ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 113 за 2018 г.

ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Кислицын А.А., Козлова А.Б., Корсакова М.Б., Машеров Е.Л., Орлов Ю.Н.

Стационарная точка уровня значимости для нестационарных функций распределения

Рекомендуемая форма библиографической ссылки: 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/prepr-2018-113
URL: http://library.keldysh.ru/preprint.asp?id=2018-113

Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук

А.А. Кислицын, А.Б. Козлова, М.Б. Корсакова,
Е.Л. Машеров, Ю.Н. Орлов

Стационарная точка уровня значимости для нестационарных функций распределения

Москва — 2018

Кислицын А.А., Козлова А.Б., Корсакова М.Б., Машеров Е.Л., Орлов Ю.Н.

Стационарная точка уровня значимости для нестационарных функций распределения

В работе исследуются свойства статистики, называемой согласованным уровнем стационарности нестационарных временных рядов. Рассматриваются такие задачи как оценка уровня значимости принятия решений, выделение характерных периодов в нестационарном случайном процессе, определение оптимальной длины выборки для построения индикаторов в задачах стохастического управления. В качестве примера строится индикатор разладки локального перехода случайного процесса из одного нестационарного режима в другой. Алгоритм тестируется на данных электроэнцефалограмм для распознавания приступа эпилепсии.

Ключевые слова: стационарная точка, уровень значимости, индикатор разладки, электроэнцефалограмма, приступ эпилепсии

Kislitsyn A.A., Kozlova A.B., Korsakova M.B., Masherov E.L., Orlov Yu.N.

Stationary point of significance level for non-stationary distribution functions

Self-consistent stationary level of non-stationary time series is investigated. The various purposes of practical use of this statistics are discussed. The main problem is an evaluation of significance level of any sample statistics. Unlike the classical problem of stationary test of two samples, in the case under consideration one should construct an indicator to predict the change in the non-stationary regime. For example, we consider special predictor of an attack of epilepsy.

Key words: stationary point, significance level, disorder indicator, electroencephalogram, epilepsy attack

Содержание

1. Согласованный уровень стационарности 3
2. Преобразование эквивалентности по Чернову 6
3. Структура вычислительного алгоритма 8
4. Стационарная точка уровня значимости для ЭЭГ 13
5. СУС как индикатор разладки 18
Литература 20

1. Согласованный уровень стационарности

В настоящей работе исследуется статистика, называемая согласованным уровнем стационарности (СУС) временного ряда [1, 2]. Смысл статистики СУС состоит в том, что для выборок определенной длины из стационарного или из нестационарного ряда СУС равен уровню значимости, на котором две выборки признаются взятыми из одного распределения. Соответствующий одномерный численный код, позволяющий проводить анализ СУС в зависимости от длины выборки, строить функцию распределения расстояний между выборочными распределениями и генерировать ансамбль траекторий, имеющих нестационарную функцию распределения, был разработан в [3, 4]. Здесь мы опишем численную процедуру конструирования индикатора разладки при анализе нестационарного случайного процесса.

Как известно [5, 6], решение задачи о принадлежности двух выборочных функций распределения (ВФР) одной генеральной совокупности использует непараметрический критерий Колмогорова-Смирнова. Строится статистика

$$S_N = \sup_x \left| F_{1,N}(x) - F_{2,N}(x) \right|, \tag{1}$$

для которой справедлива асимптотика

$$\lim_{N\to\infty} P\left\{ 0 < \sqrt{\frac{N}{2}}\, S_N < z \right\} = K(z), \tag{2}$$

где $K(z)$ есть табулированная функция Колмогорова, а N есть длина выборки. Неубывающая непрерывная функция $F_N(x)$ есть ВФР случайной величины ξ, значения которой x наблюдаются в виде последовательности событий во временном окне длины N. В формуле (2) уровень значимости Q, на котором принимается решение, считается равным $1 - K(z)$.

Пусть ε есть расстояние между выборками в норме С, определяемое формулой (1). Рассмотрим стационарную точку уровня значимости (т.е. СУС):

$$Q(\varepsilon) \equiv 1 - K\left(\sqrt{\frac{N}{2}}\varepsilon \right) = \varepsilon. \tag{3}$$

Для стационарных распределений положение стационарной точки $\varepsilon = \varepsilon_0(N)$ как решения уравнения (3) не зависит от вида распределения F. Это решение единственно в силу монотонности функции $K(z)$. Примеры решений $\varepsilon_0(N)$ для различных N содержатся в [1]. Приведем здесь более подробную таблицу (табл. 1) стационарных точек уровней значимости распределения расстояний между выборками из стационарных распределений. График поведения $\varepsilon_0(N)$ показан на рис. 1.

Табл. 1. СУС для стационарных распределений

Длина выборки N	Уровень значимости $\varepsilon_0(N)$	Длина выборки N	Уровень значимости $\varepsilon_0(N)$
100	0,18	5000	0,0291
200	0,14	5500	0,0280
300	0,10	6000	0,0268
400	0,09	6500	0,0259
500	0,081	7000	0,0250
600	0,0799	7500	0,0243
700	0,0747	8000	0,0236
800	0,0696	8500	0,0230
900	0,0644	9000	0,0224
1000	0,0593	9500	0,0219
1100	0,0576	10000	0,0214
1200	0,0560	11000	0,02043
1300	0,0543	12000	0,01962
1400	0,0527	13000	0,01892
1500	0,0510	14000	0,01829
1600	0,0495	15000	0,01772
1700	0,0481	16000	0,01722
1800	0,0466	17000	0,01676
1900	0,0451	18000	0,01634
2000	0,0436	19000	0,01593
2500	0,0401	20000	0,01556
3000	0,0364	30000	0,01293
3500	0,0340	40000	0,01132
4000	0,0322	50000	0,01023
4500	0,0305	60000	0,00941

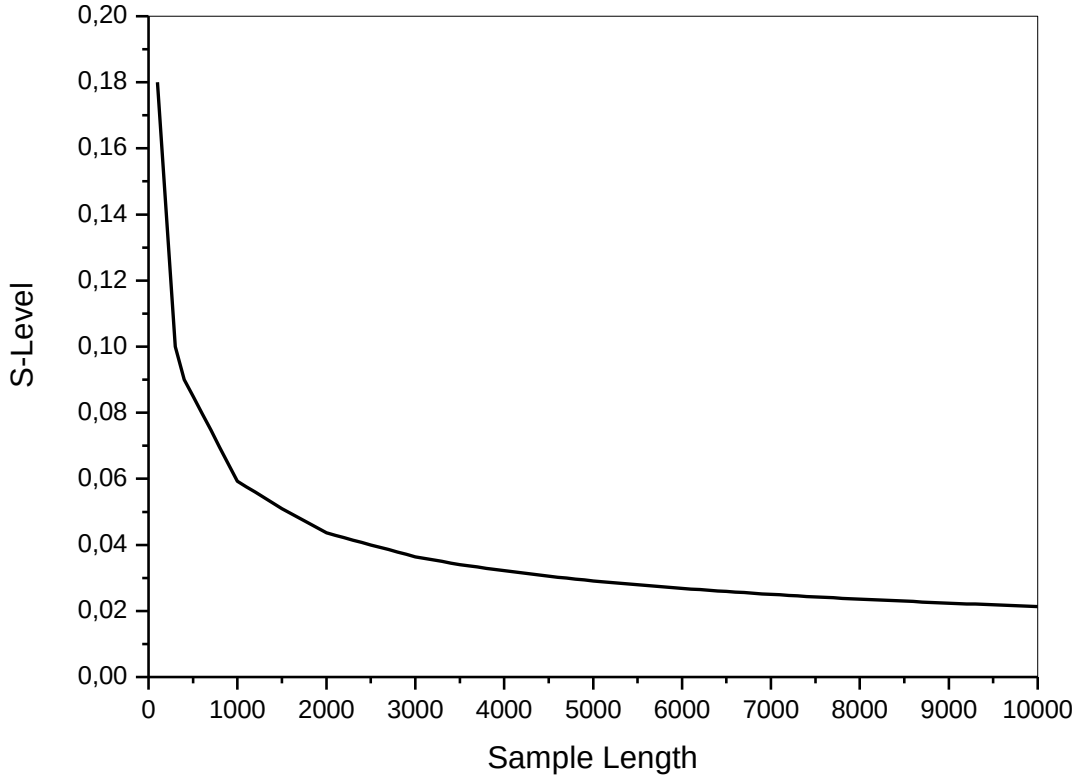

Рис. 1. СУС для стационарных распределений

Для нестационарных ВФР распределение расстояний между выборками заданной длины отличается от статистики (2) в большую сторону. Обозначим через $G_N(\rho)$ эмпирическую функцию распределения расстояний $\rho(N)$ между двумя непересекающимися выборками длины N:

$$\rho(N) = \left\| F_{1,N}(x) - F_{2,N}(x) \right\|_C . \tag{4}$$

Численное решение относительно ρ уравнения

$$G_N(\rho) = 1 - \rho \tag{5}$$

дает значение СУС $\rho^*(N)$ как вероятности того, что расстояние между выборками длины N больше ρ^*. Если оказалось, что $\rho^*(N) > \varepsilon_0(N)$, то ВФР считается нестационарной. Величина $\rho^*(N)$ есть уровень значимости, на котором можно принимать решение об однородности выборок.

