

© 2023 г. Ю.А. ДУБНОВ (yury.dubnov@phystech.edu)

(Федеральный исследовательский центр

“Информатика и управление” РАН, Москва;

Национальный исследовательский университет

“Высшая школа экономики”, Москва),

А.Ю. ПОПКОВ, канд. техн. наук (aporkov@isa.ru)

(Федеральный исследовательский центр

“Информатика и управление” РАН, Москва),

Б.С. ДАРХОВСКИЙ, д-р физ.-мат. наук (darbor2004@mail.ru)

(Федеральный исследовательский центр

“Информатика и управление” РАН, Москва)

ЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ АЛГОРИТМА ОЦЕНИВАНИЯ ГЁЛЬДЕРОВОЙ ЭКСПОНЕНТЫ НА БАЗЕ КОНЦЕПЦИИ ϵ -СЛОЖНОСТИ НЕПРЕРЫВНЫХ ФУНКЦИЙ¹

В работе приводится описание метода оценивания гёльдеровой экспоненты на базе недавно сформулированной концепции ϵ -сложности непрерывных функций. Приведены результаты вычислительных экспериментов по оценке гёльдеровой экспоненты для гладких и фрактальных функций, а также результаты вычислительных экспериментов с траекториями дискретных детерминированных и стохастических систем.

Ключевые слова: ϵ -сложность непрерывных функций, гёльдерова экспонента.

DOI: 10.31857/S0005231023040025, EDN: CEVIZP

1. Введение

Процессы фрактального типа (в частности, винеровский и фрактальный винеровский процессы) в последнее время широко используются при построении моделей различных событий, эффектов и трендов, включая так называемые процессы с длительной памятью. Большое распространение такие процессы находят в различных областях физики (см., например, [1]), биологии и других областях науки. Довольно сложно найти область науки, в которой не встречались бы представители этого класса процессов.

Важной характеристикой стохастических случайных процессов является показатель Хёрста (определение можно найти, например, в [2]). Оценка этого показателя по наблюдаемой реализации случайного процесса представляет достаточно сложную проблему. В недавней работе [3] предложен метод такого оценивания.

¹ Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 20-07-00221).

Одной из особенностей фрактальных функций является то, что они, будучи непрерывными, являются нигде недифференцируемыми, однако большой класс подобных функций удовлетворяет условию Гёльдера. Оценка гёльдеровской экспоненты (см. ниже точное определение) для фрактальных функций имеет такое же значение, как и оценка показателя Хёрста в стохастическом случае, поскольку, как известно (см. [4]), для винеровского и фрактального винеровского процессов с вероятностью единица локальная гёльдеровская экспонента сколь угодно близка к константе Хёрста. С другой стороны, гёльдеровская экспонента тесным образом связана с размерностью Хаусдорфа или показателем скейлинга, описывающего сохраняемость геометрии или статистических характеристик при изменении масштаба. Приведенные соображения делают задачу оценивания гёльдеровской экспоненты актуальной и интересной для многочисленных приложений.

В этой работе приводится экспериментальный анализ принципиально нового подхода к проблеме оценивания гёльдеровской экспоненты (и, как было отмечено выше, константы Хёрста). Этот подход основан на развиваемой в последние годы разработанной авторами концепции ϵ -сложности непрерывных функций, которая в идейном плане согласуется с подходом А.Н. Колмогорова к понятию «сложность объекта». Отметим, что концепция сложности объекта находит применения не только в теории управления, но и в широком классе иных приложений [5]. С другой стороны, применения концепции ϵ -сложности непрерывных функций в последние несколько лет вызывают определенный интерес у читателей ведущих научных журналов, в том числе в первом-втором квартилях WoS/Scopus [6–9]. Поэтому данная работа может не только познакомить читателей журнала с новой концепцией ϵ -сложности, но и потенциально привести к новым направлениям научных и практических исследований в данной области.

Основные результаты концепции ϵ -сложности непрерывных функций и метод оценивания гёльдеровской экспоненты на основе этой концепции можно найти в [8, 10]. В разделе 2, для удобства чтения, приводятся на содержательном уровне соответствующие результаты из этих работ. В разделе 3 описываются процедура оценивания и объекты анализа — некоторые гладкие и фрактальные непрерывные функции, а также последовательности, порождаемые стохастическими или детерминированными хаотическими механизмами. В разделе 4 приведены результаты вычислительных экспериментов.

2. ϵ -сложность и оценка гёльдеровской экспоненты

Неформально можно сказать, что ϵ -сложность оценивает число равномерных дискретных отсчетов непрерывной функции (заданной без ограничения общности на отрезке $[0, 1]$), которые необходимы для ее восстановления заданным набором *методов аппроксимации* с заданной точностью. Иными словами, эта величина оценивает минимальный объем информации (*на языке теории приближений*), который требуется для описания функции. В этом

отношении концепция ϵ -сложности согласуется с предложенным А.Н. Колмогоровым в середине 60-х годов понятием «сложность» объектов. Основная содержательная идея колмогоровского подхода (см., например, [11]) состоит в том, что: «сложный» объект требует для своего описания много информации, а «простой» — мало, т.е. сложность объекта целесообразно оценивать через минимальный объем информации, требующийся для его описания.

