

## Предсказание эпилептического приступа с помощью нестационарного анализа временного ряда электроэнцефалограммы

М.С. Белова<sup>1</sup>, Ю.Н. Орлов<sup>2</sup>  
margarita.s.belova@phystech.edu

<sup>1</sup>Московский физико-технический институт (национальный исследовательский университет),

<sup>2</sup>Институт прикладной математики, г. Москва

*Общепринятый и хорошо известный подход при работе с нестационарными временными рядами — пытаться привести ряд к стационарному вычитанием тренда, поиском сезонности, построением авторегрессионной модели и т.п. Такой подход хорошо исследован и поэтому прост в работе; для стационарных рядов существует множество библиотек под различные языки программирования. Вместе с тем очевидно, что адаптация ряда под стационарные критерии даётся ценой потери его свойств и ухудшением качества прогноза. Данная работа развивает другой подход: работа с нестационарными рядами как таковыми, построение инструментов для описания их свойств и прогнозирования. Апробация метода осуществлена на временных рядах электроэнцефалограммы пациента НИИ нейрохирургии им. Бурденко с эпилепсией. В результате построен алгоритм на python3 для предсказания эпилептического приступа.*

**Ключевые слова:** нестационарный временной ряд, статистический анализ, электроэнцефалограмма, ЭЭГ, эпилепсия, python.

## Construction of an epileptic seizure predictor via non-stationary analysis of electroencephalogram time series

M.S. Belova<sup>1</sup>, U.N Orlov<sup>2</sup>

<sup>1</sup>Moscow Institute of physics and technology (National Research University),

<sup>2</sup>Institute of applied mathematics, Moscow

*Common approach to handle non-stationary time series is to turn them into stationary ones — subtracting trend, seasonality, building autoregression model etc. This approach is thoroughly investigated, consequently it is practically simple; there are many of libraries for stationary time series for different programming languages. Nevertheless, adapted to stationary criteria time series partly loses its properties; thus prognosis quality is lower. In this article we develop another approach: work with non-stationary time series as given, constructing tools to describe their attributes and forecasting. Method was tested on electroencephalogram time series of a Burdenko Institute for Neurosurgery patient who has epilepsy. As a result, an python3 algorithm for epilepsy seizure prediction is written.*

**Keywords:** non-stationary time series, statistic analysis, electroencephalogram, EEG, epilepsy, python.

### Введение

Эпилепсия — хроническое неврологическое расстройство вследствие аномальной электрической активности коры головного мозга и одно из наиболее распространённых неврологических заболеваний. Более чем в половине случаев причины заболеваемости установить не удаётся. Согласно информационному бюллетеню Всемирной организации здравоохранения от 8.02.2018 на 1000 человек приходится от 4 до 10 больных активной формой эпилепсии [9]. Расстройство влечёт за собой опасные для жизни состояния: потерю сознания, судороги, травмы во время припадков вплоть до смертельных исходов. Лекарственный контроль

состояния имеет свои противопоказания и ограничения [2]. Для порядка 2/3 пациентов фармакологическое лечение даёт результат. В ещё 10% случаях помогает хирургическая операция. Для 25% эти опции неприменимы. В случаях, когда фармакологическое лечение оказывается неэффективным, а хирургическое вмешательство противопоказано, для своевременной помощи необходимо предсказать приступ заранее. Отдельный интерес представляет собой создание электронных устройств, способных детектировать признаки надвигающегося приступа, а также вычислительных алгоритмов для этих устройств.

Основным методом наблюдения пациента, ведения состояния и диагностики является энцефалография [4]. В области изучения эпилепсии энцефалограмма была и остаётся одним из наиболее ценных источников информации, хотя параллельно активно развиваются другие аппаратные средства и методики (см. [6]; [1]; [8]). Судорожный приступ на энцефалографической записи позволяет однозначно поставить диагноз «эпилепсия» (любые остальные отклонения ЭЭГ могут служить только вспомогательным средством для диагноза).

