

УДК 517.54

НОВАЯ СПЕКТРАЛЬНАЯ МЕРА СЛОЖНОСТИ И ЕЁ ВОЗМОЖНОСТИ ПО ОБНАРУЖЕНИЮ СИГНАЛОВ В ШУМЕ

© 2024 г. А. А. Галаев^{1, *}, В. Г. Бабиков^{1, **}, П. В. Лысенко^{1, ***}, Л. М. Берлин^{1, ****}

Поступило 18.12.2023 г.
После доработки 13.06.2024 г.
Принято к публикации 05.07.2024 г.

Настоящая статья посвящена совершенствованию методов распознавания сигналов на основе информационных характеристик спектра. Установлена дискретная функция нормированного упорядоченного спектра для единичной оконной функции, входящей в ДПФ. Доказаны леммы об оценках энтропии, дисбаланса и статистической сложности при обработке временного ряда независимых гауссовских величин. Предложены новые понятия одномерной и двумерной спектральных сложностей. Полученные теоретические результаты верифицированы численными экспериментами, которые подтвердили эффективность новой информационной характеристики при детектировании сигнала в смеси с белым шумом при малых отношениях сигнал/помеха.

Ключевые слова: информационная энтропия, спектральная сложность, белый гауссовский шум

DOI: 10.31857/S2686954324040122, EDN: YUEJZI

1. ВВЕДЕНИЕ

Множество прикладных задач физики и техники от квантовой механики до астрономии связано с обнаружением слабых сигналов, детектируемых на фоне естественного шума. Причем требуется иметь возможность указывать наличие детерминированных сигналов как хаотического, так и регулярного происхождения [1, 2] без их реального выделения [3]. В основе методов индикации и детектирования слабых сигналов лежат выборочные и тестовые статистики [4, 5]. В свою очередь, для оценки одного наблюдения или реализации случайной величины используются такие понятия как информационная энтропия, расстояние Кульбака–Лейблера, дивергенция Йенсена–Шеннона и некоторые другие, которые затем применяются в качестве критериев различных оптимизационных задач распознавания, классификации, фильтрации и отвечают за существо и качество принимаемых на их основе решений [6, 7]. Такие оценки возможно проводить как во временной, так и част-

ной областях исходного временного ряда [8, 9]. В настоящей работе будет показано, что для дискретных спектров сигналов оказывается удобным использовать порядковые статистики [10], где в качестве расстояния между двумя распределениями берется метрика Вассерштайна [3, 11], и будет введено понятие спектральной сложности распределения по отношению к гауссовскому белому шуму.

2. ПОСТАНОВКА ЗАДАЧИ

В работах [6, 12] было получено, что одним из лучших критериев для задачи распознавания слабого сигнала в белом шуме является статистическая сложность (C – complexity) на основе полной вариации меры со знаком (TV – total variation):

$$C_{TV}(p) = -\frac{1}{4\log_2 N} \left(\sum_{i=1}^N p_i \log_2 p_i \right) \left(\sum_{i=1}^N \left| p_i - \frac{1}{N} \right| \right)^2, \quad (1)$$

которая связана с задачей различения двух гипотез и следует из одной разновидности леммы Неймана–Пирсона. Здесь и далее рассматриваются дискретные наборы величин $p = (p_1, \dots, p_i, \dots, p_N) \in \mathbb{R}^N$, которые по определению обладают некоторыми следующими свойствами дискретного распределения вероятностей: $\forall p_i \in [0, 1]$, $\sum_{i=1}^N p_i = 1$, а N – размер ряда

¹ Институт проблем управления им. В.А. Трапезникова Российской академии наук, Москва, Россия

* E-mail: galaev@ipu.ru

** E-mail: babikov@ipu.ru

*** E-mail: pavellysen@ipu.ru

**** E-mail: berlin.lm@phystech.edu

дискретных наборов данных. Там же было показано, что максимальное значение C_{TV} достигается на распределении, состоящем всего из двух значений с оптимальным количеством отсчетов для каждого из этих значений.

Выражение (1) было предложено при условии равномерности энергетического спектра белого шума. В реальности при применении ДПФ (дискретного преобразования Фурье) к некоторому временному интервалу (далее временные интервалы будем называть фреймами) белого шума полученный спектр не является равномерным и, как следствие, уровень статистической сложности C_{TV} значительно отличается от нуля, но при этом этот критерий остается эффективным для задачи распознавания, что будет показано далее.

Стоит сказать, что распределение статистической сложности хорошо аппроксимируется кривой нормального распределения. С ростом N плотность распределения статистической сложности для гауссовского белого шума становятся уже, стандартное отклонение с ростом N уменьшается пропорционально $1/\sqrt{N}$, а значение математического ожидания монотонно увеличивается.

При появлении сигнала в шуме значение математического ожидания статистической сложности C_{TV} увеличивается и возможность обнаружения сигнала зависит от функции распределения C_{TV} как для шума, так и для смеси сигнал-шум. Таким образом, нам необходимо получить аналитические оценки функции распределения статистической сложности, и решить задачу исследования свойств спектральных энергетических распределений, которая формулируется следующим образом.

Задача 1. Пусть имеется реализация $\{x_1, \dots, x_{2N+2}\}$ последовательности независимых случайных величин $\{\xi_1, \dots, \xi_{2N+2}\}$ с нулевым математическим ожиданием, к которой применено дискретное преобразование Фурье (ДПФ)

$$X_k = \sum_{n=1}^{2N+2} x_n e^{-2i\pi k(n-1)/(2N+2)}, \quad (2)$$

определяющее случайную величину

$$\Xi_k = \sum_{n=1}^{2N+2} \xi_n e^{-2i\pi k(n-1)/(2N+2)}, \quad (3)$$

где $k = 0, \dots, N$, поскольку в силу свойства симметрии ДПФ вещественного сигнала вторая половина из $N+1, \dots, 2N+1$ комплексных амплитуд спектральных отсчетов комплексно сопряжена с первой.