Методом восстановления (аппроксимации) непрерывной функции по ее дискретным отсчетам на равномерной решетке с некоторым шагом $1 > h > 0$ называем произвольную борелевскую функцию, переводящую конечный набор отсчетов (число которых определяется величиной h) в какую-либо ограниченную функцию на $[0, 1]$ (в пространстве ограниченных функций введена равномерная метрика). Фиксируем произвольные счетные наборы борелевских функций со значениями в пространстве ограниченных функций, зависящих соответственно, от $1, 2, 3, \dots$ аргументов. Будем называть *списком* объединение этих счетных наборов. Список содержит счетный набор методов восстановления для всех $h > 0$. Фиксируем некоторый список методов восстановления \mathbb{F} . Всюду далее символом \mathcal{F} обозначается произвольное непустое подмножество \mathbb{F} , содержащее какие-то наборы борелевских функций от $1, 2, 3, \dots$ аргументов. Будем говорить, что наборы \mathcal{F} (и соответственно списки \mathbb{F} при $\mathcal{F} = \mathbb{F}$) являются *допустимыми*, если они содержат методы аппроксимации кусочно-постоянными (ступенчатыми) функциями и степенными полиномами. Положим

$$\delta^{\mathcal{F}}(h) = \inf_{\hat{x}_h(\cdot) \in \mathcal{F}} \sup_{t \in [0, 1]} |\hat{x}_h(t) - x(t)|.$$

Здесь символом $\hat{x}_h(\cdot) \in \mathcal{F}$ обозначаются оценки $x(\cdot)$ по ее отсчетам с шагом h , полученные методами семейства \mathcal{F} . В случае, когда $\mathcal{F} = \mathbb{F}$, для оценки используются все функции, входящие в \mathbb{F} .

Имеет место следующее утверждение.

Множество непрерывных функций, которые не могут быть точно восстановлены по конечному числу своих отсчетов методами произвольного, но фиксированного допустимого списка \mathbb{F} , всюду плотно в пространстве всех непрерывных функций на $[0, 1]$.

Назовем \mathcal{F} -нетривиальными функции, которые *не могут быть точно восстановлены* методами произвольного непустого допустимого подмножества $\mathcal{F} \subseteq \mathbb{F}$.

Пусть \mathbb{F} — фиксированный допустимый список и $\mathcal{F} \subseteq \mathbb{F}$ — произвольное непустое допустимое подмножество. Пусть $x(t)$ — \mathcal{F} -нетривиальная функция, $\max_{t \in [0, 1]} |x(t)| = R$. Для достаточно малых $\epsilon > 0$ положим

$$h_x^*(\epsilon, \mathcal{F}) = \inf \left\{ h \leq 1 : \frac{\delta^{\mathcal{F}}(h)}{R} > \epsilon \right\}.$$

Будем называть (ϵ, \mathcal{F}) -сложностью непрерывной функции $x(\cdot)$ величину $\mathbb{S}_x(\epsilon, \mathcal{F}) = -\log h_x^*(\epsilon, \mathcal{F})$.

Если функция не является \mathcal{F} -нетривиальной (т.е. восстанавливается точно по конечному числу своих отсчетов), то полагаем, что ее (ϵ, \mathcal{F}) -сложность равна нулю. Таким образом, из приведенного выше утверждения следует, что «почти все» непрерывные функции имеют ненулевую (ϵ, \mathcal{F}) -сложность при любом непустом допустимом $\mathcal{F} \subseteq \mathbb{F}$ для произвольного фиксированного допустимого списка \mathbb{F} .

Заметим, что $h_x^*(\epsilon, \mathcal{F}) > 0$ при $\epsilon > 0$ и $\lim_{\epsilon \rightarrow 0} h_x^*(\epsilon, \mathcal{F}) = 0$ для \mathcal{F} -нетривиальной функции. С другой стороны, $\lim_{h \rightarrow 0} \delta^{\mathcal{F}}(h) = 0$. Поэтому для любого (достаточно малого) $\epsilon > 0$ существует $\eta(\epsilon) > 0$, $\eta(\epsilon) \rightarrow 0$ при $\epsilon \rightarrow 0$, такое, что $\delta^{\mathcal{F}}(h_x^*(\cdot)) \leq \eta(\epsilon)$.

Величина $1/h_x^*(\epsilon, \mathcal{F})$ оценивает число отсчетов функции. Поэтому (ϵ, \mathcal{F}) -сложность функции есть логарифм числа ее отсчетов, которые требуются для ее восстановления методами семейства \mathcal{F} с относительной погрешностью не более $R^{-1}\eta(\epsilon)$. Иными словами, можно говорить, что это *кратчайшее описание функции* данными методами с данной точностью, что и было отмечено в начале этого раздела.

Пусть $x(t) : [0, 1] \rightarrow \mathbb{R}^1$, $\max_{t \in [0, 1]} |x(t)| = R$, непрерывная функция такая, что существуют константы $K > 0$ и $0 < p \leq 1$ (показатель Гельдера, или *гельдерова экспонента*), для которых справедливо неравенство $\sup_{(t_1, t_2) \in [0, 1], t_1 \neq t_2} |t_1 - t_2|^{-p} |x(t_1) - x(t_2)| \leq K$ (отметим, что если в этом неравенстве предполагается $|t_1 - t_2| \leq \epsilon$ для некоторого $\epsilon > 0$, то говорится о *локальной гельдеровой экспоненте*).

Пусть фиксирован некоторый допустимый список методов восстановления \mathbb{F} и пусть \mathcal{T} — множество гельдеровых функций, *не восстанавливаемых точно методами \mathbb{F} по конечному числу своих отсчетов*. Пусть функция $x(\cdot)$ принадлежит некоторому плотному подмножеству $T_0 \subseteq \mathcal{T}$ и задается своими n отсчетами (т.е. своим следом) на некоторой равномерной решетке отрезка $[0, 1]$. Из основного результата концепции ϵ -сложности следует, что для всех допустимых семейств аппроксимирующих методов $\mathcal{F} : \mathcal{F}^* \subset \mathcal{F} \subseteq \mathbb{F}$, где $\mathcal{F}^* \subset \mathbb{F}$ — достаточно богатое множество, и достаточно малых $\epsilon > 0$ существует шаг решетки $h_x^*(\epsilon, \mathcal{F})$ такой, что сложность $\mathbb{S}_x(\epsilon, \mathcal{F}) = -\log h_x^*(\epsilon, \mathcal{F})$ допускает эффективное описание.

Фиксируем (достаточно малое) число $\epsilon > 0$ и предположим, что набор методов восстановления есть $\mathcal{F} \supseteq \mathcal{F}^*$, а число отсчетов n , задающих функцию, удовлетворяет условию $n \geq n_0$, где $[h_x^*(\epsilon, \mathcal{F})n_0] = C$, $C \gg 1$ (здесь и далее символ $[a]$ означает целую часть числа a). Это означает, что если из каждых $[h_x^*(\cdot)n]$ отсчетов оставить один (а остальные выбросить), то этого будет достаточно для восстановления функции с точностью до $R^{-1}\eta(\epsilon)$.

