

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.
ЧЕРНЫШЕВСКОГО»**

Кафедра радиофизики и нелинейной динамики

Корреляционный анализ нестационарных данных

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 2232 группы

направления 03.04.03 Радиофизика

института физики

Рящикова Алексея Сергеевича

Научный руководитель

д.ф.-м.н., профессор

В.В. Астахов

Заведующий кафедрой

д.ф.-м.н., доцент

Г.И. Стрелкова

Саратов 2023

ВВЕДЕНИЕ

Многие процессы в динамике природных систем демонстрируют степенную статистику с дальнедействующими корреляциями. Их описание часто используется для понимания поведения системы и выявления изменений, вызванных переменными внешними условиями или переходами между различными типами сложной динамики. Корреляционный анализ экспериментальных данных ограничен двумя основными проблемами: (i) затуханием корреляционной функции, приближающейся к нулю, где ошибки вычислений сильно влияют на количественную оценку степенного поведения, и (ii) нестационарностью регистрируемых сигналов, связанной с зависимостью оцениваемых величин от выбора момента начала отсчета времени. Эти обстоятельства привели к разработке метода DFA (detrended fluctuation analysis или анализ флуктуаций относительно тренда) [1, 2], который был предложен для улучшения возможностей корреляционного анализа нестационарных временных рядов. Его особенность заключается в переходе от затухающей корреляционной функции к растущей функции (случайному блужданию или профилю сигнала), степенное поведение которой легче охарактеризовать. Возможности и ограничения этого подхода были исследованы в последующих работах [3–5] с использованием смоделированных наборов данных. В течение последних десятилетий DFA рассматривался как вполне универсальный подход в различных экспериментальных исследованиях наряду с вейвлет-анализом.

В своей стандартной версии, описанной Пенгом с соавторами [1, 2], DFA включает в себя разбиение профиля сигнала на неперекрывающиеся сегменты с процедурой удаления тренда, которая использует аппроксимацию тренда в пределах каждого сегмента и его удаление. Профиль с удаленным трендом далее может быть охарактеризован в терминах среднеквадратичных флуктуаций, которые обычно увеличиваются с ростом длины сегмента.

Для этой цели в основном используется кусочно-линейная аппроксимация, хотя применимы и другие типы аппроксимирующих

функций. Такой подход хорошо адаптирован для довольно однородных временных рядов со схожими особенностями тренда по всему сигналу. Однако часто реальная динамика сложных систем весьма неоднородна, особенно для переходных процессов. Эти обстоятельства являются причиной появления модифицированного подхода для неоднородных временных рядов – расширенного DFA (EDFA), который оценивает два показателя скейлинга.

Целью выпускной квалификационной работы является изучение возможностей применения метода расширенного флуктуационного анализа для изучения процессов с различными типами нестационарности, присутствующих во временных рядах, на примере случайных процессов с известными статистическими свойствами, а также его тестирование на экспериментальных данных.

Материалы исследования. Исследования проводились с использованием тестовых сигналов (шум различной статистики) и экспериментальных данных (записи ЭЭГ лабораторных животных в различных состояниях).

Выпускная квалификационная работа содержит введение, три главы (1. Флуктуационный анализ и его расширенная версия; 2. Анализ эффектов нестационарности в тестовых примерах; 3. Анализ экспериментальных данных), заключение и список использованных источников. Общий объем работы 60 стр.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Флуктуационный анализ и его расширенная версия. Алгоритм DFA включает построение одномерного случайного блуждания (профиля сигнала) и его последующий среднеквадратичный анализ. В соответствии со статьей, где этот метод был впервые предложен и протестирован [1], профиль аппроксимируется кусочно-линейной функцией (после разделения его на сегменты, аппроксимация проводится в пределах каждого сегмента), и

стандартная ошибка аппроксимации вычисляется в зависимости от длины сегмента. DFA сигнала $x(i)$, $i = 1, \dots, N$ включает следующие этапы:

1) Формирование профиля сигнала (случайного блуждания)

$$Y(i) = \sum_{k=1}^i [x(k) - \langle x \rangle], \quad (1)$$

где $\langle x \rangle$ - среднее значение.

2) Деление профиля $Y(i)$ на M неперекрывающихся сегментов длины n и определение локального тренда $Y_n(i)$ путем интерполяции $Y(i)$ в пределах каждого сегмента. Вместо кусочно-линейной функции можно выбрать полиномиальную аппроксимацию, и метод DFA часто применяют для нескольких вариантов аппроксимации тренда, вводя соответствующие обозначения DFA1, DFA2, DFA3, ... для полиномов степени 1, 2, 3, ...

