

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

КУРСОВАЯ РАБОТ

по дисциплине
«Математическая статистика»

Выполнила студентка
группы 3630102/80401

Мамаева Анастасия Сергеевна

Проверил
Доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург
2021

СОДЕРЖАНИЕ

СПИСОК ИЛЛЮСТРАЦИЙ	3
1 Постановка задачи	4
2 Теория	4
2.1 Описание корреляции	5
2.2 Взаимная корреляция и автокорреляция	6
3 Программная реализация	7
4 Ход работы	8
5 Результаты	9
5.1 Визуализация входных данных	9
5.2 Построение функций корреляции	10
6 Анализ	11
7 Приложение	11
Список литературы	12

СПИСОК ИЛЛЮСТРАЦИЙ

1	Схема ТОКАМАКа	4
2	Положение плазмы	9
3	Количество выделившихся нейтронов_14	9
4	Количество выделившихся нейтронов_12	10
5	Нормализованная функция взаимной корреляции между плазмой и нейтронами_14	10
6	Нормализованная функция взаимной корреляции между плазмой и нейтронами_12	11

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.

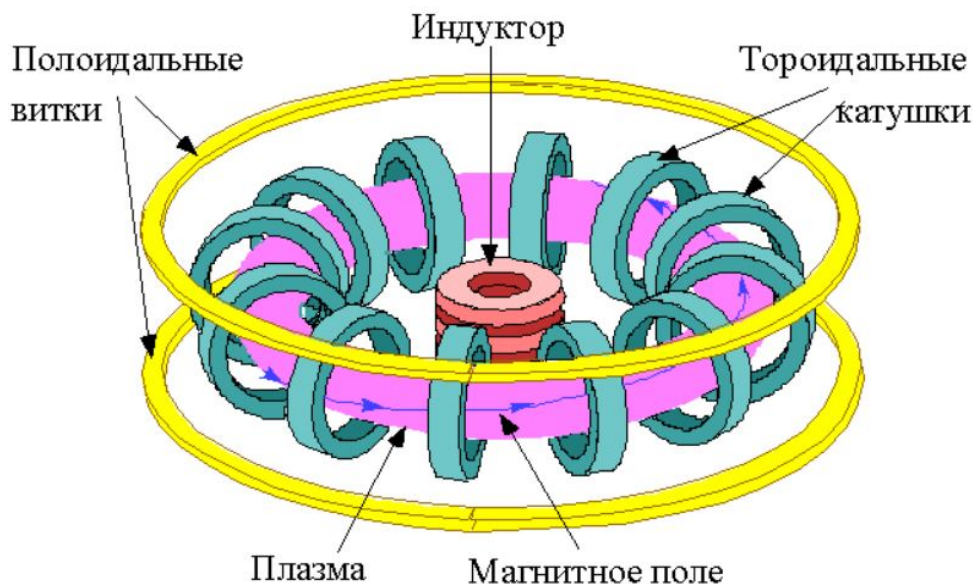


Рис. 1: Схема ТОКАМАКа

В нашей работе необходимо:

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

2 Теория

Часто бывает необходимо определить степень независимости одного процесса от другого или установить сходство одного набора данных с другими. Другими словами, искомой является функция корреляции процессов или данных, которую можно определить математически и измерить. Процесс корреляции занимает значимое место в обработке сигналов. Этот математический аппарат нашёл применение в обработке сигналов. Корреляция также является неотъемлемой частью процесса свёртки, который по сути, – та же корреляция двух последовательностей данных, при вычислении которой одна из последовательностей обращена во времени.

2.1 Описание корреляции

Рассмотрим, как можно сравнить две последовательности данных, состоящие из значений, одновременно выбираемых из двух соответствующих сигналов. Если два сигнала похоже меняются при переходе от точки к точке, то меру их корреляции можно вычислить, взяв сумму произведений соответствующих 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 организованы именно так. В непрерывной области $n \rightarrow t$ и $j \rightarrow \tau$ и

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

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 рассчитанные значения.

2.2 Взаимная корреляция и автокорреляция

Определение взаимной корреляции двух периодических последовательностей неравной длины требует аккуратности. Это объясняется тем, что результат корреляции будет повторяться с периодом более короткой последовательности. Этот результат не отражает полной периодичности более длинной последовательности, следовательно, неверен. Продemonстрируем это, найдя взаимную корреляцию $r_{ab}(j)$ последовательностей $a = \{4, 3, 1, 6\}$ и $b = \{5, 2, 3\}$. Последовательность b записывается под a и поэтапно смещается на одну позицию влево, в последнем столбце записываются соответствующие значения взаимной корреляции.