Выберем число $0 < S < 1$ и выбросим из первоначальных n отсчетов функции $x(\cdot)$ некоторую часть так, чтобы оставшиеся $[Sn]$ отсчеты были расположены *равномерно*.

Имеет место следующее утверждение.

Для произвольной гёльдеровской функции из плотного подмножества \mathcal{T}_0 , заданной своими n отсчетами на некоторой равномерной решетке отрезка $[0, 1]$, при достаточно богатом множестве методов аппроксимации \mathcal{F} , достаточно большом n и достаточно малом $\epsilon > 0$ имеет место соотношение:

$$(1) \quad \log \epsilon = A_n - p \log S + \varphi(\epsilon, n),$$

где $\lim_{n \rightarrow \infty} \frac{A_n}{\log n} = -p$ и $\varphi(\cdot)$ — ограниченная функция.

Алгоритм оценивания гёльдеровской экспоненты состоит в следующем. Задаются несколько значений параметра $0 < S < 1$, $S = S_k$, $k = 1, \dots, m$ и для каждого S_k вычисляется наименьшая (по имеющемуся списку методов аппроксимации \mathcal{F}) погрешность ϵ_k аппроксимации функции (в выброшенных точках, так как функция представлена только своим следом на решетке). Затем по набору пар $(\log \epsilon_k, \log S_k)$ стандартным методом наименьших квадратов определяются оценки \hat{A}_n, \hat{B}_n коэффициентов линейной части соотношения (1). Пусть $\hat{A}_n = A_n + \rho(\epsilon, n, m)$, где $\rho(\epsilon, n, m)$ — погрешность оценивания методом наименьших квадратов. В силу ограниченности функции $\varphi(\cdot)$ погрешность $\rho(\epsilon, n, m)$ также ограничена. Отсюда получаем следствие:

$$(2) \quad \lim_{n \rightarrow \infty} \frac{\hat{A}_n}{\log n} = -p.$$

3. Процедура оценивания. Объекты анализа

3.1. Описание алгоритмической процедуры

Пусть $\{x(k)\}_{k=1}^n$ — след непрерывной функции на некоторой равномерной решетке. В силу изложенного в предыдущем пункте для минимальной погрешности аппроксимации имеет место приближенное равенство

$$(3) \quad \log \epsilon \approx A_n + B_n \log S,$$

здесь смысл символа \approx подразумевает приближенное восстановление линейной зависимости $\log \epsilon$ от $\log S$ с точностью до остатков линейной регрессии, обозначение которых в данном выражении опущено. Коэффициенты A_n, B_n назовем *коэффициентами сложности*. Оценка этих коэффициентов и есть цель алгоритмической процедуры.

Выбираем параметр S из следующего ряда: $S = 1/2, 1/3, 1/4, 1/5, 1/6, 1/10$. В зависимости от величины S оставшиеся после выбрасывания значения наблюдаемой последовательности $\{x(k)\}$ могут располагаться разными способами. Например, при $S = 1/2$ можно оставить значения $x(1), x(3), x(5), \dots$ или значения $x(2), x(4), x(6), \dots$. Для других значений S число возможных размещений оставляемых значений функции увеличивается.

Последовательность вычислений такова:

1. При выбранном значении S для каждого варианта размещения оставшихся после выбрасывания точек последовательности $\{x(k)\}$ определяется минимальная по всем имеющимся в наличии методам аппроксимации погрешность $\epsilon(S, i)$ восстановления в выброшенных точках по оставшимся точкам. Здесь i — номер варианта размещения оставшихся точек. После этого вычисляется $\epsilon(S) = \min_i \epsilon(S, i)$ — минимальная по всем вариантам размещения оставшихся точек погрешность восстановления для данного значения S .
2. Описанный процесс повторяется при каждом значении S из указанного выше ряда.
3. В результате проведения вычислений по пп. 1 и 2 получаем наборы $(\log \epsilon(S), \log S)$. По этому набору стандартным методом наименьших квадратов находим коэффициенты сложности в (3) для длины выборки n . Затем все вычисления повторяются для исследования предельного (при $n \rightarrow \infty$) поведения этих коэффициентов.

В силу конечномерности выбор нормы для вычисления погрешности аппроксимации не принципиален. Здесь использовали норму l_1 . Набор методов аппроксимации в проведенных экспериментах содержал кусочные полиномы до четвертой степени включительно.

3.2. Объекты анализа

В экспериментах исследовались следующие функции.

1. Гладкая функция вида

$$x(t) = \sum_{i=1}^k a_i \sin \omega_i t,$$

где ω_i — несоизмеримые частоты. Хотя внешне такая функция производит впечатление «сложной», она дифференцируема и ее гельдерова экспонента равна единице.

2. Функция Вейерштрасса ([4])

$$x(t) = \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k t),$$

где $1 < \lambda$, $1 < s < 2$. Гельдеровская экспонента этой функции равна $(2 - s)$.

3. Стандартный винеровский процесс (броуновское движение). Известно (см., например, [12]), что траектория этого процесса описывается с вероятностью единица следующим рядом:

$$x(t) = \frac{1}{\sqrt{\pi}} C_0 t + \sqrt{\frac{2}{\pi}} \sum_{k=1}^{\infty} C_k \frac{\sin kt}{k},$$

где C_k — стандартные независимые нормальные случайные величины. Известно (см., например, [13]), что *локальная гёльдеровская экспонента* этого процесса с вероятностью единица сколь угодно близка к $1/2$.

При работе с функциями, заданных рядами, было необходимо оценить достаточное число членов этих рядов, так как моделировать возможно только конечные суммы. Целесообразным представляется следующее правило: погрешность (по отношению к шкале сигнала) от замены ряда конечной суммой должна быть по крайней мере на два порядка меньше, чем погрешность при аппроксимации функции в выброшенных точках при помощи имеющегося набора методов. Для стохастического процесса под погрешностью понимается величина среднеквадратичного отклонения. При анализе стохастических процессов все результаты вычислений должны усредняться по большому количеству реализаций.