Индексом нестационарности будем называть отношение

$$J(N) = \frac{\rho^*(N)}{\varepsilon_0(N)} . \tag{6}$$

Если считать, что индекс нестационарности представляет собой характеристику данного случайного процесса, то изменение его можно было бы

трактовать как разладку. Практический вопрос, возникающий в таком случае, состоит в длине окна, где считается индекс.

Предположим, что по некоторой длине L_{tot} фактически известного фрагмента временного ряда построен СУС как решение уравнения (5) для встык-выборок длины N. Следует учесть, что $L_{tot} >> 2N$, так что на длине L_{tot} укладывается достаточно много длин N, чтобы можно было набрать статистику для построения распределения $G_N(\rho)$. На этой длине L_{tot} в доле случаев $\rho^*(N)$ расстояние между встык-выборками больше, чем $\rho^*(N)$. Если теперь вместо L_{tot} рассмотреть две половины того же промежутка, то в каждой из них для встык-выборок длины N будет своя доля превышений уровня $\rho^*(N)$, например, $\rho_1(N)$ и $\rho_2(N)$. При этом, вообще говоря, $\rho^*(N) \neq \dfrac{\rho_1(N)+\rho_2(N)}{2}$. Если окажется, что $\rho_1(N) \neq \rho_2(N)$, то в той половине, где доля превышений больше, можно было бы ошибочно сделать вывод о наступлении разладки в случае, если на самом деле состояние изучаемой системы в статистическом смысле не изменилось. Насколько больше может быть превышение СУС на подмножестве расчетной области, при котором следовало бы отклонить ложную гипотезу о разладке? Иными словами, можно ли считать СУС нестационарного ряда случайной величиной, имеющей стационарное распределение, и какой квантиль этого распределения избрать в качестве индикатора разладки в окне длины L?

В последующих разделах работы мы исследуем зависимость распределения СУС от величины промежутка, на котором он вычислялся для встык-выборок определенной длины.

2. Преобразование эквивалентности по Чернову

Исследуя локальное значение СУС – то есть значение, полученное по промежутку времени L, существенно меньшему, чем первоначальный объем данных L_{tot}, естественно возникает вопрос о флуктуациях этого «локального» СУС относительно базового значения, отвечающего множеству в целом. Важно понимать, что СУС всего множества $\rho^*_{tot}(N)$ не является средним значением последовательности СУС $\rho^*_n(N,L)$, построенной для встык-выборок длины N в окнах длины L.

Для теоретического изучения поведения СУС по совокупности множеств используем следующий метод. Будем предполагать, что эмпирически построенная функция распределения $G_N(\rho)$ расстояний между встык-выборками является приближенной оценкой некоторой генеральной совокупности с непрерывной плотностью распределения, которую, собственно, мы и анализируем. Рассмотрим функцию

$$\Psi_N(\rho) = 1 - G_N(\rho). \qquad (7)$$

Поскольку $G_N(\rho)$ монотонно возрастает от 0 до 1, функция $\Psi_N(\rho)$ обладает свойствами: $\Psi_N(0)=1, \Psi'_N(0)=-a_N \le 0$. В таком случае существует предел

$$\lim_{n\to\infty}\left(\Psi_N\left(\frac{\rho}{n}\right)\right)^n = e^{-\rho a_N} \equiv \Phi_N(\rho). \tag{8}$$

Следуя определениям, данным в [7-9] для операторных функций, будем называть предельную функцию $\Phi_N(\rho)$ эквивалентной по Чернову функции $\Psi_N(\rho)$ (то есть уровню значимости распределения $G_N(\rho)$). Будем обозначать эту эквивалентность как $\Phi_N(\rho) \overset{Ch}{\propto} \Psi_N(\rho)$. Очевидно, $\left|\Phi_N(\rho)-\Psi_N(\rho)\right| = o(\rho)$.

Согласно [8], если имеется (конечный или бесконечный) набор функций $\Psi_N^k(\rho)$ вида (7), каждая из которых эквивалентна в смысле (8) функции $\Phi_N^k(\rho)$ с коэффициентом $-a_N^k$ в показателе экспоненты, и если задан соответствующий набор неотрицательных коэффициентов p_k таких, что $\sum_k p_k = 1$, то средняя функция $\overline{\Psi}_N(\rho) = \sum_k p_k \Psi_N^k(\rho)$ эквивалентна в смысле (8) функции $\Phi_N(\rho) = e^{-\rho\overline{a}_N}$, где

$$\overline{a}_N = \sum_k p_k a_N^k . \tag{9}$$

Рассмотрим последовательность непересекающихся промежутков длины L. Для каждого k-го промежутка построим эмпирическое распределение $G_N^k(\rho, L)$ расстояний между встык-выборками длины N. Пусть таких промежутков n, так что $nL = L_{tot}$ - полная длина изучаемого фрагмента ряда. Тогда распределение расстояний, построенное по всему объему данных, есть среднее распределение, полученное усреднением по отдельным выборкам:

$$\begin{aligned} G_N(\rho, L_{tot}) &= \frac{1}{n}\sum_{k=1}^{n} G_N^k(\rho, L), \\ \overline{\Psi}_N(\rho) &= \sum_k p_k \Psi_N^k(\rho) = 1 - G_N(\rho, L_{tot}), \;\; p_k = \frac{1}{n}. \end{aligned} \tag{10}$$

Что в таком случае можно сказать относительно стационарной точки функции $\overline{\Psi}_N(\rho)$? Пусть $\rho_k^*(N)$ есть стационарная точка для функции $\Psi_N^k(\rho)$, а $\tilde{\rho}_k(N)$ - для эквивалентной ей функции $\Phi_N^k(\rho)$. В силу того, что $a_N^k \ge 0$, эта стационарная точка единственная. Обозначим через $\tilde{\overline{\rho}}(N)$ стационарную точку эквивалентной средней функции. С точностью до $o(\rho)$ имеем из (8-9):

$$\tilde{\rho}_k(N) = \frac{1}{1+a_N^k}, \;\; \tilde{\overline{\rho}}(N) = \frac{1}{1+\overline{a}_N} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\tilde{\rho}_k(N)}}. \tag{11}$$

Поскольку эквивалентные функции отличаются от оригиналов на величину $o(\rho)$, то на такую же величину отличаются и соответствующие стационарные точки.

Таким образом, доказана следующая теорема о стационарных точках распределений с непрерывными плотностями.