3) Вычитание локального тренда и вычисление среднеквадратичных флуктуаций $F(n)$

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^N [Y(i) - Y_n(i)]^2}. \quad (2)$$

4) Повторение данных вычислений при варьировании длины сегмента n и вычисление степенной зависимости

$$F(n) \sim n^\alpha, \quad (3)$$

где α – показатель скейлинга метода DFA. В двойном логарифмическом масштабе ($\lg F$ от $\lg n$), показатель скейлинга легко вычислить с помощью линейной аппроксимации. Наличие линейной зависимости подтверждает существование степенных корреляций в анализируемых данных.

Дополнительное повышение точности вычислений можно обеспечить за счет следующего приема. Поскольку часть сигнала при сегментации не учитывается (последний фрагмент, длительность которого меньше величины рассматриваемого окна), то сегментация может быть проведена дважды – в прямом и обратном направлении. В результате увеличится число сегментов для анализа среднеквадратичных флуктуаций.

Вычисляемое значение α характеризует различные типы корреляций в экспериментальных данных, включая анти-коррелированную динамику ($\alpha < 0.5$, присутствует чередование больших и малых значений временного ряда) и степенные корреляции ($0.5 < \alpha < 1$, большие значения чаще следуют за большими, малые – за малыми). При $\alpha > 1$ наблюдается коррелированное поведение, которое может отличаться от степенной статистики.

Для алгоритма DFA различия локальных величин стандартных отклонений, рассчитанных для разных участков, не принимаются в расчет, хотя они способны оказать значительное влияние на оценки $F(n)$. Ранее была рассмотрена дополнительная мера, характеризующая эффекты нестационарного поведения

$$dF(n) = \max[F_{loc}(n)] - \min[F_{loc}(n)], \quad (4)$$

где $F_{loc}(n)$ – локальные значения среднеквадратичных отклонений зависимости $Y(k)$ от аппроксимирующей функции $Y_n(k)$, которые вычисляются в рамках одного сегмента. Для стационарных процессов при заданном n разброс значений $F_{loc}(n)$ будет сравнительно небольшим, и величина $dF(n)$ приближается к нулю. При наличии сильной нестационарности $dF(n)$ принимает значения в диапазоне от нуля до $\max[F_{loc}(n)]$. Обычно наблюдается рост $dF(n)$ при увеличении n , и соответствующая степенная зависимость описывается другим показателем скейлинга

$$dF(n) \sim n^\beta. \quad (5)$$

Данный метод будем называть расширенным методом DFA (или EDFA).

Анализ эффектов нестационарности в тестовых примерах.

Используя смоделированные наборы данных, было проведено обсуждение трех типов нестационарности: вариации локального среднего значения (тип 1), переключения между различными процессами (тип 2) и

нестационарность амплитудных / энергетических характеристик (тип 3).

Также рассмотрены три примера случайных процессов, а именно

- белый шум (некоррелированный процесс с $\alpha = 0,5$),
- шум $1/f$ (коррелированный процесс с $\alpha = 1,0$),
- шум с анти-корреляциями (производная от $1/f$ -шума, $\alpha \approx 0,1$).

При численном моделировании интенсивности этих процессов выбирались таким образом, что диапазоны случайных значений приблизительно соответствуют друг другу.

В качестве первого типа нестационарности рассмотрим медленные изменения локального среднего значения (низкочастотные вариации, которые можно интерпретировать как тренд). Для этого к каждому случайному процессу была добавлена гармоническая функция $A \sin(\omega t)$ с очень низкой частотой ω и амплитудой A . Частота была выбрана таким образом, чтобы вся запись (64 000 выборок) включала 2-3 периода колебаний, а сегменты данных, проанализированные в рамках EDFa, содержали небольшие фрагменты одного периода. Последнее позволяет интерпретировать эти части как тренд.