## Основные термины и определения

**Выборочная плотность функции распределения (эмпирическое распределение вероятностей)** — совокупность относительных частот  $n_i(T)$  встречаемости соответствующих значений  $x_i$  случайной величины  $\xi$  в выборке объёма  $T$ :

$$\{\nu_i(T)\}, \quad \nu_i(T) = \frac{n_i(T)}{T}, \quad i = \overline{1, n}$$

**Выборочная функция распределения** (эмпирическая функция распределения) — ступенчатая неубывающая функция  $F_T(x)$ , определяемая по эмпирическому распределению частот:

$$F(x, T) = \sum_{i=1}^k \nu_i(T)$$

## Содержание работы

### Построение индикатора приступа

Для случайной величины  $\xi$  доверительный интервал  $(\theta^* - \delta; \theta^* + \delta)$  уровня значимости  $\alpha$  определяется как

$$P(\theta^* - \delta < \xi < \theta^* + \delta) = 1 - \alpha, \quad (1)$$

где  $1 - \alpha$  — уровень доверия (или уровень достоверности). Тогда вероятность ошибки первого рода (отбросить верную гипотезу, что случайная величина принадлежит заданному распределению) равна уровню

значимости  $\alpha$ . Повышая уровень значимости, мы уменьшаем доверительный интервал; а значит, точнее локализуем значение  $\theta^*$ ; взамен растёт вероятность ошибки первого рода. Взяв в качестве случайной величины расстояние  $\rho$  между выборочными функциями распределения двух встык-выборок длины  $N$  временного ряда, поставим задачу выбрать оптимальный уровень значимости  $\alpha^*$ .

Известно, что принадлежность двух выборок размера  $N$  с непрерывными эмпирическими функциями распределения  $F_1(x)$ ,  $F_2(x)$  одному и тому же распределению определяется при помощи двухсторонней двухвыборочной статистики Смирнова [7]:

$$\begin{aligned} D_N &= \sup_{-\infty < x < \infty} |F_1(x) - F_2(x)|, \\ \lim_{N \rightarrow \infty} P \left\{ 0 < \sqrt{\frac{N}{2}} D_N < z \right\} &= \quad (2) \\ &= K(z) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2 z^2). \end{aligned}$$

Из (18) получаем, что для уровня значимости  $1 - K \left( \sqrt{\frac{N}{2}} \varepsilon \right)$  доверительный интервал для расстояния<sup>1</sup>  $D_N$  равен  $(0; \varepsilon)$ :

$$\lim_{N \rightarrow \infty} P \{ 0 < D_N < \varepsilon \} = K \left( \sqrt{\frac{N}{2}} \varepsilon \right). \quad (3)$$

Так как ВФР расстояний между выборками  $G(\rho)$  непрерывна и монотонно не убывает на  $[0; 1]$ , функция Колмогорова  $K(z)$  монотонно возрастает от единицы до нуля, существует единственное решение уравнения

$$\varepsilon = 1 - K \left( \sqrt{\frac{N}{2}} \varepsilon \right). \quad (4)$$

Поскольку практическая точность критерия принятия гипотезы равна наибольшему из двух чисел: уровню значимости и до-

<sup>1</sup>На практике статистика Смирнова применима при  $N \geq 50$  (см. [3], стр. 215)

ле меры доверительного интервала в мере множества допустимых исходов [5], решение  $\varepsilon^*$  уравнения (4) согласовывает уровень значимости и доверительный интервал. Для стационарных распределений решение  $\varepsilon^*$  не зависит от вида распределения.