Теорема. *Пусть распределения случайных величин имеют непрерывные плотности и пусть на множестве этих случайных величин задана некоторая неотрицательная мера. Тогда стационарная точка функции, эквивалентной по Чернову среднему уровню значимости данных распределений, с точностью до бесконечно малой второго порядка совпадает с обратной величиной к среднему значению обратных величин стационарных точек функций, эквивалентных по Чернову уровням значимости данных распределений:*

$$\frac{1}{\tilde{\tilde{\rho}}(N)} = \sum_{k=1}^{n} \frac{p_k}{\tilde{\rho}_k(N)} \equiv \frac{1}{\tilde{\tilde{\rho}}(N)}. \qquad (12)$$

Если не ограничиваться линейным приближением в представлении эквивалентной функции (8), а записать в общем случае решение уравнения $e^{-a\rho} = \rho$ в виде $\tilde{\rho} = R(a)$, то вместо (12) имеем

$$R^{-1}\left(\tilde{\tilde{\rho}}\right) = \sum_{k=1}^{n} p_k R^{-1}\left(\tilde{\rho}_k\right). \qquad (13)$$

В силу сказанного ранее соотношения (12 -13) с точностью $o(\rho)$ могут быть распространены и на стационарные точки распределений $\Psi_N^k(\rho)$. Формула (13) принципиальна, поскольку она позволяет значительно уменьшить количество вычислительных процедур, если требуется изучить поведение СУС на множествах, представляющих собой объединение некоторых других множеств, для которых, собственно, и считается СУС. То есть для каждой длины N встык-выборок СУС вычисляется всего один раз в наименьших доступных для этой цели непересекающихся окнах длины L, покрывающих все исследуемое множество данных L_{tot}.

3. Структура вычислительного алгоритма

Поскольку данные ЭЭГ могут быть получены в различных форматах и с использованием различного количества электродов, то для унификации работы с такими данными используется модуль библиотеки EDFbrowser, который служит для конвертации исходных файлов в формат ASCII (github.com/Teuniz/EDFbrowser).

В начале работы алгоритма задаются либо считываются из исходного файла параметры, характеризующие входные данные, и дополнительные настройки для алгоритма. К таким параметрам относятся: количество электродов; количество данных полученных с одного электрода; шаг для встык-выборок внутри двухминутного окна; длина одной встык-выборки.

Алгоритм реализован на языке C++ в среде разработки MSVS 2017. Общая блок-схема алгоритма представлена на рис 2.

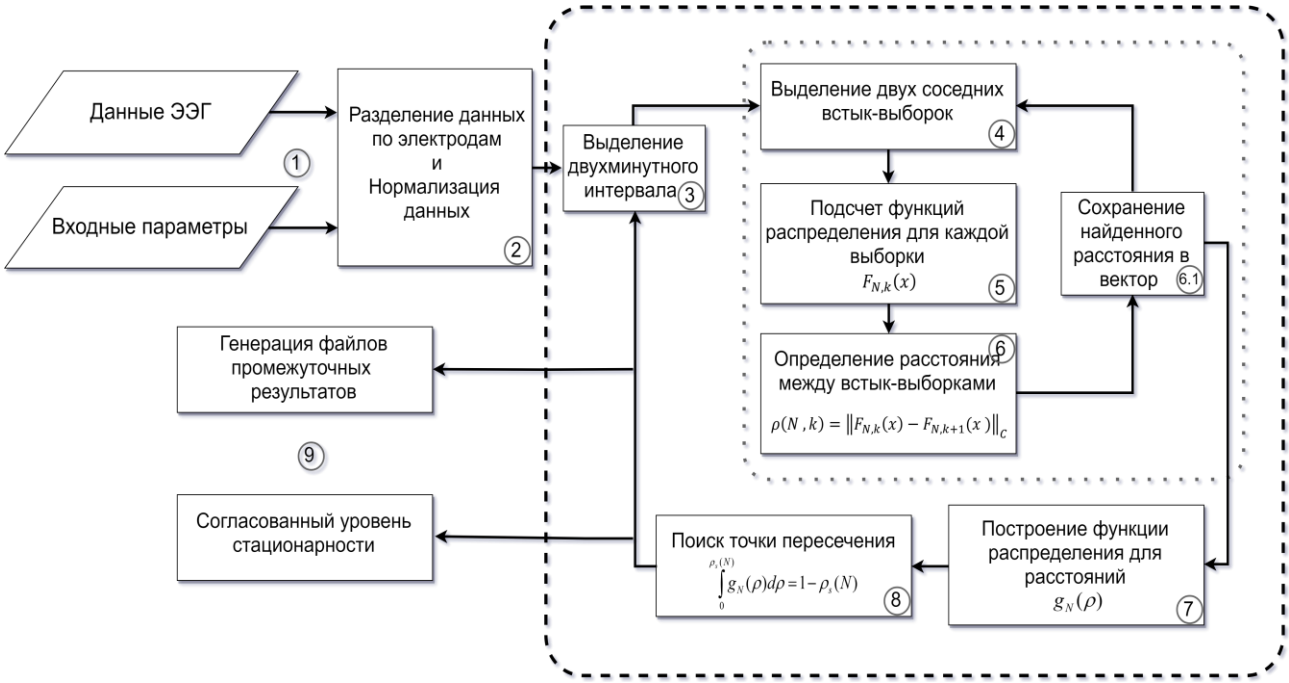

Рис. 2. Схема вычислительного алгоритма
для согласованного уровня стационарности

Реализованный алгоритм включает в себя следующие основные шаги.

Шаг 1.

После чтения данных они разделяются по векторам, отвечающим каждому (i-му) электроду. В дальнейшем удобно для каждого значения $b_{ik} = b_i(t_k)$ провести нормализацию на единичный отрезок:

$$x_{ik} = \frac{b_{ik} - \min\limits_k b_{ik}}{\max\limits_k b_{ik} - \min\limits_k b_{ik}}.$$

Время считается в единицах обратной частоты сигнала, т.е. совпадает с порядковым номером элемента ряда: $t_k = k$.

Шаг 2.

В векторе данных каждого электрода выделяются выборки, соответствующие заданному, например, двухминутному интервалу (N~30 000 данных), после чего происходит инициализация основного вычислительного модуля алгоритма. Нумеруются элементы ряда, отвечающие концевым элементам X_{ik} выборок длины N, которые сдвигаются с заданным шагом h:

$$X_{i1} = x_{iN};\ X_{i2} = x_{i,N+h};\ldots;\ X_{in} = x_{i,N+(n-1)h}.$$

Встык-выборкой с номером k при выбранных значениях параметров N и h считается множество элементов, заключенных между $x_{i,1+(k-1)h}$ и $x_{i,2N+(k-1)h}$.

Шаг 3.

Рассчитывается эмпирическая функция распределения для каждой новой пары встык-выборок из 2-х минутного интервалов. Каждая новая пара встык-выборок выделяется со сдвигом, определенным на шаге 2, относительно начала предыдущей пары. Чтобы точность вычисления ВФР не влияла бы существенно на точность последующего вычисления СУС, область изменения значений x_{ik} (единичный отрезок) разбивается сеткой на ячейки с шириной, в 100 раз меньшей, чем стационарное значение СУС для выборок длины N. Пусть H есть число ячеек (порядка 1000), на которые разбит отрезок [0; 1]. Тогда ВФР $F(k)$, где $k = 1,2,\ldots,H$ есть номер классового интервала, для множества элементов $x_1,\ldots,x_N$ вычисляется как ступенчатая функция со скачками в точках, равных концам классовых интервалов (в отличие от формального определения ВФР как ступенчатой функции со скачками в точках наблюденных значений случайной величины).

Шаг 4.

Вычисляется вектор расстояний (т.е. совокупность расстояний, вычисленных в скользящем окне с заданным шагом по времени) между эмпирическими функциями распределения двух встык-выборок, посчитанными на предыдущем шаге, в норме С. Каждое расстояние определяется по формуле

$$\rho = \max_k \left| F_1(k) - F_2(k) \right|.$$

Шаг 5.

Строится эмпирическая функция распределения из значений вектора расстояний для двух встык-выборок, посчитанного на шаге 4 по алгоритму, аналогичному описанному на шаге 3.

Шаг 6.

Решается уравнение $G_N(\rho) = 1 - \rho$. Его решение строится приближенно на сетке с шагом $1/H$. Если номер k классового интервала таков, что $G_N((k-1)/H) < 1 - (k-1)/H$, а $G_N(k/H) \geq 1 - k/H$, то этот номер фиксируется и значение $\rho^* = k/H$ считается согласованным уровнем стационарности для рассматриваемых встык-выборок.

Шаг 7.

Вычисляется эквивалентная по Чернову функция для уровня значимости $\Psi_N(\rho) = 1 - G_N(\rho)$ и с заданной точностью строится оператор R, отыскивающий ее стационарную точку.

Далее происходит возврат к шагу 1, после чего расчёты повторяются для новой пары встык-выборок другой длины.

Выводными данными являются результирующие файлы значений СУС.