В данной работе также рассмотрены траектории дискретных динамических систем как стохастических, так и детерминированных хаотического типа. Любая числовая последовательность конечной длины может рассматриваться как след некоторой непрерывной функции на конечном интервале. Действительно, например, полином соответствующего порядка будет проходить через все значения последовательности, так что множество непрерывных функций, следом которых является данная числовая последовательность, не пусто. Если оставаться на этой позиции, то для любой конечной числовой последовательности можно применить концепцию ϵ -сложности и использовать основное соотношение (3).

Многочисленные эксперименты, в том числе и с реальными данными, результаты которых были опубликованы в работах авторов в последние годы, свидетельствуют о том, что соотношение (3) работает для числовых последовательностей, порождаемых динамическими системами с дискретным временем как стохастического, так и детерминированного типа. Однако особый интерес представляет исследование поведения коэффициентов сложности для различных последовательностей при росте их длины.

4. Результаты вычислительных экспериментов

4.1. Гладкие функции

В первом эксперименте в качестве исследуемой функции рассматривалась сумма синусоид со взаимно простыми частотами, вид функции представлен на рис. 1.

На рис. 2 приведены графики зависимости вычисляемых коэффициентов ϵ -сложности при росте количества точек на отрезке $[0, 1]$. Здесь и далее результаты представлены в виде двух графиков: зависимость величины $A(n)/\ln(n)$ (слева) и зависимость коэффициента $B(n)$ (справа).

Как видно из графиков на рис. 2, обе величины стремятся к -1 при росте n в полном соответствии с исследуемой гипотезой о гёльдеровской экспоненте.

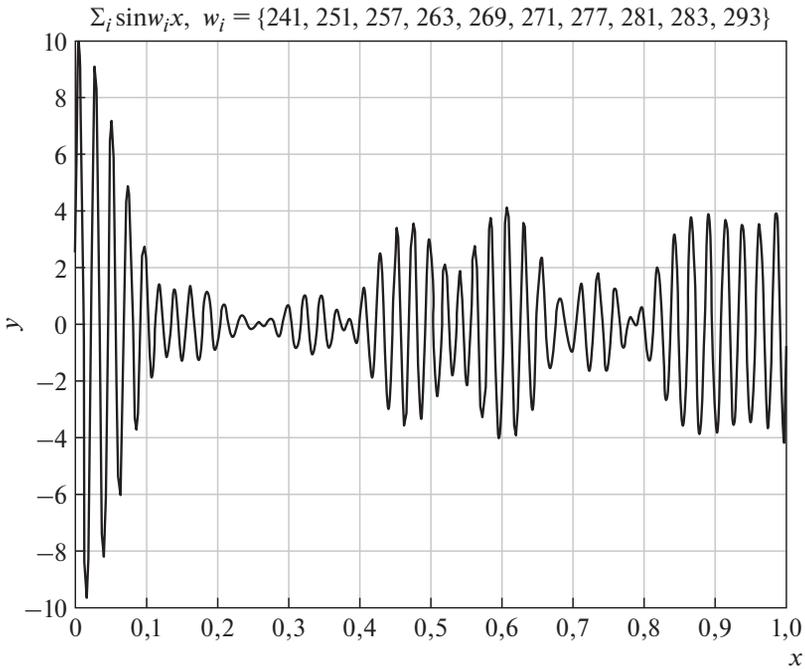


Рис. 1. Пример гладкой функции.

$\sum_i \sin w_i x, w_i = \{241, 251, 257, 263, 269, 271, 277, 281, 283, 293\}$
 $n = [10000, 300000]$

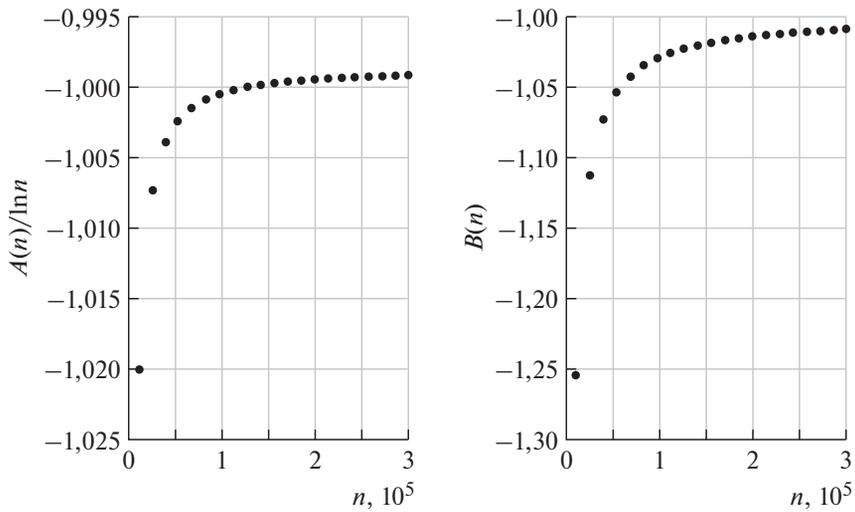


Рис. 2. Результаты вычислений для гладкой функции.