Рассмотрим в качестве примера результаты для фиксированного значения амплитуды ($A = 0,03$) с использованием белого шума со стандартным отклонением $\sigma = 0,075$. Без аддитивного сигнала, приводящего к изменениям локального среднего значения, показатели скейлинга α совпадают в диапазонах $\lg n < 3,0$ и $\lg n > 3,0$ ($\alpha \approx 0,5$). Когда к белому шуму добавляются гармонические колебания, показатель α не изменяется в первом диапазоне, но он значительно увеличивается при больших $\lg n$. Аналогичное поведение наблюдается и для показателя β . Последний обычно принимает значения меньше α , но его поведение аналогично; тренд существенно не влияет на оценки для $\lg n < 3,0$, но локальный наклон $\lg dF(n)$ от $\lg n$ изменяется для $\lg n > 3,0$. В связи с этим были выбраны два локальных значения каждого показателя скейлинга в диапазоне $\lg n < 3,0$ (соответственно, α_1 и β_1) и в диапазоне $\lg n > 3,0$ (α_2 и β_2).

Эти показатели зависят от амплитуды гармонической функции. Рассматриваемый медленный тренд не влияет на значения α_1 и β_1 , которые вычисляются с использованием коротких сегментов данных (менее 1/25-й части периода $T = 2\pi/\omega$, соответствующего $\lg n < 3,0$). Вычисляемые величины не изменяются с увеличением амплитуды A . Однако медленная нестационарность оказывает более выраженное влияние на оба показателя скейлинга в диапазоне $\lg n > 3,0$. Также α_2 увеличивается примерно в 3,3 раза при максимальном значении A , в то время как β_2 увеличивается еще сильнее (в 5 раз). Последнее может привести к неправильному толкованию результатов вычислений. Без предварительной фильтрации стохастический процесс с $\alpha_1 = 0,5$ и трендом может быть идентифицирован как процесс, близкий к броуновскому движению ($\alpha_2 = 1,5$) или даже к более “гладкому” случайному процессу. Стандартное отклонение белого шума в этом примере составляет $\sigma = 0,075$, а значения случайного процесса в основном соответствуют диапазону $[-0,2, 0,2]$, следовательно, $A \leq \sigma$.

В качестве второго типа нестационарности рассмотрены периодические переключения (с периодом T) и проанализированы результаты в зависимости от τ (длительности участков 1-го режима) и отношения τ/T . Были выбраны переключения между белым шумом и шумом с анти-корреляциями, белым шумом и $1/f$ – шумом, $1/f$ – шумом и шумом с анти-корреляциями. Три случая, а именно переключения между белым шумом и антикоррелированным случайным процессом, белым шумом и $1/f$ – шумом, $1/f$ – шумом и шумом с антикорреляциями, анализировались с помощью разности значений показателей скейлинга в зависимости от A . Чередования между сегментами белого шума и шума с антикорреляциями приводят к относительно небольшим смещениям показателя α от уровня $\alpha_0 = 0,5$, который соответствует некоррелированному процессу. Изменения значений β более сильные, и этот показатель уменьшается примерно с 0,3 до отрицательных значений. Наиболее выраженные реакции возникают, когда τ выбирается вблизи значения $T/10$. Чередования между белым шумом и

шумом $1/f$ приводят к сдвигу α с $\alpha_0 = 0,5$ для коротких сегментов шума $1/f$ примерно до 1, когда сегменты каждого случайного процесса сопоставимы. В этом примере показатель β изменяется от отрицательных значений примерно до 0,7. Переключения между шумом $1/f$ и шумом с антикорреляциями не приводит к сильным изменениям обоих показателей в зависимости от τ . Чтобы наглядно показать и сравнить эффекты, изучены различия в показателях $\alpha - \alpha_0$ и $\beta - \beta_0$. Процесс с более высокими значениями α_0 оказывает более сильное влияние на оцениваемые характеристики дальнедействующих корреляций.

Анализ экспериментальных данных. В динамике природных систем, например, в физиологии, могут сосуществовать несколько типов нестационарности, и их устранение или коррекция является сложной проблемой даже для простейшего типа нестационарности – низкочастотного тренда. Сложная динамика таких систем не всегда позволяет устранить низкочастотные вариации локального среднего значения, например, при изучении медленных волновых процессов, связанных с механизмами физиологической регуляции. В последнем случае часто неясно, как отличить интересующие медленные ритмы от тренда. Как правило, тренд связан с полосой частот ниже исследуемых ритмов. Его удаление с помощью фильтра верхних частот не влияет на более быструю динамику, но удается избежать влияния этого тренда на количественную оценку состояния системы. Такая фильтрация может оказаться неприемлемой при изучении медленной динамики. Здесь мы рассмотрим один из таких примеров, а именно активацию очистительной функции головного мозга во время сна. Эта активация связана с диапазоном частот $f < 1$ Гц, который обычно не рассматривается в традиционно выполняемом анализе ЭЭГ-ритмов. Без предварительной фильтрации данных нельзя быть уверенным в том, что нестационарные составляющие данных не повлияют на результаты EDFA.