Требуется найти дискретную функцию нормированного упорядоченного спектрального распределения $\eta_k(N)$ как нормированное среднее для каждого k -го значения случайной величины

$$\eta_k(N) = \frac{(T I)_k}{E_X}, \quad (4)$$

где $I_k = \Xi_k \Xi_k^*$ (квадрат модуля амплитуды или энергия спектрального отсчета), E_X — половина энергии сигнала, а T — оператор упорядочивания ряда в порядке невозрастания, и исследовать свойства полученного распределения на различных информационные меры.

Замечание 1. Сдвиг максимального количества отсчетов временного ряда выбран для дальнейшего удобства изложения материала.

Последовательность изложения результатов статьи имеет следующий вид. Сначала на основе известных результатов о распределении случайных отсчетов энергетического спектра белого шума, доказывается лемма о его упорядоченном дискретном распределении. Затем доказывается несколько лемм о поведении информационных характеристик, таких как энтропия и дисбаланс, на этом распределении. Затем вводится понятие спектральной сложности, и проводится сравнение статистической и спектральной сложностей для индикации сигнала при малом отношении сигнал/шум.

3. ФОРМУЛИРОВКА РЕЗУЛЬТАТОВ

Для решения Задачи 1 нам потребуется несколько математических утверждений, сформулированных и доказанных далее.

Лемма А. [8] В Задаче 1 распределения случайных величин I_k при $k = 1, \dots, N$ являются экспоненциальными, а при $k = 0$ — нормальным распределением, если случайные величины ξ_n , $n = 1, \dots, 2N+2$ — независимые одинаково распределенные гауссовские с нулевым средним и дисперсией σ_0^2 (Шустер, 1898).

Из Леммы А следует, в частности, что при использовании прямоугольной оконной функции для ограничения интервала анализа и формирования отсчетов сигнала, плотность распределения квадратов амплитуд имеет вид

$$\rho_{I_k}(y) = \frac{1}{N\sigma_0^2} \exp\left(-\frac{y}{N\sigma_0^2}\right), y \geq 0. \quad (5)$$

Лемма 1. Пусть y_1, \dots, y_N — результаты наблюдений случайной величины Y , имеющей экспоненциальное распределение $F(y) = 1 - \exp(-y)$. Тогда

значения последовательности $y_{(1)}, \dots, y_{(N)}$ тех же результатов, но расположенных в порядке убывания, которые в свою очередь являются обратным вариационным рядом, где случайная величина $Y_{(k)}$ является невозрастающей k -ой порядковой статистикой, удовлетворяют следующему условию:

$$E[\exp(-Y_{(k)})] = \frac{k}{N+1},$$

$$D[\exp(-Y_{(k)})] = \frac{k(N-k+1)}{(N+1)^2(N+2)}.$$

Доказательство. Согласно [5] в выборке объема N из генеральной совокупности с функцией распределения $F(y)$, обладающей непрерывной плотностью $f(y)$, распределение G_k случайной величины $Y_{(k)}$ выражается формулой

$$dG_k = \frac{(F(Y_{(k)}))^{N-k}(1-F(Y_{(k)}))^{k-1}}{B(N-k+1, k)} dF(Y_{(k)}),$$

где $B(N-k+1, k) = \frac{(N-k)!(k-1)!}{N!}$ – Бета-функция.

Согласно условиям леммы распределение $F(y)$ – экспоненциальное, значит $dF = e^{-y} dy$. Сделаем замену $\Theta = \exp(-Y)$. Отсюда получаем следующее выражение

$$G_k \Big|_0^1 = \int_0^1 \frac{(1-\theta)^{N-k} \theta^{k-1}}{B(N-k+1, k)} d\theta = 1.$$

Далее найдем математическое ожидание случайной величины Θ :

$$E(\Theta_{(k)}) = \int_0^1 \frac{(1-\theta)^{N-k} (\theta)^k}{B(N-k+1, k)} d\theta = \frac{k}{N+1}.$$

Итак, можно записать, что справедливо равенство

$$E[\exp(-Y_{(k)})] = \frac{k}{N+1}.$$

Теперь найдем математическое ожидание случайной величины Θ^2 :

$$E(\Theta_{(k)}^2) = \int_0^1 \frac{(1-\theta)^{N-k} (\theta)^{k+1}}{B(N-k+1, k)} d\theta = \frac{k(k+1)}{(N+1)(N+2)}.$$

Тогда для дисперсии $D(\Theta_{(k)}) = E(\Theta_{(k)}^2) - (E(\Theta_{(k)}))^2$ справедлива формула

$$D[\exp(-Y_{(k)})] = \frac{k(N-k+1)}{(N+1)^2(N+2)},$$

вид зависимости которой для различных значений N приведен на рис. 1.

Лемма 2. Пусть справедлива Лемма 1. Тогда нормированное упорядоченное распределение спектра имеет вид

$$n_k(N) = -\frac{1}{N \cdot K_N} \ln \frac{k}{N+1},$$

где $K_N = -\frac{1}{N} \sum_{k=1}^N \ln \frac{k}{N+1}.$ (6)