Назовём временной ряд  $\xi_t$  значений случайной величины  $\xi$  стационарным (в широком смысле), если матожидание и дисперсия

$$\begin{aligned}\mu_\xi &= E[\xi] = \sum_i \xi_i p_i, \\ \sigma^2[\xi] &= \sum_i (\xi_i - E[\xi])^2 p_i\end{aligned}\quad (5)$$

не меняются со временем. Статистика Смирнова и приведённые выше выкладки применимы для стационарных рядов. Для нестационарных ВФР решение (4) всегда отличается в большую сторону. Для работы с нестационарными ВФР запишем аналогичное по форме уравнение

$$G_N(\rho) = 1 - \rho \quad (6)$$

и назовём его численное решение  $\rho^*(N)$  согласованным уровнем стационарности (СУС) — вероятность того, что расстояние между выборками длины  $N$  больше  $\rho^*$ . Предлагаемый в работе индикатор эпилептического приступа — СУС временного ряда ЭЭГ.

## Алгоритм построения предиктора приступа эпилепсии

### Общее описание

Ряд ЭЭГ делится на статистические окна. По каждому статистическому окну алгоритм проходит двумя меньшими встык-окнами, между которыми считает расстояние в  $C_1$  норме и строит ВФР  $G(\rho)$  расстояний между встык-выборками внутри статистического окна. После чего численно решается уравнение (6). Результат работы программы — ряд СУС и ВФР ряда СУС. Длина статистического окна выбирается из соображений, что за это время состояние больного не успевает сильно измениться (слишком длинное окно), при этом

```
Time,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,
8.843750,-79.301479,65.757277,10.879224,-21.913759,-58.387587,12.217713,-3
8.847656,-87.499724,43.337584,-8.863490,-39.648739,-84.655435,4.019467,-13
8.851562,-123.806241,26.104537,-30.781249,-69.597433,-120.961951,-8.696179
8.855469,-148.066356,9.708046,-38.812184,-85.659302,-131.167931,-16.894425
8.859375,-145.556688,-11.707780,-47.010429,-94.526792,-130.833309,-26.7657
8.863281,-148.735600,-22.750315,-61.566498,-103.226971,-134.681465,-34.964
8.867188,-154.089556,-41.154540,-77.461056,-107.242438,-129.494819,-43.496
8.871094,-134.514154,-62.737676,-81.141901,-96.534526,-110.755972,-41.6564
8.875000,-121.798507,-73.278278,-70.433988,-80.472657,-83.316946,-35.13133
8.878906,-125.479352,-65.749277,-48.516229,-56.045231,-51.695141,-24.42342
8.882812,-127.989019,-42.158406,-32.788983,-32.454360,-30.112005,-14.38475
8.886719,-132.339109,-16.392491,-32.119738,-27.936960,-26.598471,-12.20971
8.890625,-129.160197,6.194512,-32.119738,-29.275449,-38.310250,-12.544335,
```

Рис. 1: Часть файла входных данных

данных должно быть достаточно для набора статистики. Для предсказания приступа по ряду ЭЭГ рекомендуемое время порядка 2 минут. Для тестируемого входного файла это 30тыс. точек. Длина каждого скользящего окна по 5000 точек (порядка 20 сек.). Для предсказания приступа следует наблюдать индикативный электрод — электрод с наиболее выраженными изменениями СУС перед приступом по сравнению с нормальным состоянием. Причина в том, что эпилептический очаг может локализоваться возле одного-нескольких электродов, тогда изменения на остальных будут невыраженными.

## Структура алгоритма

В результате работы написан алгоритм на языке python3.7.2 с использованием библиотек numpy1.16.2, matplotlib3.0.3, scipy1.2.0. Входные данные — файл формата .txt параллельной записи 2340000 значений напряжений с 64-х электродов с частотой дискретизации 256 Гц (длительность записи 9140с), вживлённых в голову пациента (см. рис 1). Точность входных значений —  $10^{-6}$ . Размер файла 1,6 Гб. Нормальное состояние пациента в начале файла переходит в эпилептический приступ в конце. Файл получен из практики НИИ нейрохирургии им. Бурденко. Программа реализована в среде разработки Pycharm 2019.1.1 Community Edition, OS Ubuntu 18.04.2 LTS 64bit, количество строк кода ~500. Время работы программы на процессоре Intel Core i7-8550U CPU 1.80GHz × 8, 8 Gib RAM порядка 2970 сек.