При обработке большого количества векторов данных применялось распараллеливание вычислений. Была проведена программная оптимизация времени вычисления описанного выше алгоритма с использованием интерфейса OpenMP. Расчеты производились на 2, 4, 6 и 8 параллельных потоках для двух вариантов входных данных, которые представляли собой 64 временных ряда длинной 690000 и 2340000 данных (стандартные размеры файлов данных, используемых в исследовании). Каждому потоку выделялся определенный набор временных рядов, которые в дальнейшем на нем обрабатывались, и результаты вычислений записывались в отдельный файл. Собранная статистика представлена на рис. 3-4.

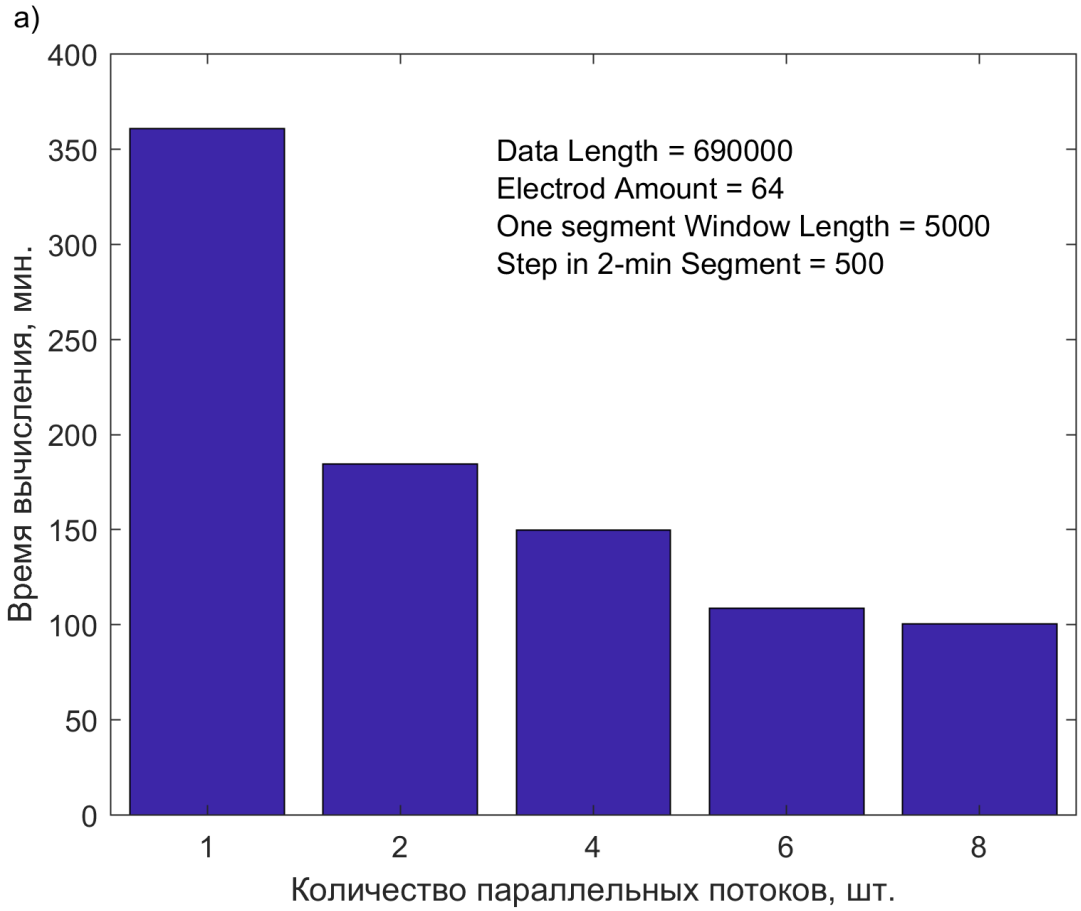

Рис. 3. Время работы алгоритма при разном количестве параллельных потоков для входного файла ЭЭГ размером 64x690000 данных

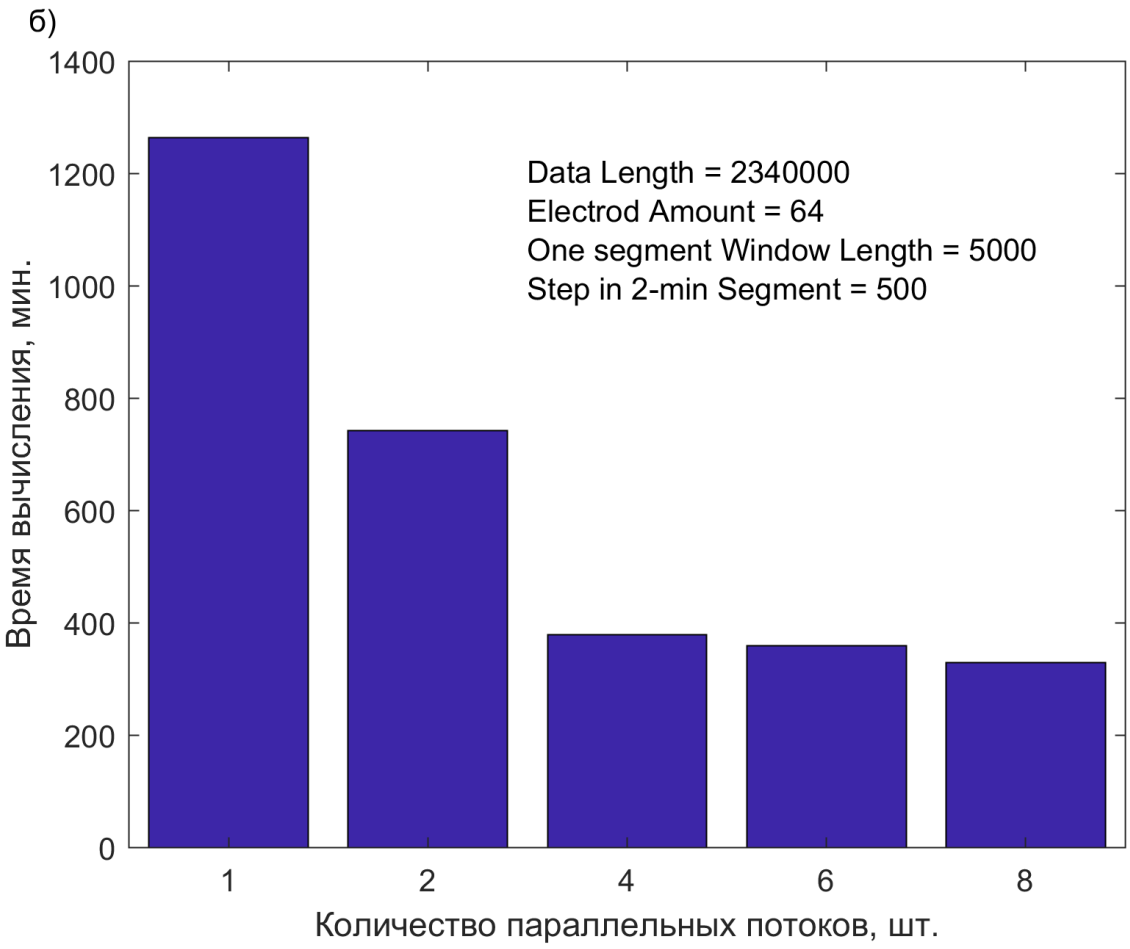

Рис. 4. Время работы алгоритма при разном количестве параллельных потоков для входного файла ЭЭГ размером 64x2340000 данных

Тестирование проводилось на системе со следующими характеристиками: Процессор Intel® Core™ i7-2670QM 6 МБ кэш-памяти, 2.20 ГГц, ОЗУ 16 ГБ.

Из рис. 3-4 видно, что качественно картина не изменилась, а время счета пропорционально количеству данных (то есть во втором случае в 4 раза больше).

Видно также, что на 8 потоках удалось достигнуть ускорения алгоритма примерно в 3,6 раз по сравнению с однопоточной последовательной обработкой.

Данный вычислительный процесс обладает определенным потенциалом для дальнейшего распараллеливания, что будет учтено при будущей разработке программного комплекса.