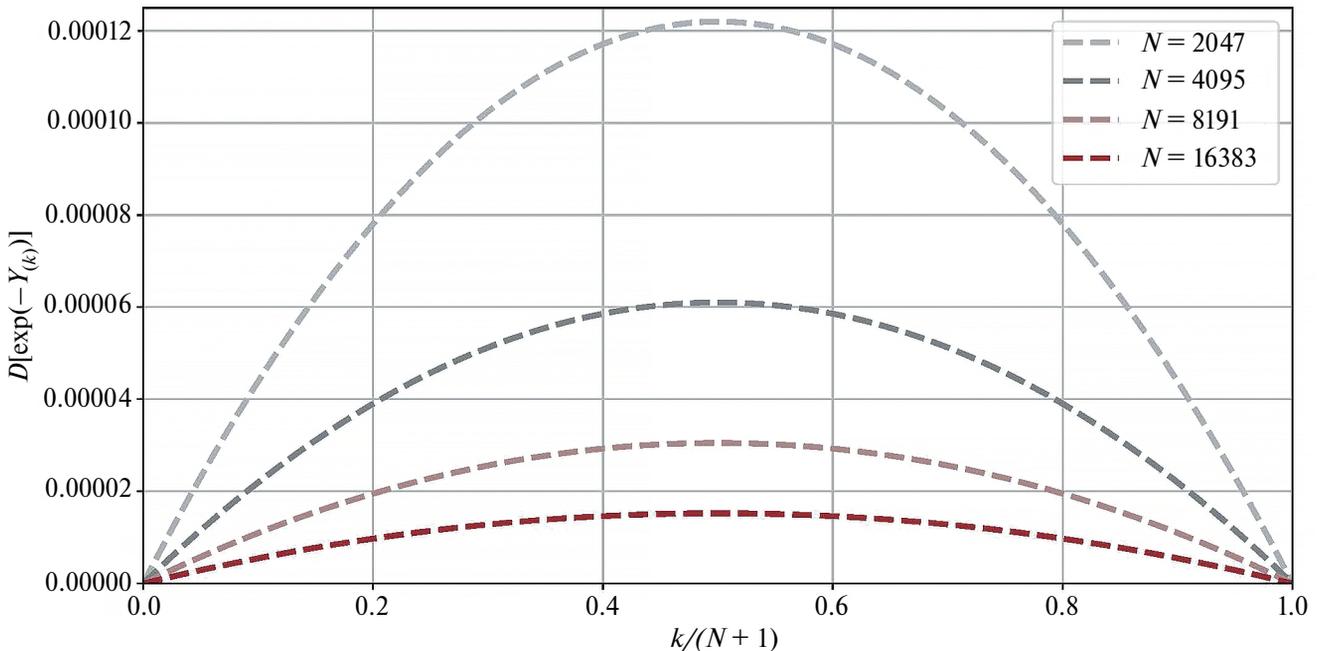


Рис. 1. Графики дисперсии случайной величины $\exp(-Y_{(k)})$ для разных значений k, N .

Доказательство. Воспользуемся формулой (5), заметив, что $2E_X = 2N\sigma_0^2$ соответствует полной энергии сигнала. Тогда

$$\rho_{n_k(N)}(y) = \exp(-y), y \geq 0.$$

Нормированный спектр плотности распределения квадратов амплитуд будет подчиняться экспоненциальному распределению, а по Лемме 1 его упорядоченная дискретная зависимость от номера отсчета в спектре будет определяться так: $\frac{k}{N+1} \sim e^{-n_k(N)}$, т.е. $n_k(N) \sim -\ln \frac{k}{N+1}$. Откуда требуется найти только нормировочный коэффициент K_N .

Нормировочный коэффициент $K_N = -\frac{1}{N} \sum_{k=1}^N \ln \frac{k}{N+1}$ может быть оценен при помощи формулы Стирлинга, причем

$$\begin{aligned} \lim_{N \rightarrow \infty} K_N &= \lim_{N \rightarrow \infty} -\frac{1}{N} \sum_{k=1}^N \ln \frac{k}{N+1} = \\ &= \lim_{n \rightarrow \infty} -\frac{1}{N} \ln \frac{N!}{(N+1)^N} = \\ &= \lim_{n \rightarrow \infty} -\frac{N \ln N - N + \mathcal{O}(\ln N) - N \ln(N+1)}{N} = 1. \end{aligned}$$

Для оценки скорости сходимости в табл. 1 приведены значения K_N и его непрерывного аналога

$$\tilde{K}_N = -\frac{N+1}{N} \int_{\frac{1}{N+1}}^1 \ln x dx.$$

Как видно из табл. 1 достаточно близкой оценкой суммы $K_N = -\sum_{k=1}^N \frac{1}{N} \ln \frac{k}{N+1}$ является ин-

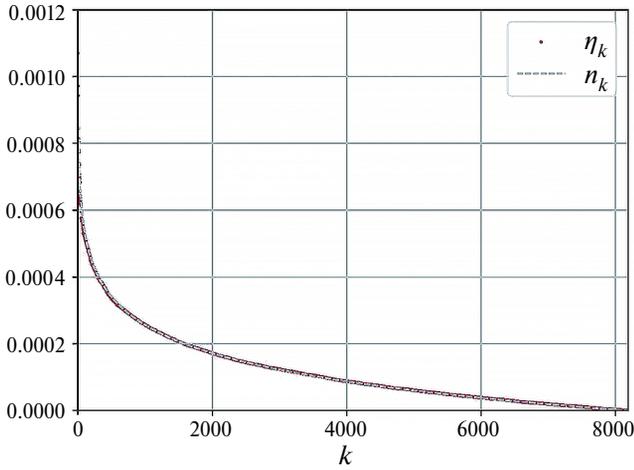
теграл $\tilde{K}_{2N} = -\frac{2N+1}{2N} \int_{\frac{1}{2N+1}}^{\frac{2N}{2N+1}} \ln x dx = -\ln 2N +$
 $+ \left(1 - \frac{1}{2N}\right)(1 + \ln(1 + 2N)) = 1 - \frac{\ln 2N}{2N} + \mathcal{O}\left(\frac{\ln N}{N}\right)$
 который в свою очередь есть предел сумм Дарбу. Поэтому исходная сумма близко оценивается интегралом.

