируем полученные данные, сопоставив каждому значению τ получившееся значение $r_{12}(\tau)$, это и будет график функции взаимной корреляции.

5 Результаты

5.1 Визуализация входных данных

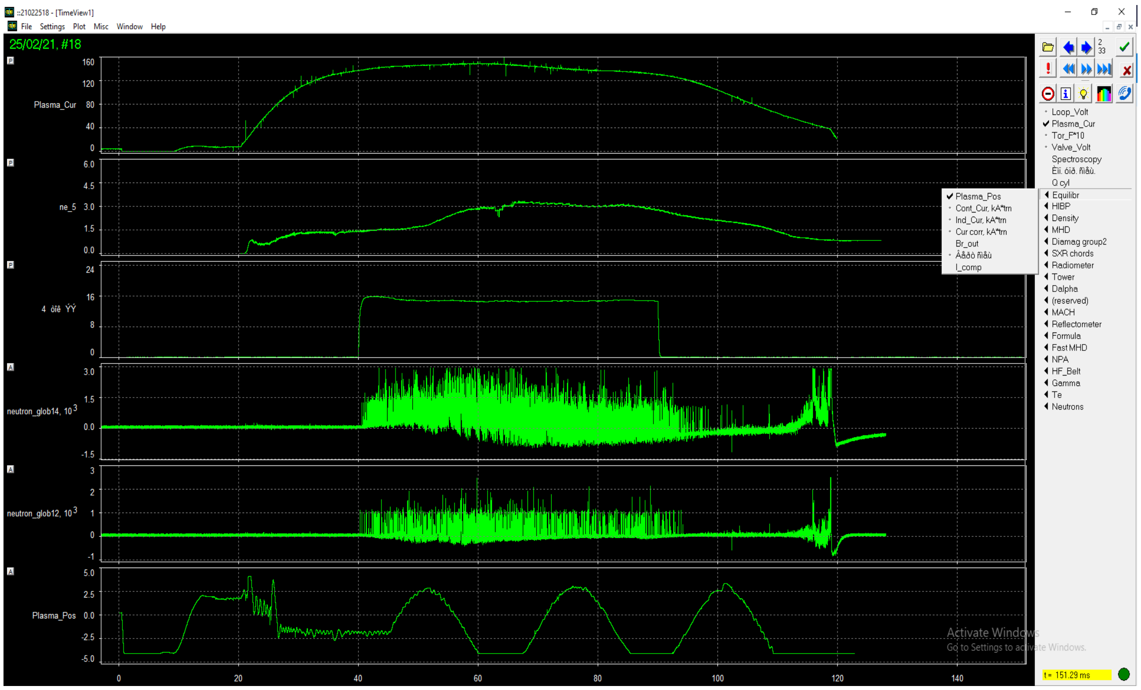

Рис. 1: Данные с датчиков

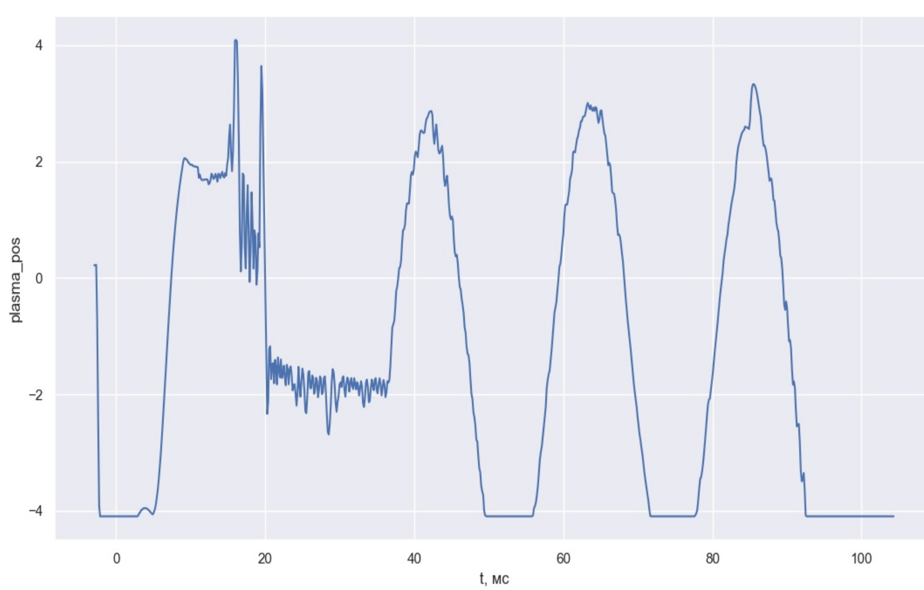