последовательность	задержка	$r_{ab}(j)$
4 3 1 6		
3 5 2 3	0	47
5 2 3 5	1	59
2 3 5 2	2	34
3 5 2 3	3	47
5 2 3 5	4	59

Результат показывает, что $r_{ab}(j)$ циклично с периодом в три задержки, т.е. период $r_{ab}(j)$ равен периоду более короткой последовательности b . Описанная процедура называется *циклической корреляцией*. Чтобы получить правильное значение, в котором каждое значение a усножается на каждое значение b , все элементы b нужно последовательно сместить под каждым значением a :

```

      4  3  1  6
        5  2  3
          5  2  3
            5  2  3
              5  2  3
                5  2  3
                  5  2  3

```

Видно, что для того, чтобы последовательность b стала повторяться, требуется 6 задержек. Длины последовательностей равны 4 и 3, а число требуемых задержек $- 4 + 3 - 1 = 6$. Таким образом, получаем общее правило нахождения линейной взаимной корреляции двух периодических последовательностей длины N_1 и N_2 : дополнить нулями каждую последовательность, чтобы их длины были равны $N_1 + N_2 - 1$ (т.е. добавить $N_2 - 1$ нулей к последовательности длиной N_1 и $N_1 - 1$ нулей к последовательности длиной N_2 .) Проиллюстрируем для указанных выше последовательностей:

последовательность	задержка	$r_{ab}(j)$
4 3 1 6 0 0		
5 2 3 0 0 0	0	29
2 3 0 0 0 5	1	17
3 0 0 0 5 2	2	12
0 0 0 5 2 3	3	30
0 0 5 2 3 0	4	17
0 5 2 3 0 0	5	35
5 2 3 0 0 0	6	29

Следовательно, искомая взаимная корреляция a и b равна

$$r_{ab}(j) = \{29, 17, 12, 30, 17, 35\}$$

3 Программная реализация

Код программы был написан на языке Python в среде разработки PyCharm. В ходе работы использовались дополнительные библиотеки:

1. seaborn (визуализация)
2. matplotlib (визуализация)
3. numpy (вычисление ряда числовых характеристик)

В приложении находится ссылка на GitHub репозиторий с исходным кодом.

4 Ход работы

На вход подаётся .txt файл со входными данными, полученными с датчиков ТОКАМАКа.

Первый этап работы – обработка информации. Необходимо считать построчно файл, убирая лишние знаки, пробелы и табуляции. Из получившегося массива создать 12 векторов, соответствующих снятым показаниям.

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

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

Четвёртый шаг – нормировка данных. Воспользуемся минмаксной нормализацией, где каждый элемент данных x вычисляется по формуле:

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

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

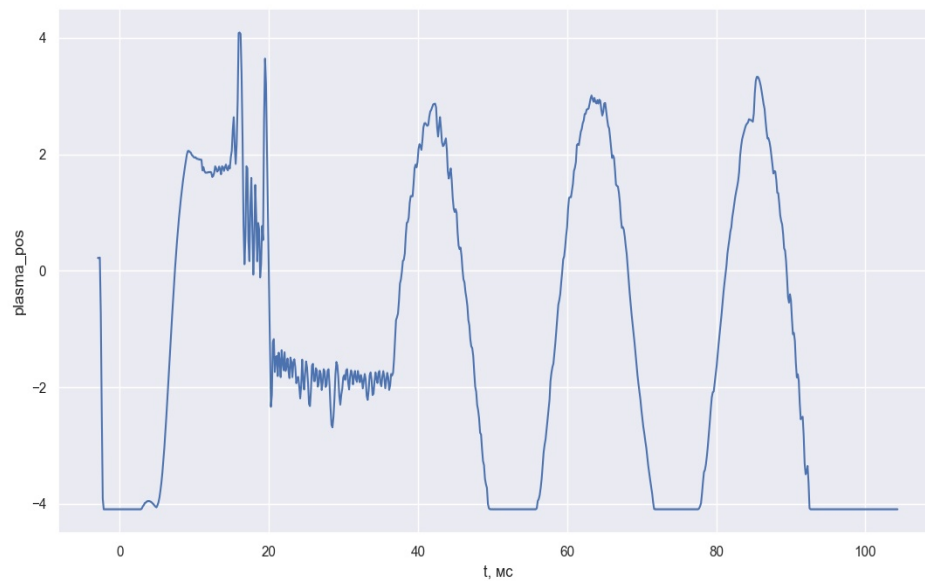


Рис. 2: Положение плазмы

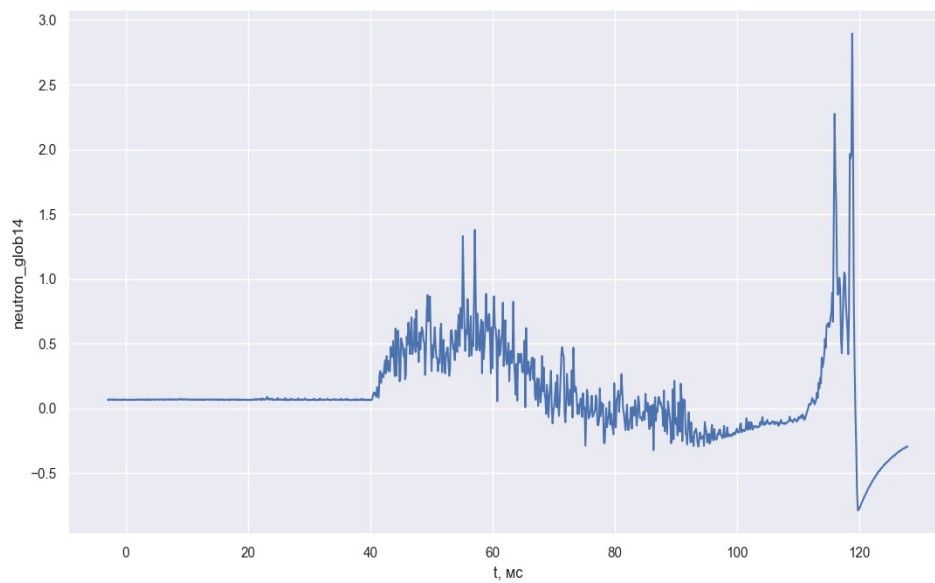


Рис. 3: Количество выделившихся нейтронов_14

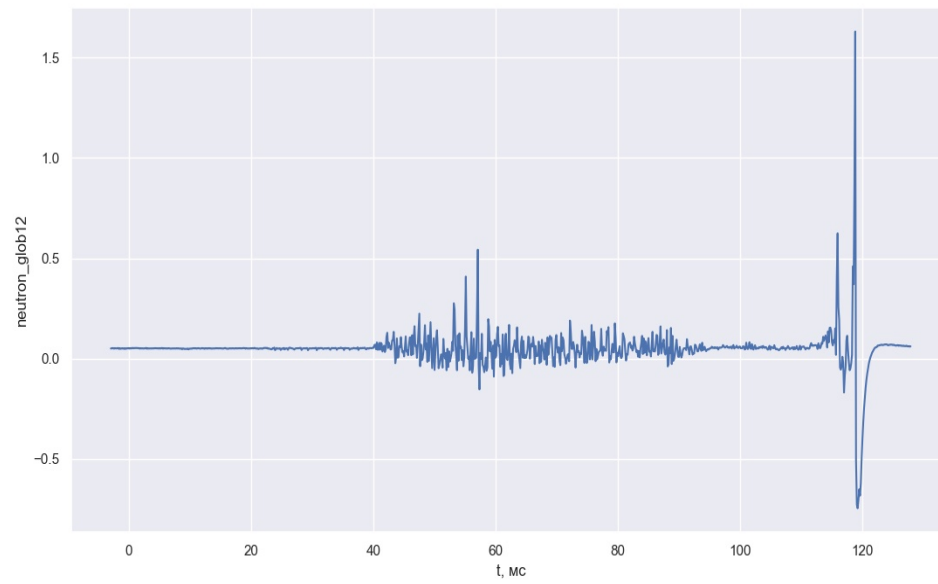


Рис. 4: Количество выделившихся нейтронов_12

5.2 Построение функций корреляции

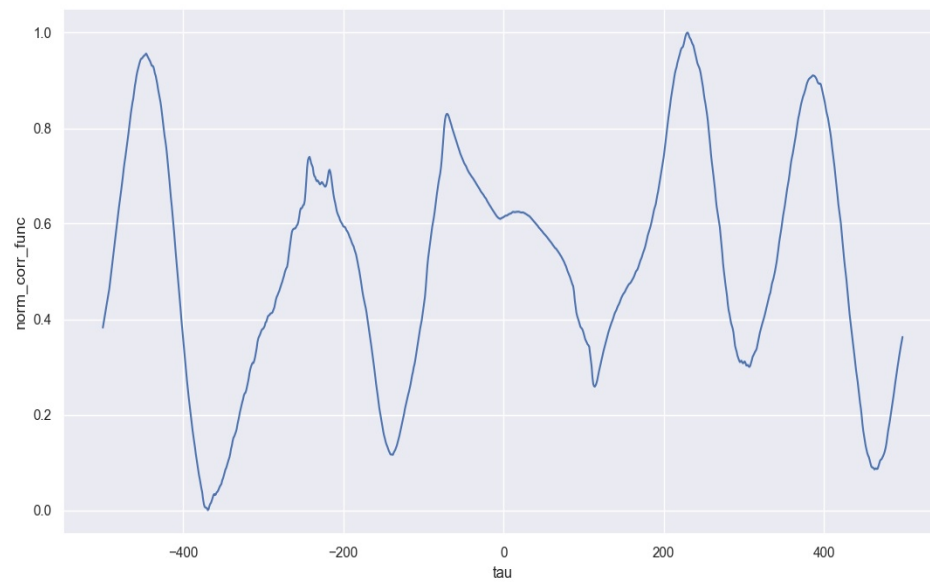


Рис. 5: Нормализованная функция взаимной корреляции между плазмой и нейтронами_14

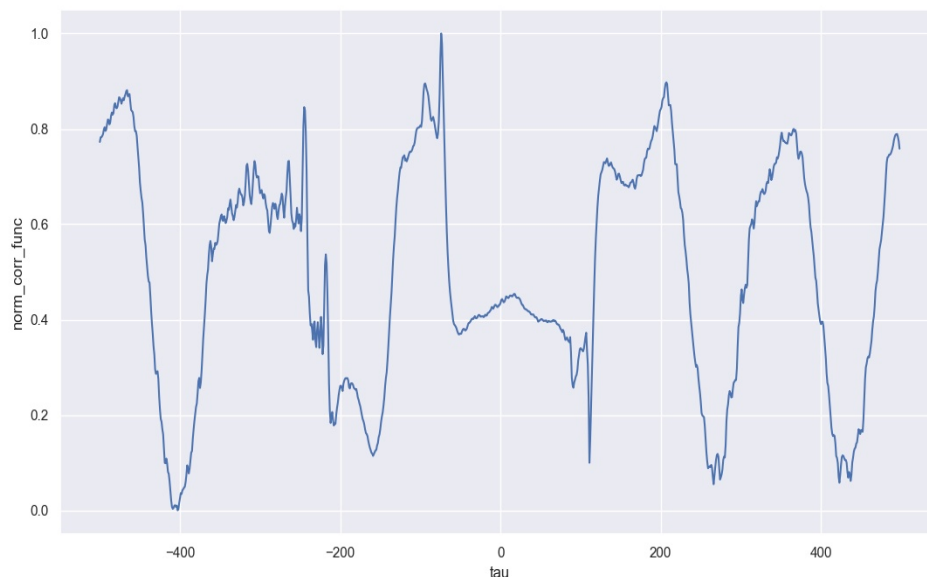


Рис. 6: Нормализованная функция взаимной корреляции между плазмой и нейтронами_12

6 Анализ

1. При изучении графика (5) получаем:

периодичность $\tau = 210$

основной период колебаний $T = \tau \cdot \Delta t = 210 \cdot 0.1311 \text{мс} = 27.531 \text{мс}$

частота колебаний $f = \frac{1}{T} = \frac{1}{27.531} \text{кГц} \approx 0.036 \text{кГц}$

2. Аналогичными выкладками для графика (6) получаем:

периодичность $\tau = -90$

основной период колебаний $T = \tau \cdot \Delta t = 90 \cdot 0.1311 \text{мс} = 11.799 \text{мс}$

частота колебаний $f = \frac{1}{T} = \frac{1}{11.799} \text{кГц} \approx 0.085 \text{кГц}$

7 Приложение

Код программы GitHub URL:

<https://github.com/Brightest-Sunshine/correlation-analysis-of-plasma/tree/master/scr>

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>