Леммы 1 и 2 определяют распределение нормированного упорядоченного спектра $n_k(N)$ и отклонения амплитуд от нее, которые быстро убывают с ростом N . Поэтому в модельном эксперименте энтропию белого шума предлагается оценивать по энергетическому спектру $e_k(N)$ ($k \in 1, \dots, N$), который предварительно нормируется, так чтобы сумма амплитуд спектра по всем частотам равнялась единице $\sum_{k=1}^N e_k(N) = 1$. Упорядочив получившийся спектр по амплитудам, получаем некоторую зависимость энергии $\tilde{e}_k(N)$ соответствующей части спектра от номера k этой части в упорядоченном массиве ($k = 1, \dots, N$), а именно

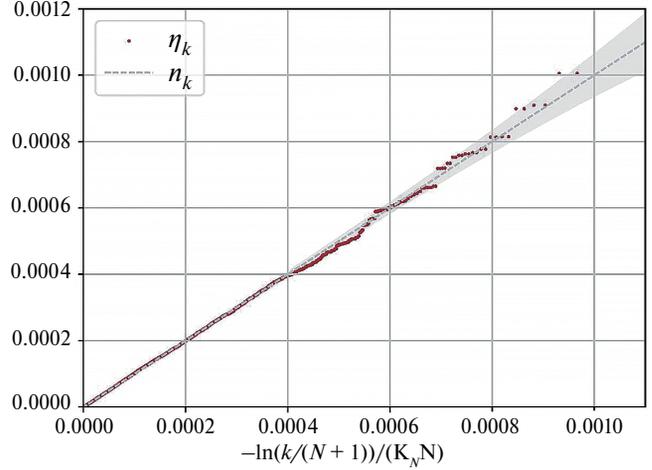
$$\tilde{e}_k(N) = \frac{(TJ)_k}{\sum_{i=1}^N J_n} = \frac{\eta_k(N)E_X}{\sum_{i=1}^N \eta_i(N)E_X} = \eta_k(N). \quad (7)$$

Таблица 1. Значения нормировочных коэффициентов K_N, \tilde{K}_N и $1 - K_N, 1 - \tilde{K}_N$ для различных значений N .

$N + 1$	$\log_2(N + 1)$	K_N	$1 - K_N$	$-\log_2(1 - K_N)$	\tilde{K}_N	$1 - \tilde{K}_N$	$-\log_2(1 - \tilde{K}_N)$
2	1	0.75204	0.24796	2.01	0.35616	0.64384	0.64
4	2	0.81492	0.18508	2.43	0.57078	0.42922	1.22
8	3	0.87165	0.12835	2.96	0.71813	0.28187	1.83
16	4	0.91622	0.08378	3.58	0.82105	0.17895	2.48
32	5	0.94782	0.05218	4.26	0.89026	0.10974	3.19
64	6	0.96863	0.03137	4.99	0.93465	0.06535	3.94
128	7	0.98164	0.01836	5.77	0.96200	0.03800	4.72
256	8	0.98948	0.01052	6.57	0.97832	0.02168	5.53
512	9	0.99406	0.00594	7.40	0.98781	0.01219	6.36
1024	10	0.99669	0.00331	8.24	0.99323	0.00677	7.21
2048	11	0.99818	0.00182	9.10	0.99628	0.00372	8.07
4096	12	0.99900	0.00100	9.97	0.99797	0.00203	8.94
8192	13	0.99946	0.00054	10.85	0.99890	0.00110	9.83
16384	14	0.99971	0.00029	11.75	0.99941	0.00059	10.72



(а) Упорядоченный спектр $\eta_k(N)$ и дискретное распределение $n_k(N)$ одной реализации фрейма длины $N + 1 = 8192$.



(б) Те же распределения, горизонтальная ось в логарифмическом масштабе (показаны стандартные отклонения).

Рис. 2. Пояснение к Леммам 1 и 2.

На рис. 2 видно, что даже в одном фрейме упорядоченные амплитуды спектра лежат на аналитическом распределении вида (6).

Поскольку в (1) $C_{TV}(p) = S(p)D_{TV}(p)$, где $S(p)$ – энтропия, а $D_{TV}(p)$ – дисбаланс, то в целях детектирования наличия сигнала требуется оценить энтропию энергетического спектра белого шума и его дисбаланс. Дальнейшие аналитические оценки помогут определить качество положения этих величин на кривых, получаемых в вычислительном эксперименте.

Лемма 3. На спектральном распределении вида (6) энтропия равна

$$S(N) = \frac{\ln(NK_N)}{\ln N} + \frac{1}{NK_N \ln N} \sum_{k=1}^N \ln \frac{k}{N+1} \left(\ln \left(-\ln \frac{k}{N+1} \right) \right). \quad (8)$$

Ее приближенную оценку $\tilde{S}(N)$ можно получить следующим образом:

$$\tilde{S}(N) = 1 + \frac{\gamma - 1}{\ln N} + \mathcal{O}\left(\frac{1}{N \ln N}\right), \quad (9)$$

где γ – постоянная Эйлера-Маскерони.

При этом для стандартного отклонения энтропии справедлива оценка

$$D_{TV} = \frac{1}{4N^2} \left(-2\lceil Ne^{-KN} \rceil + N + \frac{-2\ln\lceil Ne^{-KN} \rceil! + 2\lceil Ne^{-KN} \rceil \ln(N+1)}{K_N} + \frac{\ln N! - N \ln(N+1)}{K_N} \right)^2. \quad (11)$$

При больших N на распределении вида (6) его можно оценить следующим выражением

$$\tilde{D}_{TV}(N) = \frac{1}{e^2} + \mathcal{O}\left(\frac{\ln N}{N}\right). \quad (12)$$

Таблица 2. Расчет энтропии \tilde{S} по формуле (9). Оценка энтропии S^* при помощи численного эксперимента (10 000 испытаний) для белого шума. Расчет стандартного отклонения энтропии $\Delta\tilde{S}$ по формуле (10). Оценка стандартного отклонения энтропии ΔS^* при помощи численного эксперимента (10 000 испытаний) для белого шума.