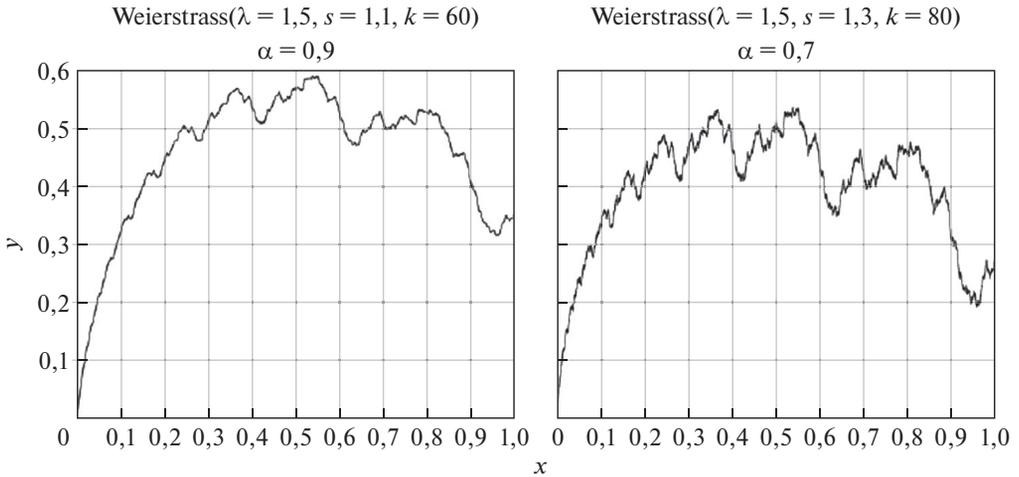


Рис. 3. Пример моделирования функции Вейерштрасса со значениями $s = 1,1$ и $s = 1,3$.

4.2. Функция Вейерштрасса

Моделирование функции Вейерштрасса производится по формуле, приведенной в разделе 3.2 для некоторого конечного числа слагаемых k_0 . Приведем кратко формулу для оценки предельной величины остаточной суммы, учитывая, что $\lambda^{(s-2)} < 1$ при $1 < \lambda$, $1 < s < 2$:

$$\begin{aligned}
 S_{rest} &= \sum_{k=k_0+1}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k t) \leq \\
 &\leq \sum_{k=k_0+1}^{\infty} \lambda^{(s-2)k} = \frac{\lambda^{(s-2)(k_0+1)}}{1 - \lambda^{(s-2)}} = \frac{A}{2} \lambda^{(s-2)k_0},
 \end{aligned}$$

где константа $A = \frac{2}{\lambda^{(2-s)} - 1}$ используется для нормировки шкалы сигнала.

В работе было рассмотрено два варианта функции Вейерштрасса со значениями $s = 1,1$ и $s = 1,3$ (рис. 3), т.е. со значениями гельдеровской экспоненты $\alpha = 0,9$ и $\alpha = 0,7$ соответственно. Результаты вычислений для функции Вейерштрасса приведены на рис. 4.

Стоит отметить, что при недостаточном количестве слагаемых ряда, когда остаточная сумма оказывается велика, наблюдается ситуация, при которой коэффициенты сложности сходятся к -1 , как в случае гладких функций, т.е. моделируемая функция теряет свои фрактальные свойства.

В обоих рассмотренных примерах предельные значения коэффициентов превышают -1 . Так, в эксперименте при $s = 1,1$ значения коэффициента B сходятся приблизительно к $-0,8$, а при $s = 1,3$ — к $-0,57$. Что касается коэффициента A , то, вероятно, с учетом предельного характера отношения

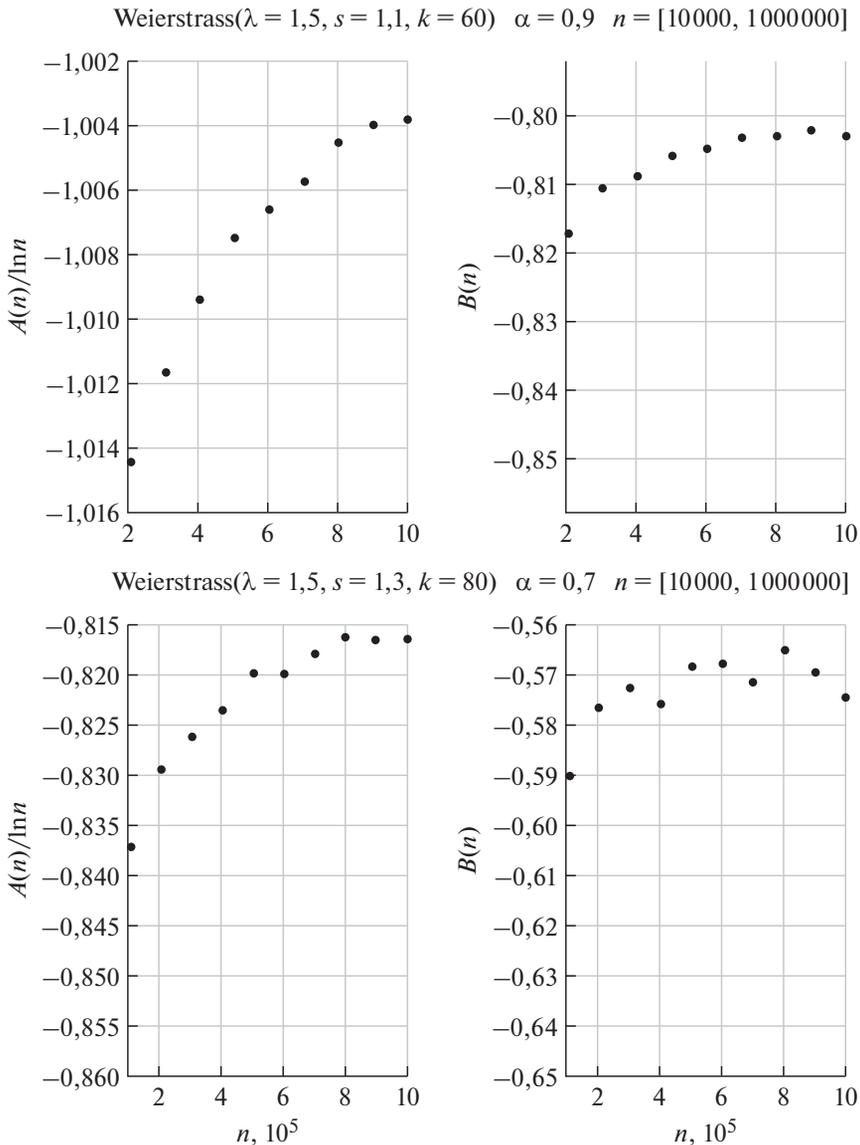


Рис. 4. Результаты вычислений для функции Вейерштрасса со значениями $s = 1,1$ и $s = 1,3$.