Рис. 2: Положение плазмы

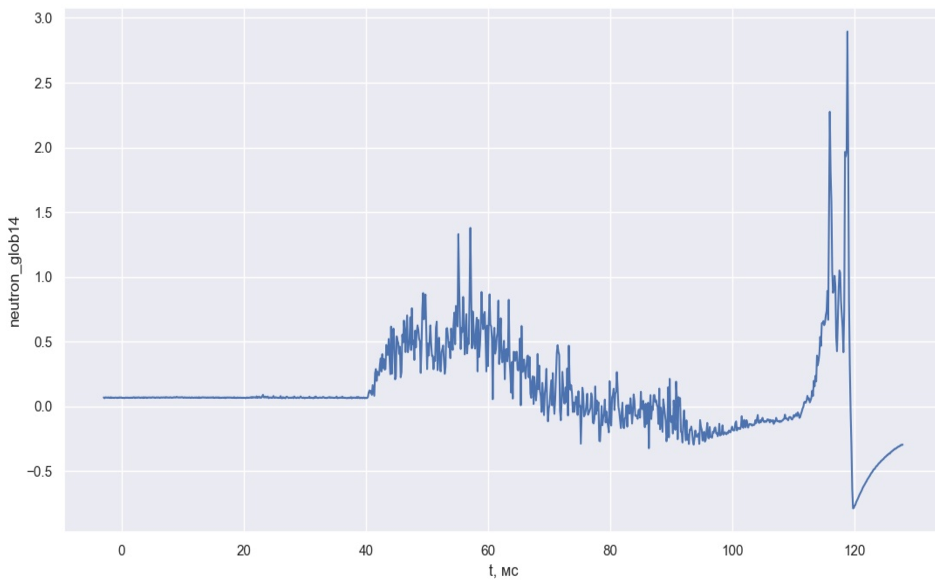

Рис. 3: Количество выделившихся нейтронов

Локализуем период активности:

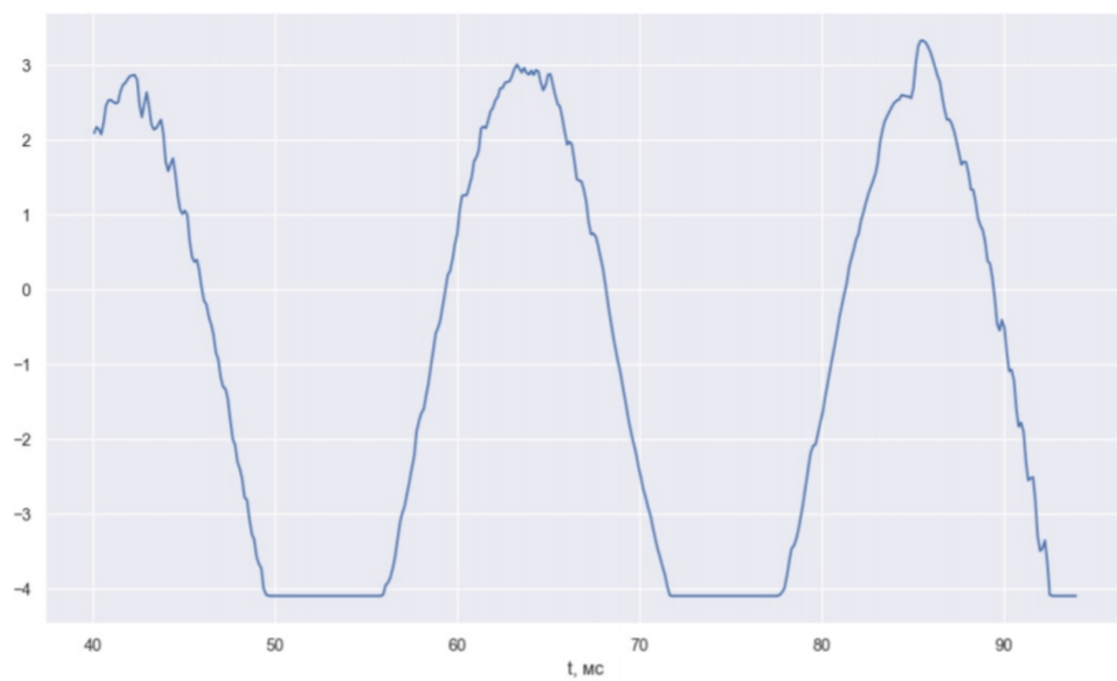

Рис. 4: Локалиация сигнала в периоде активности

Далее необходимо избавиться от выбросов. Воспользуемся медианным фильтром.