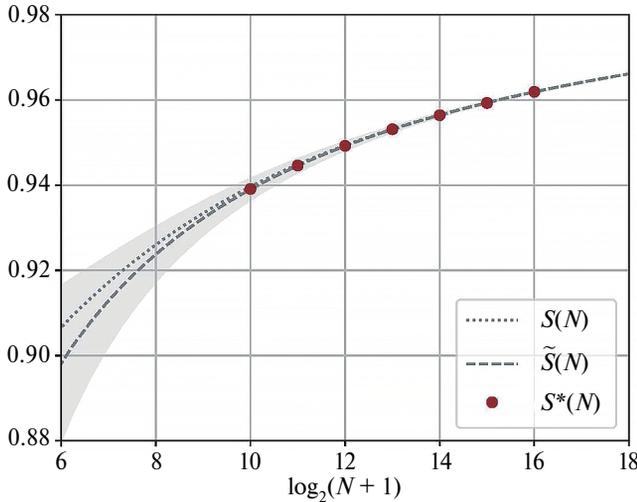
$N + 1$	\tilde{S}	S^*	$\Delta\tilde{S}$	ΔS^*
1024	0.9390	0.9390	0.00250	0.00247
2048	0.9446	0.9446	0.00159	0.00157
4096	0.9492	0.9492	0.00102	0.00101
8192	0.9531	0.9531	0.00066	0.00066
16384	0.9564	0.9564	0.00043	0.00043

$$\Delta\tilde{S} = \frac{\ln 2N}{2\sqrt{N} \ln^2 N} + \mathcal{O}\left(\frac{\ln N}{N^{\frac{5}{2}}}\right). \quad (10)$$

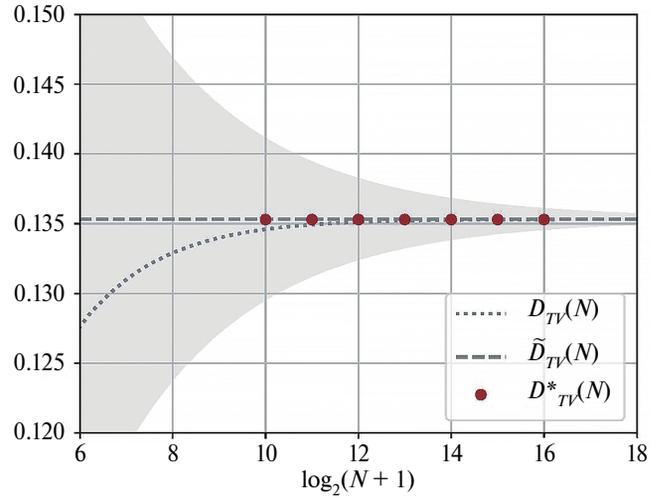
Доказательство Леммы носит технический характер.

Лемма 4. На распределении (6) дисбаланс

$$D_{TV} = \frac{1}{4} \left(\sum_{k=1}^N p_k - \frac{1}{N} \right)^2 \text{ принимает вид}$$



(а) Энтропия S



(б) Дисбаланс D_{TV}

Рис. 3. Зависимость математических ожиданий и стандартных отклонений энтропии и критерия разности от $\log_2 N$ (иллюстрация Леммы 3 и Леммы 4).

Таблица 3. Расчет \tilde{D}_{TV} по формуле (12). Оценка D_{TV}^* при помощи численного эксперимента (10 000 испытаний) для белого шума. Расчет стандартного отклонения энтропии $\Delta \tilde{D}_{TV}$ по формуле (13). Оценка стандартного отклонения энтропии ΔD_{TV}^* при помощи численного эксперимента (10 000) для белого шума.

$N+1$	\tilde{D}_{TV}	D_{TV}^*	$\Delta \tilde{D}_{TV}$	ΔD_{TV}^*
1024	0.1353	0.1354	0.0057	0.0057
2048	0.1353	0.1353	0.0041	0.0040
4096	0.1353	0.1354	0.0029	0.0028
8192	0.1353	0.1353	0.0020	0.0020
16384	0.1353	0.1353	0.0014	0.0014

Стандартное отклонение для дисбаланса определяется как

$$\Delta \tilde{D}_{TV} = \frac{1}{2e\sqrt{N}} + \mathcal{O}\left(\frac{\ln N}{N^{\frac{3}{2}}}\right) \quad (13)$$

Заметим, что $\lim_{N \rightarrow \infty} D_{TV}(N) = e^{-2}$. На рис. 3б показан этот уровень ~ 0.13534 .

Опираясь на распределение (6) введем новые понятия спектрального дисбаланса и спектральной сложности.

Определение 1. Спектральным дисбалансом относительно спектрального распределения (6) назовем

$$D_S(p) = \frac{1}{4} \left(\sum_{k=1}^N \left| p_k + \frac{1}{N \cdot K_N} \ln \frac{k}{N+1} \right| \right)^2, \quad (14)$$

а спектральной сложностью величину

$$C_S(p) = -\frac{1}{4 \log_2 N} \left(\sum_{k=1}^N p_k \log_2 p_k \right) \times \left(\sum_{k=1}^N \left| p_k + \frac{1}{N \cdot K_N} \ln \frac{k}{N+1} \right| \right)^2. \quad (15)$$

В случае единственности смены знака выражения под модулем в выражении (15), т.е. при

$$p_k + \frac{1}{NK_N} \ln \frac{k}{N+1} < 0, k \in 1, \dots, Z, \quad (16)$$

$$p_k + \frac{1}{NK_N} \ln \frac{k}{N+1} \geq 0, k \in Z+1, \dots, N,$$

формулу (14) для вычисления спектрального дисбаланса можно привести к виду

$$D_S(p) = \frac{1}{4} \left(-\sum_{k=1}^Z \left(p_k + \frac{\ln \frac{k}{N+1}}{NK_N} \right) + \sum_{k=Z+1}^N \left(p_k + \frac{\ln \frac{k}{N+1}}{NK_N} \right) \right)^2 = \frac{1}{4} \left(\left(-\sum_{k=1}^Z p_k + \sum_{k=Z+1}^N p_k \right) + \frac{1}{NK_N} \left(\ln \frac{N!}{Z!(N+1)^{N-Z}} - \ln \frac{Z!}{(N+1)^Z} \right) \right)^2, \quad (17)$$

что аналогично выражению (45) из [6], которое является точной аналитической формулой при определении максимума C_{TV} .