Проведенный анализ основан на экспериментах, выполненных на девяти взрослых крысах с использованием двухканальных записей ЭЭГ в

бодрствующем состоянии (нормальные условия) и во время сна. Все измерения проводились с частотой дискретизации 2 кГц. После удаления артефактов выбирались 10-минутные сегменты со сравнительно однородной структурой.

Применение EDFA выявляет различия в обоих показателях скейлинга. В отличие от показателя α обычного DFA, который обычно снижается во время сна, показатель β может менять знак и становится отрицательным в диапазоне больших $\lg n$. Для одного из примеров α уменьшается с 0,78 до 0,43, количественно отражая изменение типа корреляций (положительные степенные корреляции у бодрствующего животного и анти-корреляции во время сна). Показатель β уменьшается с 0,41 до -0,29. Аналогичный эффект наблюдается для всей группы животных.

Также изучались индуцированные звуком изменения электрической активности головного мозга. Существует несколько защитных элементов, которые защищают мозг от различных повреждений, и гематоэнцефалический барьер (ГЭБ) является одним из важнейших из них. Возможность временного открытия ГЭБ для лечения активно обсуждалась в последние годы. Важным вопросом является обратимая процедура: вскрыть барьер для доставки лекарства с дальнейшим восстановлением его нормальной функции. Недавние исследования предложили и изучили возможности довольно безопасного подхода, основанного на использовании громкой музыки или звука в течение 2 часов. Экспериментальные исследования были проведены на 8 крысах в СГУ. Был применен подход, основанный на использовании громкого звука (100 дБ, 370 Гц) в течение 2 часов в соответствии с алгоритмом: 1 мин – звук, 1 мин – пауза. Регистрировалась двухканальная ЭЭГ. Активность ЭЭГ регистрировали и сравнивали у бодрствующих крыс в нормальных условиях (ГЭБ закрыт) и через 1 ч после громкого звука (ГЭБ открыт).

Так, рис. 1 иллюстрирует разности между значениями α для закрытого и открытого ГЭБ и разности между соответствующими значениями β . Здесь

выполняется усреднение для 10 оценок локальных наклонов $\lg F$ от $\lg n$ и $\lg dF$ от $\lg n$ в диапазоне $\lg n \in [3.4, 4.5]$. Несмотря на некоторую вариабельность показателей в зависимости от диапазона, значения α и β изменяются, когда ГЭБ открыт.

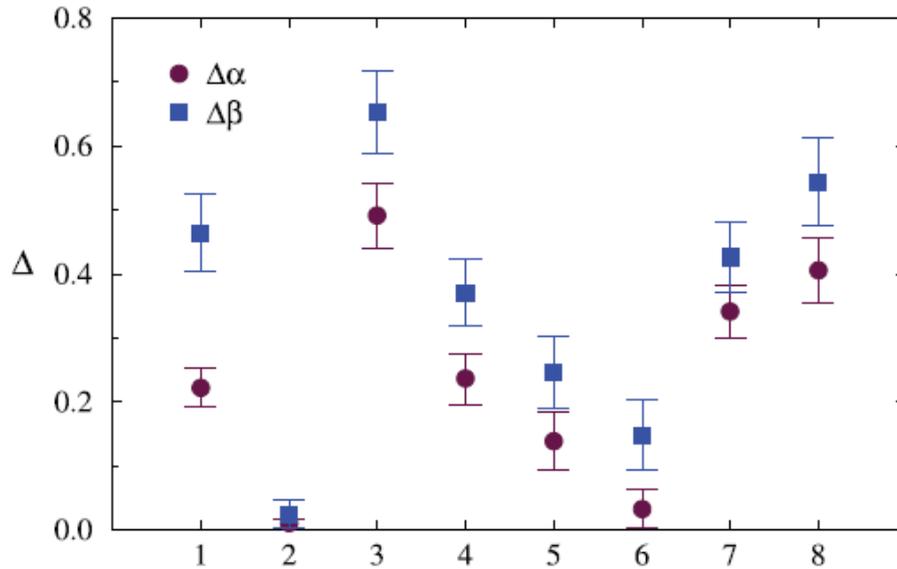


Рисунок 1 – Статистический анализ, который показывает более сильные различия между группами, характеризующиеся расширенным DFA ($\Delta\beta$) по сравнению с традиционным подходом ($\Delta\alpha$).