$A(n)/\ln(n)$ рассмотренного объема точек недостаточно для достижения сходимости к показателю $-p$ согласно (1).

4.3. Винеровский процесс

Моделирование винеровского процесса отличается наличием случайной компоненты, поэтому в данном случае оценивается не величина остаточной суммы ряда, а ее дисперсия. Итак, согласно формуле для винеровского про-

цесса, приведенной в предыдущем разделе, сумма всех слагаемых, начиная после некоторого номера k_0 , равна

$$S_{rest} = \sqrt{\frac{2}{\pi}} \sum_{k=k_0+1}^{\infty} C_k \frac{\sin(kt)}{k},$$

где $C_k \propto \mathcal{N}(0, 1)$ — независимые случайные величины со стандартным нормальным распределением.

Тогда дисперсия данной суммы выражается следующим образом:

$$\begin{aligned} \mathcal{D}[S_{rest}] &= \mathcal{D} \left[\sqrt{\frac{2}{\pi}} \sum_{k=k_0+1}^{\infty} C_k \frac{\sin(kt)}{k} \right] = \frac{2}{\pi} \sum_{k=k_0+1}^{\infty} \mathcal{D} \left[C_k \frac{\sin(kt)}{k} \right] = \\ &= \frac{2}{\pi} \sum_{k=k_0+1}^{\infty} \frac{\sin^2(kt)}{k^2} \mathcal{D}[C_k] = \frac{2}{\pi} \sum_{k=k_0+1}^{\infty} \frac{\sin^2(kt)}{k^2} \leq \frac{2}{\pi} \sum_{k=k_0+1}^{\infty} \frac{1}{k^2} = \\ &= \frac{2}{\pi} \left(\frac{\pi^2}{6} - \sum_{k=1}^{k_0} \frac{1}{k^2} \right) = \frac{\pi}{3} - \frac{2}{\pi} \sum_{k=1}^{k_0} \frac{1}{k^2}, \end{aligned}$$

и тогда

$$\delta = 3 * \text{std}(S_{rest}) = 3\sqrt{\mathcal{D}[S_{rest}]} \leq \sqrt{\frac{\pi}{3} - \frac{2}{\pi} \sum_{k=1}^{k_0} \frac{1}{k^2}}.$$

В силу гауссовости исследуемого процесса именно величина δ , вычисляемая как трехкратное среднеквадратическое отклонение остатка ряда, используется для определения количества слагаемых при усечении ряда. Так, например, δ принимает значения: $2,4 \times 10^{-2}$ при использовании $k_0 = 10\,000$ слагаемых, $3,4 \times 10^{-3}$ — при $k_0 = 500\,000$ и $7,5 \times 10^{-4}$ — при $k_0 = 10$ млн. слагаемых. Разумным допущением будет выбор k_0 таким образом, чтобы трехкратное стандартное отклонение остаточной суммы оказалось как минимум на 2 порядка ниже погрешности аппроксимации при вычислении коэффициентов сложности.

Итак, при моделировании использовалось значение $k_0 = 10^7$, кроме того, в силу случайности процесса все результаты усреднялись по реализациям процесса в количестве $n_{mc} = 100$. Пример реализации продемонстрирован на рис. 5, а результаты вычислений представлены на рис. 6.

Результаты расчетов демонстрируют чувствительность технологии вычисления коэффициентов сложности к типу процесса. Как и в случае с функцией Вейерштрасса, в данном примере коэффициенты сложности сходятся к некоторому значению, отличающемуся от -1 . Согласно графику на рис. 6 коэффициент B принимает значения около $-0,405$ вместо теоретического значения $-0,5$, соответствующего гёльдеровой экспоненте для винеровского процесса.

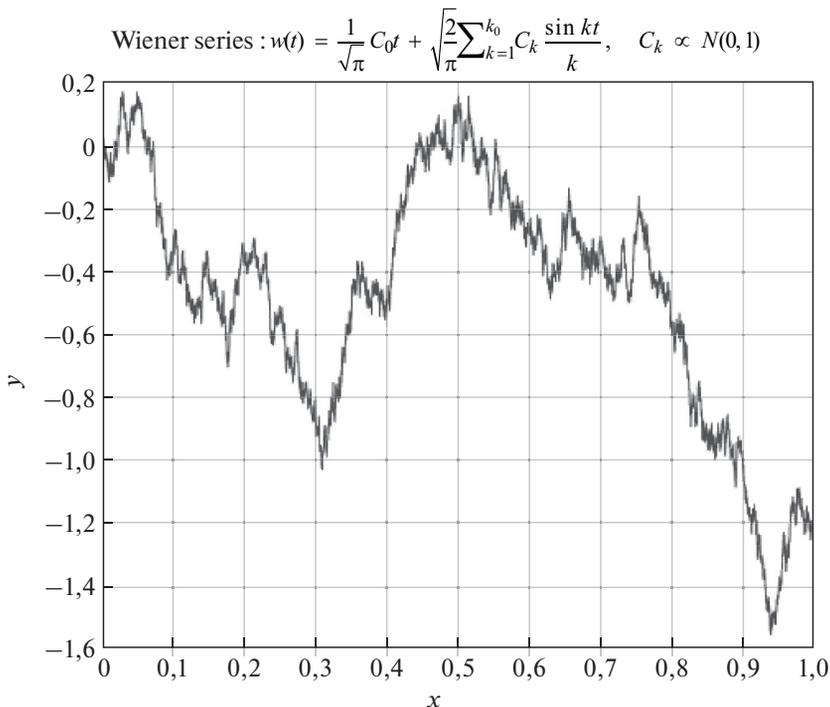


Рис. 5. Пример моделирования винеровского процесса.