Для больших N формулу (17) можно упростить до вида

$$\begin{aligned} \tilde{D}_S(p) = & \frac{1}{4} \left(\left(-\sum_{k=1}^Z p_k + \sum_{k=Z+1}^N p_k \right) + \right. \\ & \left. + \frac{1}{NK_N} \left(-N - 1 + 2Z \ln \left(\frac{e(N+1)}{Z} \right) \right) \right)^2. \end{aligned} \quad (18)$$

Определение 2. Спектральным дисбалансом двумерного распределения $p = \{p_{ki} = p_k^2 p_i, k = 1, \dots, N, i = 1, \dots, M\}$ двух независимых случайных величин относительно двумерного спектрального распределения размерности $N \times M$, порожденного распределениями вида (6), назовем

$$\begin{aligned} D_{S^2}(p) = & \frac{1}{4} \left(\sum_{k=1}^N \sum_{i=1}^M |p_k^2 p_i - \right. \\ & \left. - \frac{1}{NMK_N K_M} \ln \frac{k}{N+1} \ln \frac{i}{M+1} \right)^2, \end{aligned} \quad (19)$$

а спектральной сложностью величину

$$C_{S^2}(p) = S(p) D_{S^2}(p). \quad (20)$$

Например, двумерная спектральная сложность может быть вычислена по двум последовательным фреймам или двум независимым каналам приема сигнала.

4. СРАВНЕНИЕ

ИНФОРМАЦИОННЫХ МЕР C_{TV} И C_S

В данном разделе приводится статистическое сравнение предложенной авторами информационной характеристики C_S (15) со стандартной статистической сложностью C_{TV} (1) как критериев в задаче обнаружения полезного детерминированного сигнала s_n , которая традиционно сводится к задаче различения двух гипотез о принятой последовательности данных

$$\begin{cases} \Gamma_0 : x_n = w_n, \\ \Gamma_1 : x_n = s_n + w_n, n = 1, \dots, N. \end{cases} \quad (21)$$

Гипотеза Γ_0 связана с принятием решения о приеме только шума, а гипотеза Γ_1 – о приеме смеси полезного сигнала и шума, где последовательность $\{x_n\}, n = 1, \dots, N$ – временной ряд из принятых данных, $\{s_n\}$ – полезный сигнал, $\{w_n\}$ – аддитивный белый гауссовский шум, N – длина временного ряда данных (фрейма).

Для проверки качества обнаружения полезного детерминированного сигнала была набрана статистика на $Q = 10000$ численно сгенериро-

ванных фреймах $\{x_n\}$ сигнально-шумовой смеси длины $N + 1 = 16384$. Во всех реализациях сигнал $\{s_n\}$ оставался одним и тем же, а именно фиксированным по количеству и амплитудам набором синусоид со случайными фазами. Аддитивный белый гауссовский шум $\{w_n\}$ был получен генератором гауссовских последовательностей с одними и тем же параметрами $\mu = 0$ и σ (в рамках одного набора Q фреймов).

Для каждой полученной последовательности $\{x_n\}$ вычисляются коэффициенты БПФ X_k (2) (количество коэффициентов в спектральном разложении также равно N), по которым затем считается нормированный упорядоченный дискретный набор величин p_k , построение которого подробно описано в [6, 12]. Далее на основании p_k вычисляются значения C_{TV} и C_S для шума и смеси шума с сигналом, отвечающим двум гипотезам из выражений (21).

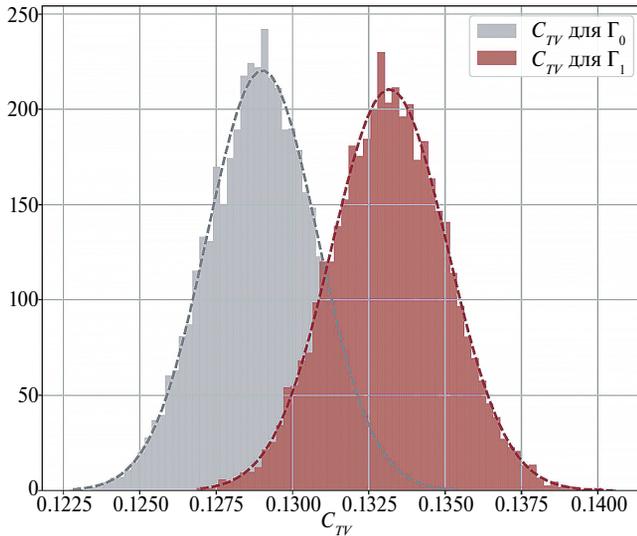
Основной количественной характеристикой “зашумленности” сигнала является отношение сигнал/шум – SNR (Signal-To-Noise Ratio), которое описывается формулой