4. Стационарная точка уровня значимости для ЭЭГ

В работе [10] была исследована возможность предсказания приступа эпилепсии на основе анализа электроэнцефалограмм (ЭЭГ) и вычисления для них уровня стационарности. Предположение заключалось в том, что уровни стационарности пациента в спокойном состоянии и в момент приступа различаются. На основе этой идеи был построен предиктор приступа и определены эмпирически те отведения, по которым СУС в последние две минуты перед приступом заметно (более, чем на величину самого СУС) превосходил СУС спокойного состояния. Однако анализа зависимости СУС от величины промежутка, по которому он вычислялся, не проводилось. Настоящая работа восполняет указанный недостаток.

Чтобы выяснить, действительно ли СУС является индикатором состояния пациента, или же последовательность $\rho_k^*(N)$ просто представляет собой реализацию некоторого случайного процесса, не связанного с этим состоянием напрямую, а вызванного случайным изменением окружающей среды, на которое и реагирует мозг, необходимо проверить на стационарность выборку $\rho_k^*(N)$ на промежутке, где состояние пациента стабильно.

Напомним, что в п.2 были введены распределения $G_N^k(\rho, L)$ расстояний между встык-выборками длины *N*, построенные по промежуткам длины *L*, для которых $\rho_k^*(N) \equiv \rho_k^*(N, L)$ есть СУС *k*-го промежутка длины *L*. Если объем данных $nL = L_{tot}$ достаточен для того, чтобы построить распределения величин СУС $\rho_k^*(N, L)$, то можно сравнить эти распределения для одного и для разных состояний пациента и при фиксированной длине встык-выборки *N* исследовать зависимость этих распределений от длины фрагмента *L*.

В работе [10] минимальная длина окна, в котором вычислялся индикатор разладки, была принята равной 2 мин (при частоте 250 Гц в этом промежутке укладывалось *L* = 30 тыс. данных ЭЭГ по каждому отведению). Оптимальная длина встык-выборок, между которыми вычислялось расстояние ρ, была принята равной *N* = 5000. Расстояния между встык-выборками вычислялись в скользящем окне длиной 10000 с шагом 500. Таким образом, для длины *L* имеем 40 вычисленных расстояний между ВФР встык-выборок. Общий объем данных, отвечающих визуально не меняющемуся состоянию бодрствования пациента, составил 2,5 часа, т.е. примерно 2,3 млн данных или *n* = 78 независимых двухминутных отрезков. Для каждого из этих двухминутных промежутков и для каждого отведения по имеющимся сорока данным вычислялся СУС $\rho_k^*(5000, 30000)$, $k = 1,..n = 78$. Типичный пример ряда указанных величин СУС для одного из электродов приведен ниже на рис. 5.

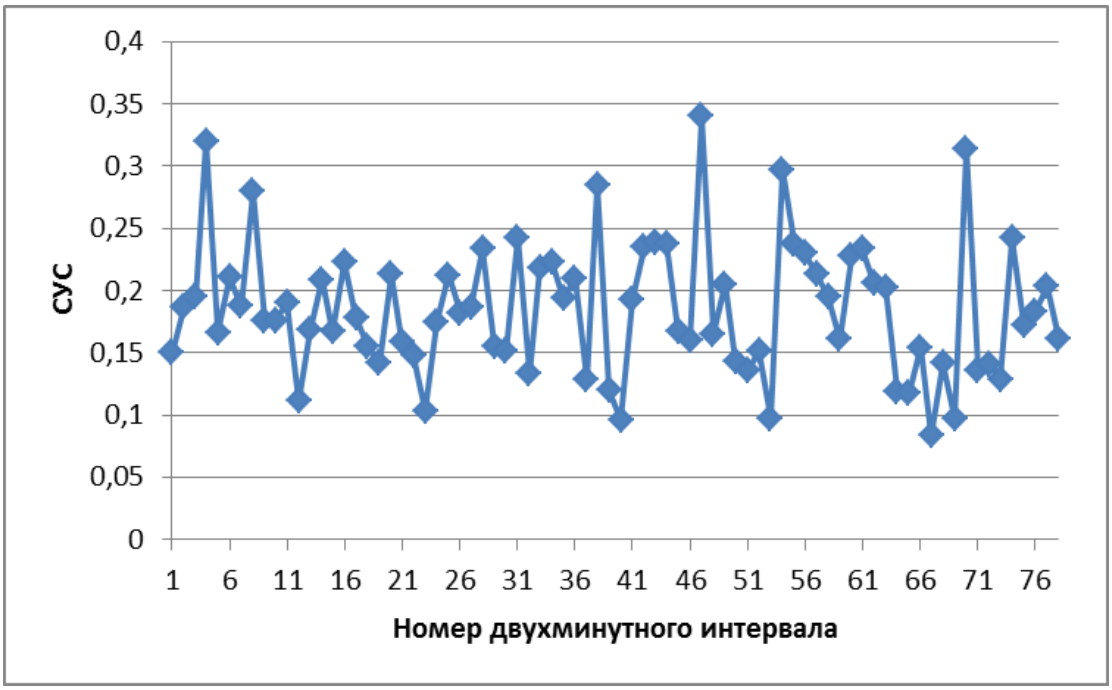

Рис. 5. Пример ряда СУС $\rho_k^*(5000, 30000)$

Эмпирические распределения значений СУС по первой и второй половинам данных, а также по всей выборке в целом приведены на рис. 6.

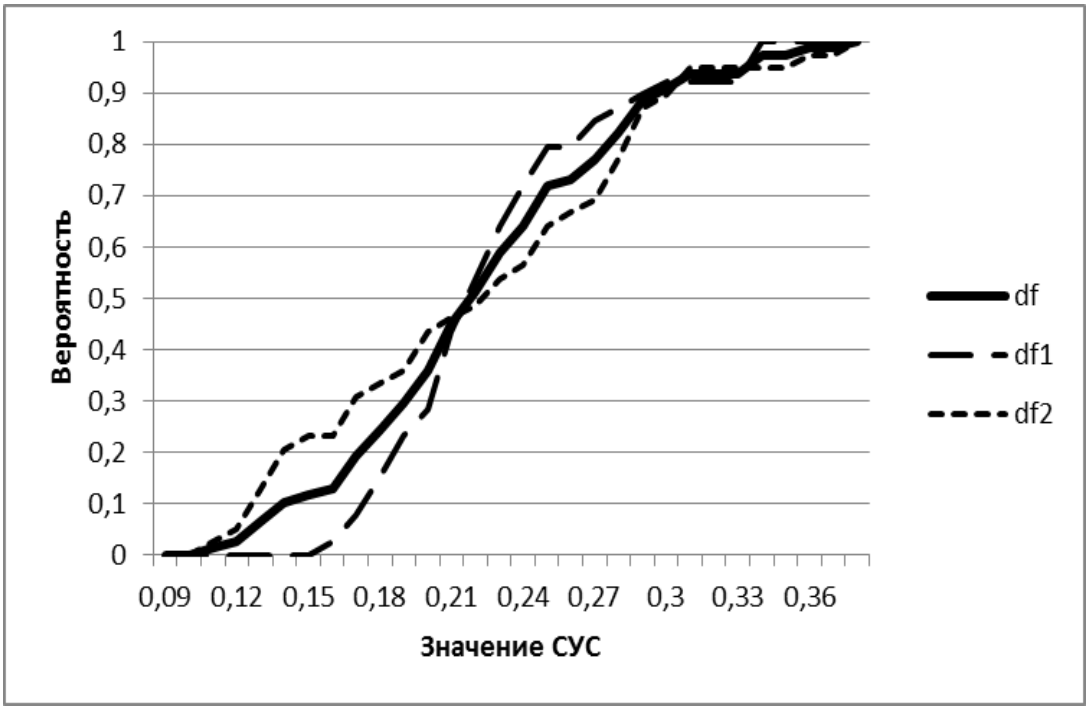

Рис. 6. Эмпирические распределения СУС $\rho_k^*(5000, 30000)$

Для стационарных выборок длиной 40 данных теоретический согласованный уровень значимости составляет $\varepsilon_0(40) = 0{,}25$. В приведенном примере (как и для всех 64 отведений рассматриваемых данных ЭЭГ) СУС меняется от 0,245 до 0,255. Это означает, что ряд СУС стационарный, хотя сам исходный ряд показаний ЭЭГ, для которых этот СУС посчитан, не является

стационарным. Индекс нестационарности (6) для встык-выборок ряда ЭЭГ длиной 5000 на промежутке L_{tot} приблизительно равен $J = 10$, т.е. ряд данных ЭЭГ сильно нестационарный. Это, в частности, означает, что применение к нему различных стационарных критериев и методов по факту даст ошибку существенно большую, чем уровень значимости критерия.