5.2 Построение функций корреляции

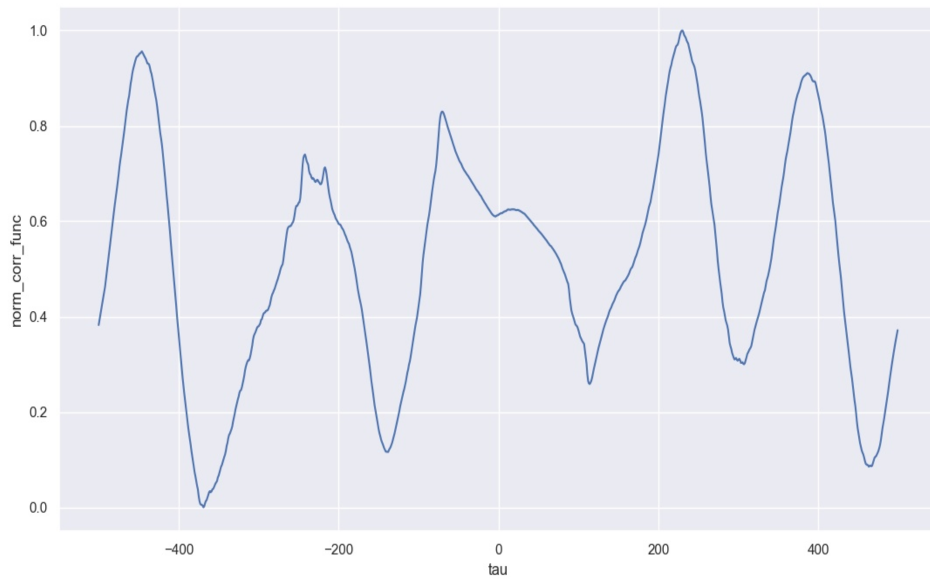

Рис. 5: Нормализованная функция взаимной корреляции между плазмой и нейтронами

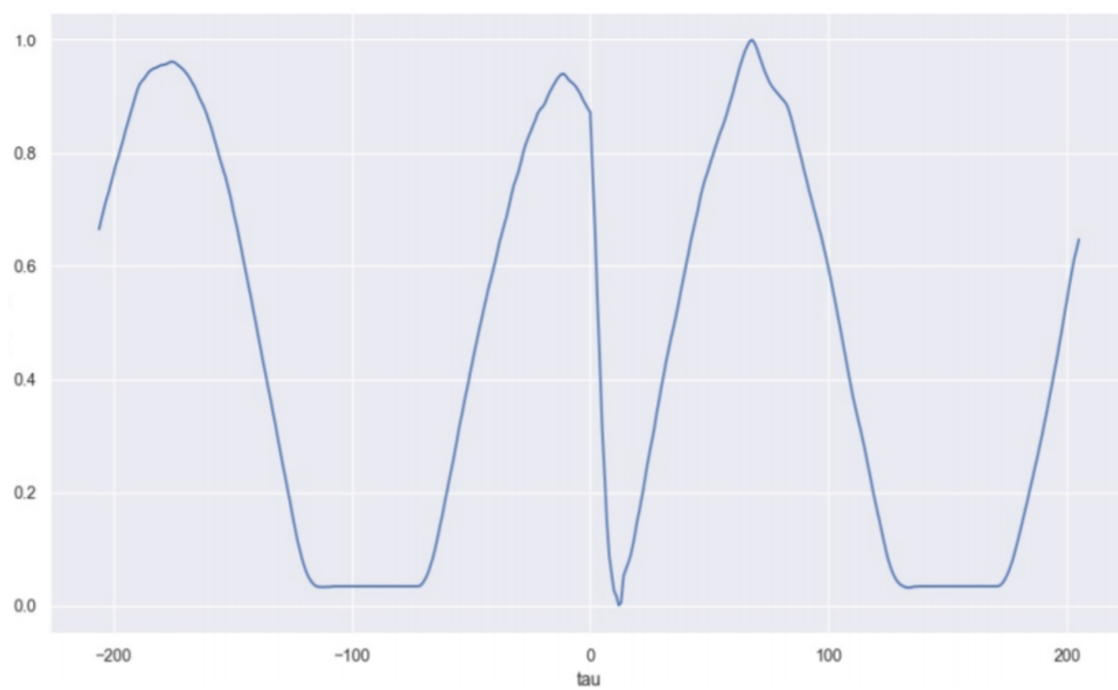

Рис. 6: Нормализованная функция взаимной корреляции между плазмой и нейтронами

6 Выводы

Из последнего графика можем наблюдать, что максимальных значений функция корелляции достигает при tau равных -176,-12, 69.

7 Приложения

Ссылка на репозиторий с кодом: https://github.com/HellInsider/MathStat

Список литературы

[1] Вероятностные разделы математики. Учебник для бакалавров технических направлений.//Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.

[2] Эконометрика: Учебник/Под ред. И.И. Елисеевой. — М.: Финансы и статистика, 2003. — 344 с.: ил.

[3] Beauchamp K. G. (1973) Signal Processing Using Analog and Digital Techniques. London: Allen and Unwin

[4] Электронный ресурс: http://www.williamspublishing.com/PDF/5-8459-0710-1/part.pdf