$$SNR = 10 \log_{10} \left(\frac{E_{signal}}{E_{noise}} \right), \quad (22)$$

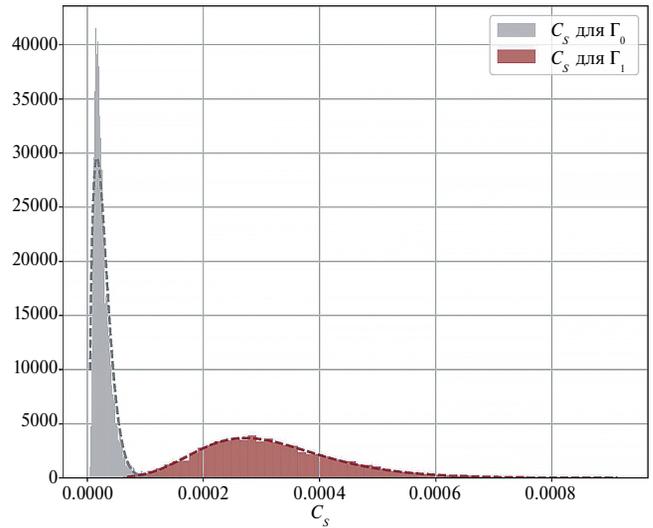
где E_{signal} , E_{noise} – полные энергии сигнала и шума соответственно, рассчитанные как сумма мощностей спектрального разложения последовательностей $\{s_n\}$ и $\{w_n\}$.

На рис. 4 показаны результаты набранной статистики для среднего значения $SNR = -15$ dB. Как видно, качество разделения шума и смеси сигнала с шумом возрастает при использовании новой функции, так как носители распределений C_S в этом случае практически не пересекаются, в отличие от распределений C_{TV} . Количественно данный эффект отражен на графиках ROC [17] (рабочая характеристика приемника) на рис. 5. Метрика $AUC_{ROC} \in [0,1]$ (площадь под ROC-кривой) позволяет оценить вероятность, с которой классификатор сможет разделить классы. Чем она выше, тем эффективнее критерий бинарной классификации. Из графика на рис. 5 следует, что значение AUC_{ROC} при том же значении SNR для спектральной сложности C_S выше, чем у классической статистической сложности C_{TV} (0.99 против 0.94).

Для получения более полной картины сравнения информационных характеристик требуется провести описанную выше процедуру сбора и анализа статистики для разных значений SNR . Только тогда можно быть уверенным в эффективности новой функции. На рис. 6 приведены зависимости AUC для двух рассматриваемых

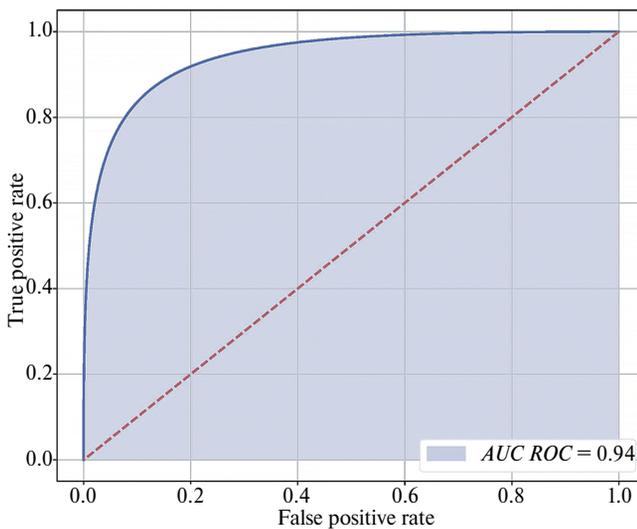


(а) Гистограмма распределения значений C_{TV}

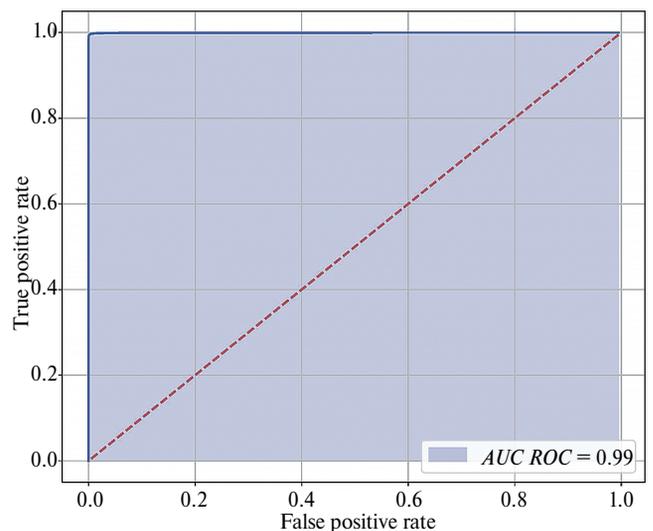


(б) Гистограмма распределения значений C_S

Рис. 4. Гистограммы информационных характеристик для среднего значения $SNR = -15$ dB, полученные для $Q = 10000$ реализаций сигнално-шумовой смеси.



(а) ROC-кривая для C_{TV}



(б) ROC-кривая для C_S

Рис. 5. ROC-кривые информационных характеристик для среднего значения $SNR = -15$ dB, полученные для $Q = 10000$ реализаций сигнално-шумовой смеси.

информационных характеристик в зависимости от величины SNR . Видно, что для значений SNR от -15 dB до -20 dB C_S демонстрирует устойчивое преимущество над C_{TV} . Значение в -20 dB является граничным для данного эксперимента, поскольку ни один из рассмотренных критериев уже не позволяет сколь-нибудь эффективно разделить классы (значения $AUC ROC$ меньше ≈ 0.65). Что подтверждает эффективность и обоснованность применения спектральной сложности C_S для индикации появления сигнала в шумовой смеси.