Значимость выявленных различий была подтверждена с использованием как непараметрических, так и параметрических статистических тестов. Для диапазона $\lg n > 3,5$ тест Стьюдента дает t-значения 2,53 для α ($p = 0,017$) и 3,06 для β ($p = 0,005$), которые превышают критическое значение $t_c = 2,042$, связанное с уровнем значимости 0,05. В качестве непараметрической альтернативы t-критерию использовался тест Манна-Уитни, который не имеет ограничений в отношении распределения сравниваемых величин. Этот тест также подтвердил значимость межгрупповых различий и позволил получить значения $p < 0,05$ для α и $p < 0,01$ для β .

Использование EDFA не означает, что диагностика с помощью показателя α стандартного DFA не очень хорошая. Более того, зависимости $\lg F$ от $\lg n$ часто показывают более узкое распределение оценочных значений вокруг аппроксимирующей линии. Для показателя β значения $\lg dF$ от $\lg n$ могут демонстрировать более сильные отклонения от аппроксимирующей функции, и их распределение шире, что приводит к большим ошибкам при вычислении β . Для того, чтобы надежно разделить оба состояния с помощью β , эти состояния должны характеризоваться показателями скейлинга, разность которых превышает возможные вычислительные ошибки. Вот почему большее значение $\Delta\beta$ по сравнению с $\Delta\alpha$ важно для достоверного разделения групп. Оба подхода могут быть использованы одновременно для улучшения анализа данных, когда различия довольно невелики. Так, в частности, оценки $\Delta\alpha$ для крысы № 6 (рис. 1) приводят к относительно небольшим величинам, которые рассматриваются как несущественные. Обработка данных, выполненная с помощью $\Delta\beta$, улучшает диагностику реакции на открытие ГЭБ у этого животного.

Таким образом, в работе было изучено влияние разных типов нестационарности на результаты флуктуационного анализа сигналов. Было проанализировано влияние тренда, переключений между разными типами случайных процессов и изменение энергетических характеристик для нескольких вариантов стохастических процессов. Было показано, что каждый вариант нестационарного поведения приводит к изменениям показателей скейлинга расширенного метода (EDFA). Медленные вариации локального среднего уровня (тренд), оказывают наибольшее влияние на антикоррелированные процессы и сопровождаются значительными смещениями обоих показателей, α и β . С ростом α и связанного с этим перехода от антикорреляций к положительным степенным корреляциям изменения показателей скейлинга становятся менее выраженными. При переключениях между случайными процессами показатель скейлинга α принимает значения,

которые ближе к максимальной величине α для двух рассматриваемых процессов.

Полученные результаты показывают, что знание эффектов нестационарности имеет решающее значение для надежной характеристики сложных процессов. Несмотря на мнение, что DFA можно применять к экспериментальным данным без предварительной обработки, устранение нестационарности на первом этапе (если эта процедура возможна) является важной задачей, способной существенно повысить надежность интерпретации результатов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Buldyrev, S. Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis / S. Buldyrev, A. Goldberger, S. Havlin, R. Mantegna, M. Matsuoka, C.-K. Peng, M. Simons, H. Stanley // *Phys. Rev. E*. - 1995. - Vol. 51. - P. 5084-5091.
- [2] Peng, C.-K. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series / C.-K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger // *Chaos*. – 1995. – Vol. 5. – P. 82–87.
- [3] Ma, Q. D. Y. Effect of extreme data loss on long-range correlated and anticorrelated signals quantified by detrended fluctuation analysis / Q. D. Y. Ma, R. P. Bartsch, P. Bernaola-Galvan, M. Yoneyama, P. Ch. Ivanov // *Phys. Rev. E*. – 2010. – Vol. 81. – P. 031101.
- [4] Havlin, S. Scaling in Nature: from DNA through heartbeats to weather / S. Havlin, S. Buldyrev, A. Bunde, A. Goldberger, P. Ivanov, C.-K. Peng, H. Stanley // *Physica A*. - 1999. - Vol. 273. - P. 46-69.
- [5] Pavlov, A. N. Detrended fluctuation analysis of cerebrovascular responses to abrupt changes in peripheral arterial pressure in rats / A. N. Pavlov, A. S. Abdurashitov, A. A. Koronovskii Jr., O. N. Pavlova, O. V. Semyachkina-Glushkovskaya, J. Kurths // *Commun. Nonlinear Sci. Numer. Simulat.* – 2020. – V. 85. – P. 105232.