Wiener series : $w(t) = \frac{1}{\sqrt{\pi}} C_0 t + \sqrt{\frac{2}{\pi}} \sum_{k=1}^{k_0} C_k \frac{\sin kt}{k}$, $C_k \propto N(0, 1)$ $n = [1000, 10\ 000]$

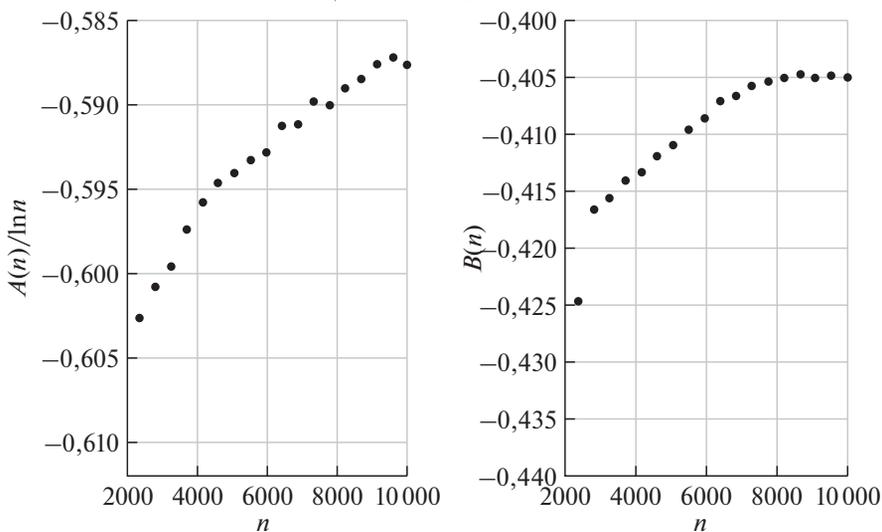


Рис. 6. Результаты вычислений для винеровского процесса.

Таким образом, в рассмотренных примерах погрешность экспериментального вычисления гёльдеровой экспоненты составила от 12 до 19%. Поскольку проверяемая гипотеза имеет предельный характер, то, очевидно, повысить точность вычислений позволит увеличение объема точек n . Кроме того, как было указано выше, в экспериментах в качестве методов аппроксимации использовалась лишь кусочно-полиномиальная аппроксимация различных порядков. Поэтому расширение данного списка (скажем, сплайнами, вейвлет преобразованиями и пр.), приведет к более точному соответствию расчетов с теорией.

4.4. Траектории дискретных динамических систем

В данной серии экспериментов траектории дискретных динамических систем анализировались с помощью коэффициентов эpsilon сложности. В отличие от предыдущих экспериментов в данном случае представляют интерес не предельные значения коэффициентов сложности, а разброс их значений при увеличении длины последовательности.

Рассмотрим две системы:

- 1) классическая авторегрессионная модель со скользящим средним $ARMA(p, q)$ с параметрами $p = 3$ и $q = 2$ (рис. 7);
- 2) детерминированная последовательность хаотического типа, так называемая логистическая карта (рис. 8):

$$x_{k+1} = 4x_k(1 - x_k), \quad x_0 = 0,2.$$

В случае 2) вид траектории существенно зависит от начального значения, что является одним из свойств хаотических последовательностей. Поэтому при моделировании пропускались первые 1000 значений последовательности прежде, чем приступить к вычислению коэффициентов сложности.

Здесь результаты вычислений удобнее представить в виде таблицы, чем графиков. В таблице для каждого из рассмотренных процессов приведены средние значения коэффициентов сложности и их среднеквадратическое отклонение при варьировании длины последовательности.

В результате вычислений было установлено, что относительная величина стандартного отклонения обоих коэффициентов уменьшается от 1,7–2% до

Результаты расчетов для последовательностей

n	100–1000		1000–10 000		10 000–100 000	
Назв. послед-ти	A_{mean} A_{std}	B_{mean} B_{std}	A_{mean} A_{std}	B_{mean} B_{std}	A_{mean} A_{std}	B_{mean} B_{std}
$ARMA(3, 2)$	–0,4420 0,0071	–0,1856 0,0039	–0,4458 0,0030	–0,1880 0,0015	– –	– –
<i>Logistic</i>	–1,0199 0,0173	–0,0223 0,0121	–1,0432 0,0089	–0,0247 0,0048	–1,0384 0,0031	–0,0263 0,0014

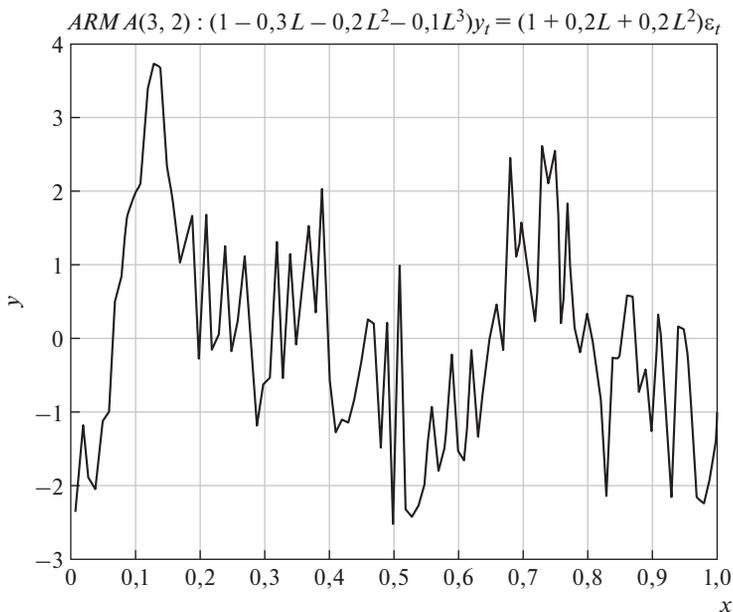


Рис. 7. Пример стохастической последовательности.