5. ЗАКЛЮЧЕНИЕ

Таким образом, можно сделать вывод об эффективности как статистической сложности, так и новой меры, названной авторами спектральной сложностью (15), в задачах обнаружения (предобнаружения) детерминированных сигналов в гауссовском шуме в одном фрейме, что является основой для решения последующих задач классификации и распознавания. Полученные теоретические утверждения позволяют точно оценивать значения информационных характеристик без полного расчета по реальным распре-

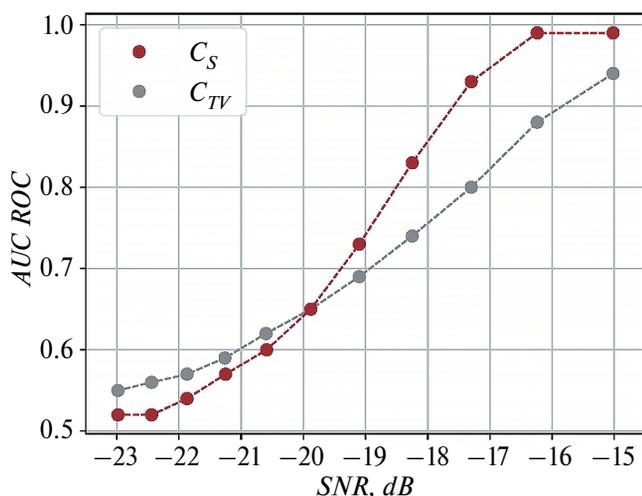


Рис. 6. Зависимость $AUC ROC$ от SNR .

делениям, что является важным для прикладных задач детектирования.

Введенная двумерная спектральная сложность относительно белого шума (20) является перспективной в теоретическом и прикладном аспектах исследования в рассматриваемых задачах, что будет предметом дальнейших научных разработок.

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Исследование выполнено за счет гранта Российского научного фонда (проект № 23-19-00134) в ИПУ РАН им. В. А. Трапезникова.

СПИСОК ЛИТЕРАТУРЫ

1. *Amigo J.M.* Ordinal methods: Concepts, applications, new developments, and challenges In memory of Karsten Keller (19612022) / J. M. Amigo, O. A. Rosso // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2023. Vol. 33, no. 8. P. 080401. <https://pubs.aip.org/cha/article/33/8/080401/2905538/Ordinal-methods-Concepts-applications-new>.
2. Distinguishing Noise from Chaos / O.A. Rosso, H.A. Larrondo, M. T. Martin et al. // *Phys. Rev. Lett.* 2007. Oct. V. 99. P. 154102. <https://link.aps.org/doi/10.1103/PhysRevLett.99.154102>.
3. *Perkey S.* Using Fourier Coefficients and Wasserstein Distances to Estimate Entropy in Time Series / S. Perkey, A. Carvalho, A. Krone-Martins // 2023 IEEE 19th International Conference on e-Science (e-Science). Limassol, Cyprus: IEEE, 2023. P. 1–2. <https://ieeexplore.ieee.org/document/10254949/>.
4. *Statistical Distributions* / C. Forbes, M. Evans, N. Hastings, B. Peacock. 1 edition. Wiley, 2010. <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470627242>.
5. *Klenke A.* Probability Theory: A Comprehensive Course / A. Klenke. Universitext. London: Springer London, 2014. <https://link.springer.com/10.1007/978-1-4471-5361-0>.
6. *Галаяев А.А.* Статистическая сложность как критерий задачи обнаружения полезного сигнала / А.А. Галаяев, П.В. Лысенко, Л.М. Берлин // *Автоматика и телемеханика*. 2023. С. 121–145.
7. *Distances in Probability Space and the Statistical Complexity Setup* / A. M. Kowalski, M. T. Mart'ın, A. Plastino et al. // *Entropy*. 2011. V. 13. №. 6. P. 1055–1075. <http://www.mdpi.com/1099-4300/13/6/1055>.
8. *Richards M.A.* The Discrete-Time Fourier Transform and Discrete Fourier Transform of Windowed Stationary White Noise / M.A. Richards // *Technical Memorandum*. 2013. P. 1–24.
9. *Kay S.M.* Fundamentals Of Statistical Processing, Volume 2: Detection Theory / S.M. Kay. Prentice-Hall signal processing series. Pearson Education, 2009. <https://books.google.ru/books?id=wwmnY9xyt9MC>.
10. *Орлов И.Я.* Оценка потерь обнаружения сигналов приемником с адаптивным порогом на основе метода порядковых статистик / И.Я. Орлов, Е.С. Фитасов // *Известия вузов. Радиофизика*. 2018. Т. 61. № 7. С. 596–604.
11. *Cazelles E.* The Wasserstein-Fourier Distance for Stationary Time Series / E. Cazelles, A. Robert, F. Tobar // *IEEE Transactions on Signal Processing*. 2021. V. 69. P. 709–721. <https://ieeexplore.ieee.org/document/9303405/>.
12. *Berlin L.M.* Comparison of Information Criteria for Detection of Useful Signals in Noisy Environments / L.M. Berlin, A.A. Galyaev, P.V. Lysenko // *Sensors*. 2023. V. 23. № 4. <https://www.mdpi.com/1424-8220/23/4/2133>.

A NEW SPECTRAL MEASURE OF COMPLEXITY AND ITS CAPABILITIES FOR DETECTING SIGNALS IN NOISE

Corresponding Member of the RAS A. A. Galyaev^a, V. G. Babikov^a, P. V. Lysenko^a, L. M. Berlin^a

^aInstitute of Control Sciences of RAS, Moscow, Russian Federation

This article is devoted to the improvement of signal recognition methods based on the information characteristics of the spectrum. A discrete function of the normalized ordered spectrum is established for a single window function included in the DFT. Lemmas on estimates of entropy, imbalance and statistical complexity in processing a time series of independent Gaussian quantities are proved. New concepts of one-dimensional and two-dimensional spectral complexities are proposed. The theoretical results obtained were verified by numerical experiments, which confirmed the effectiveness of the new information characteristic when detecting a signal mixed with white noise at low signal-to-noise ratios.

Keywords: information entropy, spectral complexity, additive white Gaussian noise