Поскольку же ряд значений СУС стационарный, имеет смысл исследовать его статистические характеристики в зависимости от длины окна L, в котором этот СУС вычисляется. Это необходимо для того, чтобы понимать, с какой вероятностью такой локальный СУС превосходит СУС, посчитанный по всему множеству. Длина встык-выборок при этом фиксирована и равна $N = 5000$.

В работе [10] было выяснено, что лишь некоторые отведения (порядка 10) могут служить индикаторами-предвестниками приступа эпилепсии. Это те отведения, для которых относительное превышение индекса нестационарности в последние две минуты перед приступом по сравнению со спокойным состоянием составляет величину не менее, чем СУС. Поэтому имеет смысл сравнить статистические свойства распределений СУС для индикативных и остальных отведений.

Введем среднее значение и стандартное отклонение для ряда СУС как функции длины окна L:

$$avr(n) = \frac{1}{n}\sum_{k=1}^{n}\rho_k^*(N,L),\ sd(n) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\rho_k^*(N,L) - avr(n)\right)^2},\quad n = \frac{L_{tot}}{L}. \tag{14}$$

На рис. 7 представлена зависимость нормированного среднего значения СУС от количества промежутков n (т.е. в зависимости от обратной длины окна) для индикативного электрода и дана ее степенная аппроксимация.

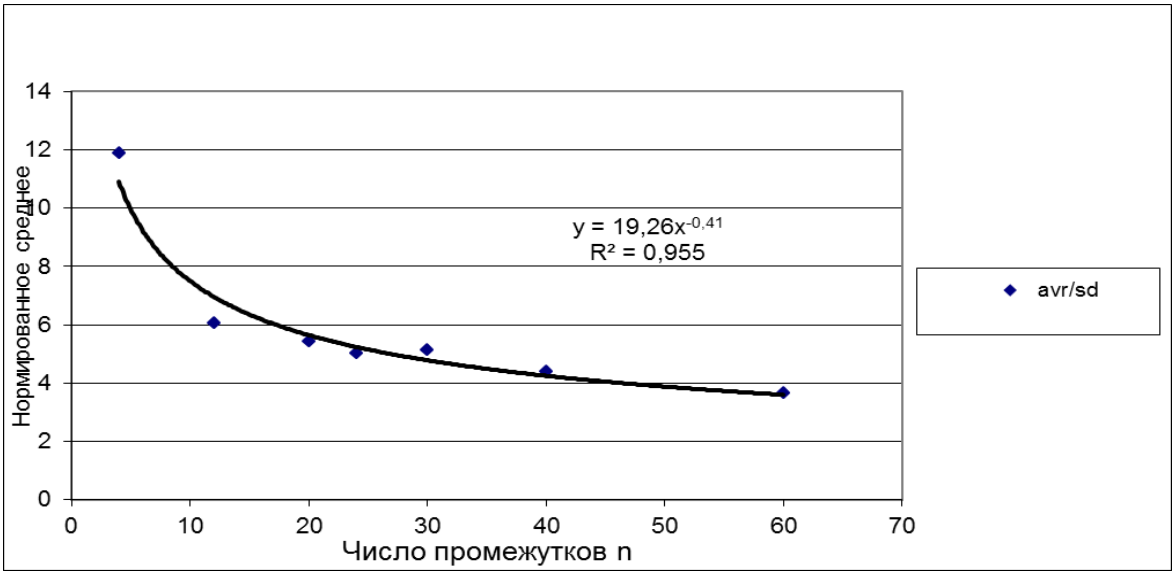

Рис. 7. Зависимость нормированного среднего СУС от длины окна, индикативный электрод

Отметим, что, хотя длина выборки ряда СУС невелика (менее 100), все же степенная аппроксимация близка к зависимости $\frac{avr(n)}{sd(n)}\sqrt{n} \approx const$. Это согласуется со сделанным ранее выводом о стационарности ряда СУС, поскольку есть снование считать, что отклонение локального СУС от среднего значения распределено асимптотически нормально в соответствии с центральной предельной теоремой. Плотность функции распределения СУС, обозначенной как df на рис. 6, представлена на рис. 8. Видно, что она имеет приближенно симметричный треугольный вид:

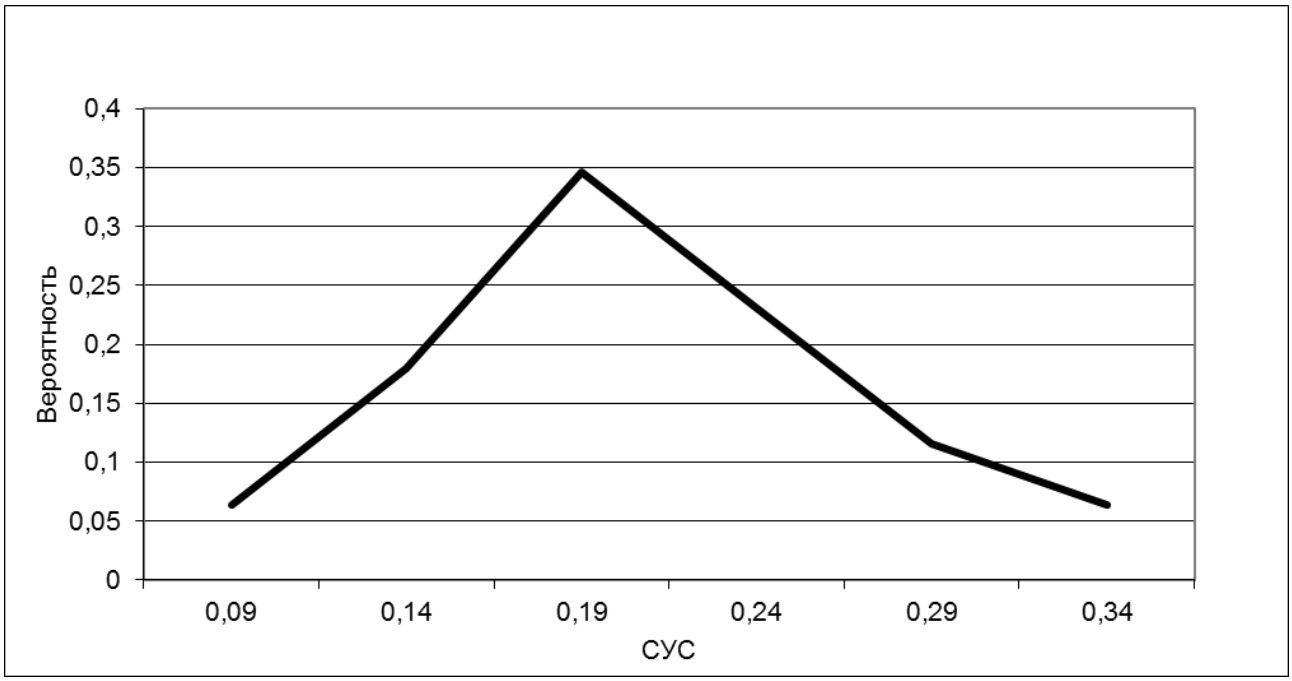

Рис. 8. Стационарная плотность распределения СУС для окна 30000 (2 мин.), индикативный электрод

Отметим здесь, что среднее значение СУС, равное для данного примера $avr(n) \approx 0{,}21$, достаточно близко к стационарной точке функции значимости $\Psi_{5000}(\rho)$, отвечающей всему множеству в целом: $\rho^*_{tot}(5000) \approx 0{,}22$. Значение же стационарной точки эквивалентной ей по Чернову функции $\Phi_{5000}(\rho)$ составило величину $\tilde{\rho}_{tot}(5000) \approx 0{,}20$, а ее оценка, полученная по формуле (13), приближенно равна $\tilde{\tilde{\rho}}(5000) \approx 0{,}19$.

Согласно рис. 6, вероятность превышения уровня $\rho^*_{tot}(5000) \approx 0{,}22$ в окне длины L на величину, равную $\left(\rho^*_{tot}(5000)\right)^2 \approx 0{,}05$, есть $1 - df(0{,}27) \approx 0{,}23$. Эта величина равна вероятности ошибки ложной тревоги.