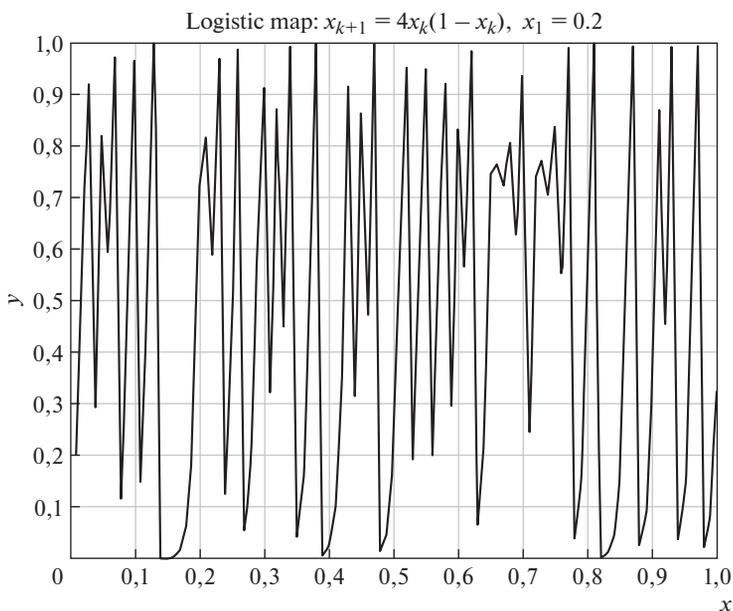


Рис. 8. Пример последовательности хаотического типа.

0,7–0,8% уже при достижении 10 тыс. точек. для модели *ARMA*, а для хаотической последовательности — до 0,3% для коэффициента *A* и 5,3% для коэффициента *B* при достижении 100 тысяч точек.

Отметим, что статистики среднего и стандартного отклонения в данном случае вычисляются не по различным реализациям случайного процесса, а по различным длинам последовательности в указанных диапазонах. Именно такой эксперимент имеет практический смысл для анализа реальных данных, когда неизвестны модель процесса и тем более истинное значение гёльдеровой экспоненты для него.

5. Заключение

В работе приводятся описание метода оценивания гёльдеровой экспоненты на базе концепции ϵ -сложности непрерывных функций и соответствующие вычислительные эксперименты. Эта концепция согласуется с общей идеей А.Н. Колмогорова о том, как следовало бы измерять «сложность» тех или иных объектов. Для функций гёльдера класса ϵ -сложность эффективно описывается так называемыми *коэффициентами сложности*. Использование коэффициентов сложности открывает возможность оценивания гёльдеровой экспоненты, которая широко применяется при анализе данных фрактального типа в самых разных приложениях. В частности, гёльдерова экспонента практически совпадает с константой Хёрста для фрактального броуновского движения, которое используется для моделирования многих физических и финансовых процессов. Предлагаемый метод оценивания гёльдеровой экспоненты достаточно прост в реализации и не требует статистических процедур. В проведенных экспериментах коэффициенты сложности вычислялись на основе весьма бедного набора методов аппроксимации (набор таких методов составляет основу концепции ϵ -сложности непрерывных функций), и полученные результаты свидетельствуют об эффективности метода даже при таких условиях. Эффективность метода может быть существенно улучшена при расширении набора алгоритмов аппроксимации непрерывных функций по их дискретным отсчетам. Были проведены также эксперименты для оценки поведения коэффициентов сложности для траекторий дискретных динамических систем. Хотя строгой математической теории для этого случая пока нет, оказалось, что коэффициенты сложности имеют предельные значения при росте объема выборки. Этот факт позволяет применять технологию вычисления коэффициентов сложности дискретных процессов произвольной природы в скользящем окне сравнительно малого объема при решении задач сегментации и классификации данных.

СПИСОК ЛИТЕРАТУРЫ

1. Павлов А.Н., Анисченко В.С. Мультифрактальный анализ сложных сигналов // Успехи физических наук. 2007. Т. 177. С. 859–876.
2. Ширяев А.Н. Основы стохастической финансовой математики. 2016. МЦНМО.
3. Савицкий А.В. Метод оценки показателя Хёрста фрактального броуновского движения // Доклады РАН. 2019. Т. 489. № 5. С. 456–460.

4. *Falconer K.J.* Fractal Geometry: Mathematical foundations and Applications. Wiley, 2003.
5. *Ming L., Vitanyi P.* An Introduction to Kolmogorov Complexity and Its Applications (англ.). 2nd ed. Springer, 1997.
6. *Piryatinska A., Darkhovsky B., Kaplan A.* Binary classification of multichannel-EEG records based on the ϵ -complexity of continuous vector functions // Comput. Method. Program. Biomedicin. 2017. V. 152, P. 131–139.
7. *Piryatinska A., Darkhovsky B.* Retrospective change-points detection for multidimensional time series of arbitrary nature: Model-free technology based on the ϵ -complexity theory // Entropy. 2021. V. 23. No. 12. P. 1626.
8. *Darkhovsky B.S.* Estimate of the Holder Exponent Based on the ϵ -Complexity of Continuous Functions // Mathematical Notes. 2022. V. 111. No. 3–4. P. 628–631. [пер. *Дарховский Б.С.* Оценка показателя Гёльдера на основе концепции ϵ -сложности непрерывных функций // Математические заметки. 2022. Т. 111. Вып. 4. С. 620–623.]
9. *Dahan A., Dubnov Y.A., Popkov A.Y. et al.* Brief Report: Classification of Autistic Traits According to Brain Activity Recoded by fNIRS Using ϵ -Complexity Coefficients // J. Autism Dev Disord. 2020. Vol. 51. Iss. 9. P. 3380–3390.
10. *Дарховский Б.С.* О сложности и размерности непрерывных конечномерных отображений // Теория вероятностей и ее применения. 2020. Т. 65. Вып. 3. С. 479–497.
11. *Колмогоров А.Н.* Комбинаторные основания теории информации и исчисления вероятностей // Успехи математических наук. 1983. Т. 38. № 4. С. 27–36.
12. *Ито К., Маккин Г.* Диффузионные процессы и их траектории. М.: Мир, 1968.
13. *Mörters P., Peres V.* Brownian Motion. Cambridge University Press, 2010.

Статья представлена к публикации членом редколлегии Б.М. Миллером.

Поступила в редакцию 22.08.2022

После доработки 18.11.2022

Принята к публикации 30.11.2022