**Чтение из файла:** Из файла ЭЭГ считываются показания выбранного элект-

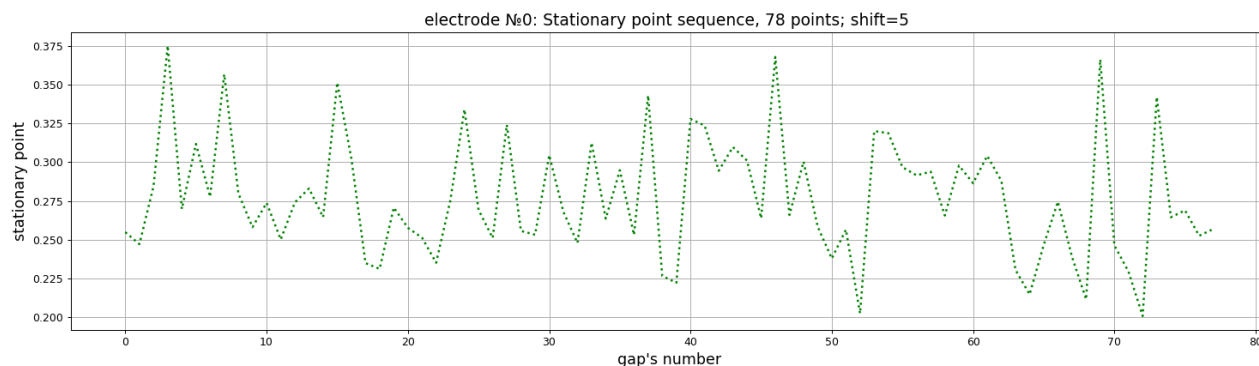


Рис. 2: Ряд значений СУС для электрода №0

трода длиной 30 тыс. точек — статистическое окно.

**Набор статистики:** Программа проходит по статистическому окну двумя встык-выборками длиной 5 тыс. точек, смещая их на 5 точек вправо. На каждом шаге смещения строятся ВФР встык-выборок с расчётом оптимального числа интервалов разбиения для численной аппроксимации ВФР. Вычисляется расстояние между ВФР в  $C_1$  норме. Шаг смещения и длины встык-выборок подбираются так, чтобы статистическое окно было прочитано полностью вместе с конечными значениями.

**Расчёт СУС:** После того как программа прошла по статистическому окну, строится ВФР полученного ряда расстояний между встык-выборками. Численно решается уравнение (6), полученное значение СУС сохраняется в массив.

**Повтор:** Повторяем вновь с пункта «Чтение из файла», продолжаем, пока ряд напряжений с электрода не будет считан полностью.

**Обработка результата:** Строится график поведения полученного в результате предыдущих шагов ряда СУС (см. рис. 2). Строится ВПФР ряда СУС. (см. рис. 3).

Полученный ряд значений СУС вновь проверяется на стационарность. Поскольку ряд содержит только 78 точек, построение автокорреляционной функции не будет иллюстративным. Поэтому воспользуемся следу-

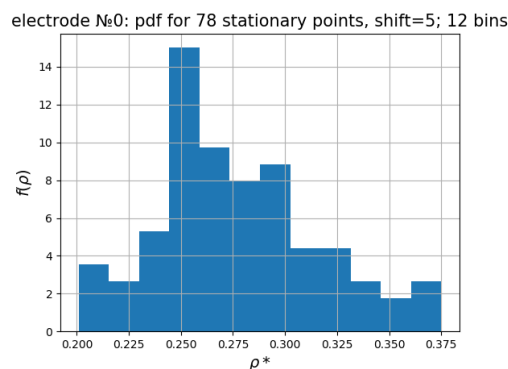


Рис. 3: ВПФР ряда значений СУС для электрода №0

ющим способом: поделим выборку пополам, построим для ВФР для обеих частей и посчитаем расстояние между ними в норме  $L_1$ . Для стационарных распределений расстояние  $\rho_{L1} < 0.4$ . Для сравнения, возьмём также 78 точек исходного ряда ЭЭГ и проведём этот же тест:

$$\rho(\text{СУС})_{L1} = 0.0116 \quad (7)$$

$$\rho(\text{ЭЭГ})_{L1} = 100 \quad (8)$$

Отсюда следует, что ряд СУС является стационарным. Запуск алгоритма для других электродов даёт аналогичный результат.