Рассмотрим теперь, как меняется максимальное значение СУС, наблюдаемое на полном объеме данных L_{tot} для встык-выборок длины 5000, в зависимости от длины окна L. На рис. 9 показана соответствующая эмпирическая зависимость и ее аппроксимация.

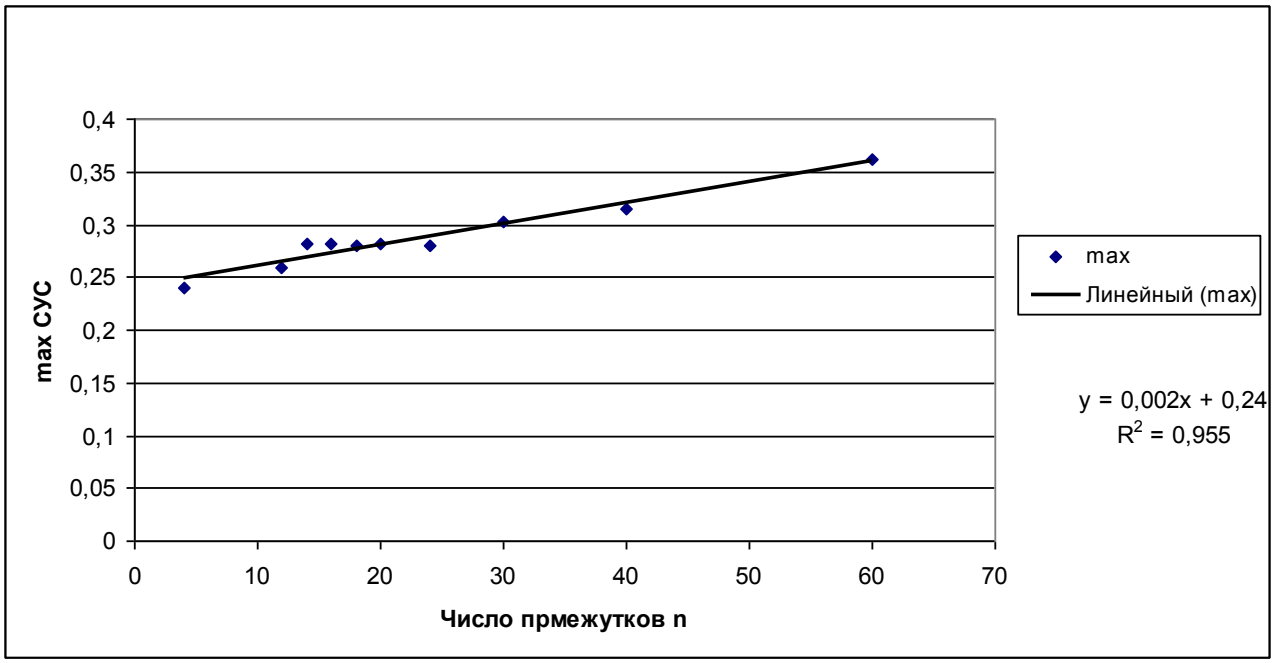

Рис. 9. Зависимость максимального значения СУС для встык-выборок длины 5000 от длины окна, индикативный электрод

Из рис. 9 видно, что указанная зависимость носит линейный характер (обратная пропорциональность длине L).

На рис. 10-11 показаны аналогичные зависимости для обычных электродов, не служащих индикаторами разладки.

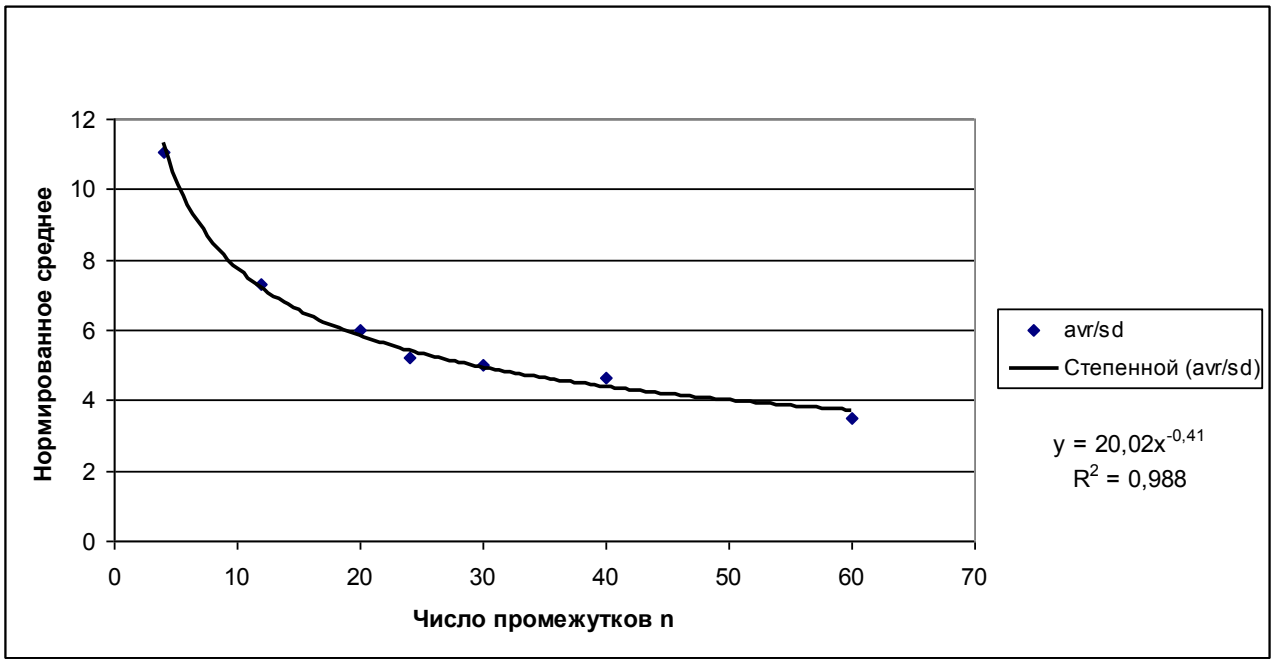

Рис. 10. Зависимость нормированного среднего СУС от длины окна, «обычный» электрод

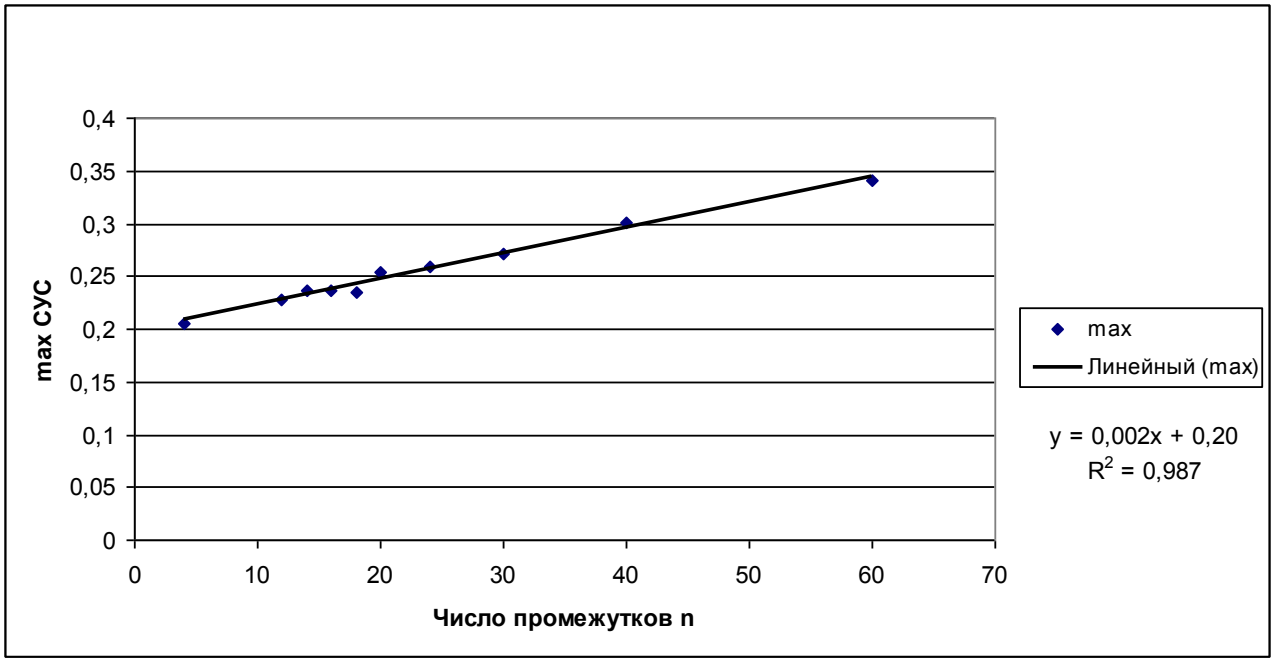

Рис. 11. Зависимость максимального значения СУС для встык-выборок длины 5000 от длины окна, «обычный» электрод

Таким образом, выяснилось, что по каждому типу электродов – индикативному и не индикативному – распределение согласованного уровня стационарности является стационарным. Иными словами, тип нестационарного поведения временного ряда не меняется в течение того промежутка времени, пока состояние пациента визуально считается стабильным. При этом детерминация аппроксимаций рассмотренных зависимостей для «обычных» электродов заметно выше, чем для индикативных, то есть эти «обычные» электроды порождают более стационарный ряд данных. Важно подчеркнуть, что при этом сам ряд данных ЭЭГ для визуально стабильного состояния пациента является нестационарным, поэтому индикатором смены состояния пациента следует считать не отклонение показаний ЭЭГ (ибо стационарные критерии оценки уровня значимости оказываются некорректными), а изменение уровня нестационарности выборочных распределений показаний ЭЭГ, если этот уровень действительно различен для разных состояний. В следующем разделе приводятся данные по расчетам СУС для разных, но условно стабильных состояний.

5. СУС как индикатор разладки

Рассмотрим уровни нестационарности для состояний, которые предположительно можно считать различными: это, во-первых, состояния бодрствования и сна без приступов эпилепсии, и, во-вторых, состояния, предшествующие двум приступам эпилепсии. Выяснилось, что распределения СУС этих состояний различаются значительно сильнее, чем это можно ожидать для выборок из одной генеральной совокупности. При этом существенно, что для каждого состояния показания ЭЭГ имеют стационарно распределенный ряд

значений СУС. Поскольку плотности распределения визуально более информативны, чем сами ВФР, мы приведем на рис. 12 именно эти функции, хотя расстояния между распределениями различных состояний указаны в норме С для ВФР.

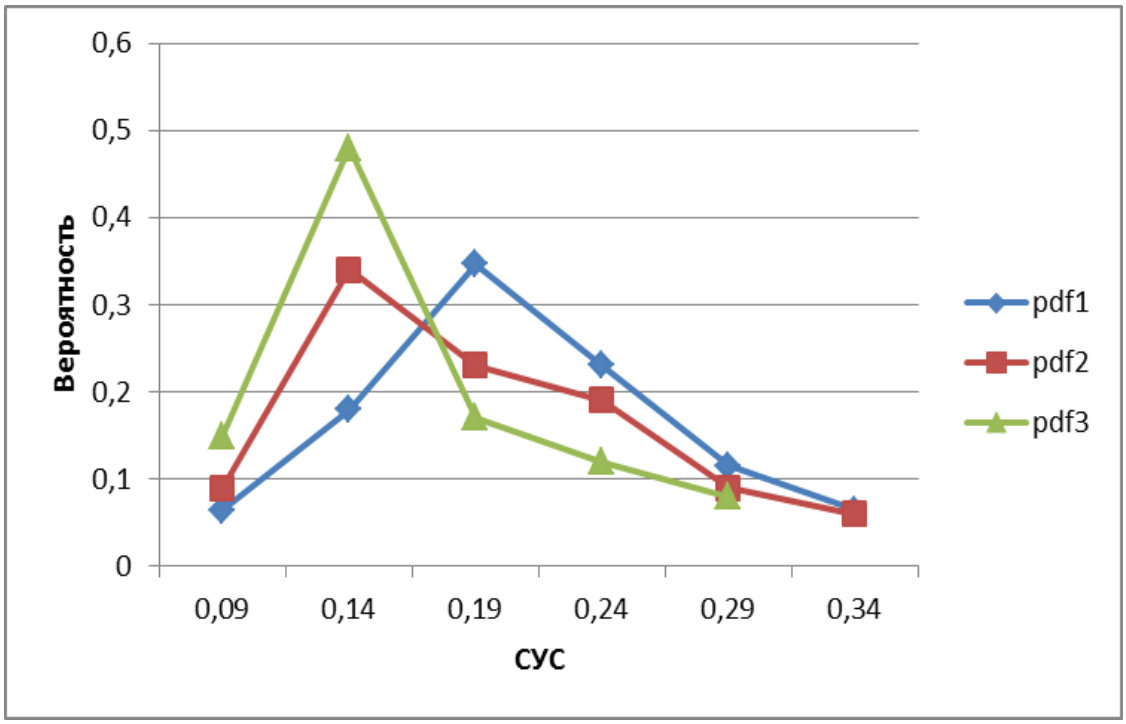

Рис. 12. Плотности распределений СУС для встык-выборок длины 5000 в окне длиной 2 минуты для трех состояний: 1- бодрствование до засыпания, 2- сон, 3- бодрствование между приступами эпилепсии

Расстояния между этими состояниями (ВФР в норме С) следующие: $\delta_{12} = 0{,}36$; $\delta_{23} = 0{,}33$; $\delta_{31} = 0{,}45$. Количество данных, по которым построены эти распределения: pdf1 – 78 значений, pdf2 – 90 значений, pdf3 – 55 значений. Типичными же значениями расстояний между стационарными распределениями таких длин являются величины порядка 0,25. Следовательно, можно сделать вывод, что величина СУС действительно различна для разных чередующихся состояний пациента, так что она может служить индикатором разладки. Разладкой в данном случае следует считать смену одного нестационарного режима показаний ЭЭГ другим нестационарным же, но с отличающимся от предыдущего состояния уровнем СУС.

Литература

1. Орлов Ю.Н. Кинетические методы исследования нестационарных временных рядов. – М.: МФТИ, 2014. – 276 с.
2. Орлов Ю.Н., Федоров С.Л. Методы численного моделирования процессов нестационарного случайного блуждания. – М.: МФТИ, 2016. – 112 с.
3. Орлов Ю.Н., Федоров С.Л. Моделирование распределений функционалов на ансамбле траекторий нестационарного случайного процесса // Препринты ИПМ им. М.В. Келдыша. 2016. № 101. 14 с.
URL: http://library.keldysh.ru/preprint.asp?id=2016-101
4. Орлов Ю.Н., Федоров С.Л. «Программный комплекс NonStatBox для статистического анализа и моделирования нестационарных временных рядов». Свидетельство о регистрации программы для ЭВМ № 2017619117 от 15.08.2017.
5. Kolmogoroff A.N. Sulla determinazione empirica di una legge di distribuzione // Giornale dell` Istituto Italiano degly Attuari. 1933. – Vol. 4. – № 1. – P. 83-91.
6. Гнеденко Б.В. Курс теории вероятностей. – М.: Физматлит, 1961. – 406 с.
7. Smolyanov O.G., Weizsacker H., Wittin O. Chernoff's theorem and discrete time approximation of Brownian motion on manifolds // Potential Anal., 26, 2007. P. 1-29.
8. Орлов Ю.Н., Сакбаев В.Ж., Смолянов О.Г. Формулы Фейнмана как метод усреднения случайных гамильтонианов // Труды МИРАН, 2014. Т. 285. С. 232-243.
9. Бутко Я.А., Смолянов О.Г. Формулы Фейнмана в стохастической и квантовой динамике / В сб. Современные проблемы математики и механики, 6, № 1, 2011. С. 61-75.
10. Kislitsyn A.A., Kozlova A.B., Masherov E.L., Orlov Yu.N. Numerical Algorithm for Self-consistent Stationary Level for Multidimensional Non-stationary Time-series // Keldysh Institute Preprints. 2017. № 124. 14 p.
URL: http://library.keldysh.ru/preprint.asp?id=2017-124