Таким образом, получен результат: у нестационарного ряда ЭЭГ найдена стационарно распределённая статистика. Перед эпилептическим приступом за время порядка двух минут на индикативном электроде наблюдается выраженный скачок предложенного индикатора тревоги. В дальнейшем предполагается проверка алгоритма на электроэнцефалограммах других пациентов НИИ

нейрохирургии им. Бурденко и сбор статистики по количеству положительных, ложноположительных и ложноотрицательных тревог предложенного индикатора.

[10] Стационарная точка уровня значимости для нестационарных функций распределения / А.А.Кислицын [и др.] // Препринты ИПМ им. М.В.Келдыша. 2018. № 113. 20 с. doi:10.20948/prep.

## Литература

- [1] A. V. de Vel, K. Cuppens, B. Bonroy, M. Milosevic, K. Jansen, S. V. Huffel, B. Vanrumste, L. Lagae, and B. Ceulemans. Non-eeG seizure-detection systems and potential sudep prevention: State of the art. *Seizure*, 22(5):345 – 355, 2013. ISSN 1059-1311. doi: <https://doi.org/10.1016/j.seizure.2013.02.012>. URL <http://www.sciencedirect.com/science/article/pii/S1059131113000575>.
- [2] C. Deckers, P. Genton, G. Sills, and D. Schmidt. Current limitations of antiepileptic drug therapy: a conference review. *Epilepsy Research*, 53(1):1 – 17, 2003. ISSN 0920-1211. doi: [https://doi.org/10.1016/S0920-1211\(02\)00257-7](https://doi.org/10.1016/S0920-1211(02)00257-7). URL <http://www.sciencedirect.com/science/article/pii/S0920121102002577>.
- [3] А. И. Кобзарь. *Прикладная математическая статистика. Для инженеров и научных работников*. М.: ФИЗМАТЛИТ. - 816с., 2006. ISBN 5-9221-0707-0.
- [4] D. Longo, A. Fauci, D. Kasper, S. Hauser, J. Jameson, and J. Loscalzo. *Harrison's Principles of Internal Medicine, 18th Edition*. McGraw-hill, 2011. ISBN 9780071748896. URL <https://books.google.ru/books?id=9uffAgpvgmEC>.
- [5] Ю. Н. Орлов. *Кинетические методы исследования нестационарных временных рядов*. Москва: МФТИ. - 276 с., 2014. ISBN 978-5-7417-0521-6.
- [6] S. Ramgopal, S. Thome-Souza, M. Jackson, N. E. Kadish, I. S. Fernández, J. Klehm, W. Bosl, C. Reinsberger, S. Schachter, and T. Loddenkemper. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy & Behavior*, 37:291 – 307, 2014. ISSN 1525-5050. doi: <https://doi.org/10.1016/j.yebeh.2014.06.023>. URL <http://www.sciencedirect.com/science/article/pii/S1525505014002297>.
- [7] Н. В. Смирнов. Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках. *Бюллетень Московского государственного университета*. Математика, II(2):3 – 14, 1939.
- [8] A. Ulate-Campos, F. Coughlin, M. Gainza-Lein, I. S. Fernández, P. Pearl, and T. Loddenkemper. Automated seizure detection systems and their effectiveness for each type of seizure. *Seizure*, 40:88 – 101, 2016. ISSN 1059-1311. doi: <https://doi.org/10.1016/j.seizure.2016.06.008>. URL <http://www.sciencedirect.com/science/article/pii/S1059131116300711>.
- [9] World Health Organization. Epilepsy (bulletin). <https://www.who.int/news-room/fact-sheets/detail/epilepsy>, 02 2018. Online; accessed 07 January 2019.