ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2024, Vol. 85, No. 12, pp. 1168–1183.

STOCHASTIC SYSTEMS

Order Statistics of the Normalized Spectral Distribution for Detecting Weak Signals in White Noise

A. A. Galyaev[*,a], **L. M. Berlin**[*,b], **P. V. Lysenko**[*,c], **and V. G. Babikov**[*,d]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a]*galaev@ipu.ru,* [b]*berlin.lm@phystech.edu,* [c]*pavellysen@ipu.ru,* [d]*vl.babikov@gmail.com*

Received September 8, 2024
Revised October 8, 2024
Accepted October 24, 2024

Abstract—This paper further develops the topic of the authors' previous research. In particular, order statistics of the discrete normalized spectral distribution of the additive white Gaussian noise are investigated to detect a deterministic useful signal in a noise mixture using information features. An additional relationship is established between the discrete spectral distribution of the single-window realization statistics of the white noise. Another novel result consists in exact formulas for calculating the mean and variance of normalized order statistics. Based on the analytical expressions derived, a new formula for calculating the spectral complexity is proposed and the already known one is refined. The theoretical results are verified by statistical simulation.

Keywords: order statistics, signal processing, Fourier transform, signal detection in noise

DOI: 10.31857/S0005117924120038

1. INTRODUCTION

The problems of detecting deterministic, chaotic, and random signals have attracted the attention of researchers since the time of studying space objects [1, 2]. When solving such problems, the main difficulty lies in the unknown properties and characteristics of the observed signal. Therefore, the periodic repetition of a signal over time is often treated as a feature of the appearance of a deterministic signal. Then, based on the results of observations, the received energy is accumulated and averaged, and the presence/absence of the desired signal is concluded accordingly [3].

The statistical foundations of the theory of signal detection and classification in noise were laid in the 1950s–1960s [4]. In turn, machine learning methods allowed solving classification problems based on extracting a set of individual features [5]. Over the last 30–40 years, research results have appeared in the scientific literature in which information characteristics of temporal or spectral distributions are used as detection and classification criteria. Since entropy or information is insensitive to the permutation of discrete samples of distributions, the gaze of the scientific community is directed to studying the distributions of order statistics and describing their properties [6, 7]. In recent years, entropy and related indicators have been frequently applied in the analysis of electroencephalography (EEG) and brain activity [8]. Another example of using signal entropy and related spectral complexity is the problem of signal classification, as shown in [5, 9]. Lately, order statistics have found applications in generative neural networks [10] and the related Wasserstein distance and theory of optimal transport [11, 12].

In many detection problems, the signal being detected is often supposed known [13]. In practice, the pre-detection problem is also popular when it is important, e.g., to determine the presence of a deterministic signal in white noise [14, 15].

This paper aims to establish statistical regularities in the single-window observations of white noise and studying the evolution of these regularities when a signal appears in the mixture. An expected result is a new tool for detecting (and, subsequently, classifying) a signal under small values of the signal-to-noise ratio (SNR).

The remainder of this paper is organized as follows. In Section 2, we present the mathematical formulation of the problem and the background necessary for further considerations. Sections 3 and 4 contain the main analytical results of this research, namely, the lemmas and propositions for finding the mean and variance of the normalized order statistics of the spectral distribution of the white noise when processing the observations of a single rectangular window. In Section 5, new information characteristics of signals are introduced based on the analytical results obtained. Section 6 describes the results of numerical simulations confirming the theoretical constructs of the previous sections. In Section 7, we summarize the outcomes of the paper.

2. PROBLEM STATEMENT AND BACKGROUND

The problem of detecting a signal $s(n)$ is traditionally reduced to that of distinguishing between two hypotheses [15]:

$$\begin{cases} \Gamma_0 : x(n) = w(n) \\ \Gamma_1 : x(n) = s(n) + w(n), \ n = 1, \ldots, 2N+2. \end{cases}$$

The hypothesis Γ_0 is associated with the decision to receive only noise while the hypothesis Γ_1 with the decision to receive a mixture of a useful signal and noise, where the sequences $\{x(n)\}$, $\{s(n)\}$, and $\{w(n)\}$, $n = 1, \ldots, 2N+2$, denote the time series of the received data, useful signal, and additive random noise, respectively, and $2N+2$ is the time series length. The random variables $(x(1), \ldots, x(n), \ldots, x(2N+2))$ of the time series take values $(x_1, \ldots, x_n, \ldots, x_{2N+2}) \in \mathbb{R}^{2N+2}$.

To estimate the probability of erroneous hypotheses distinction, we can apply a modification of the Neyman–Pearson lemma in which the error function is described by the exact formula

$$\mathcal{E}r(N; \Gamma_0, \Gamma_1) = 1 - \frac{1}{2}\|P_0^{(N)} - P_1^{(N)}\| = 1 - TV(P_0, P_1), \tag{1}$$

where $P_0^{(N)}$ and $P_1^{(N)}$ are the multivariate distribution functions of observation statistics for the hypotheses Γ_0 and Γ_1, respectively, $TV(P_0, P_1)$ is the total variation of the measure with a sign, and $\|Q\| = 2\sup_A |Q(A)|$. Thus, if the carrier sets of the measures P_0 and P_1 do not overlap, then error-free hypotheses distinction is possible. If the measures $P_0^{(N)}$ and $P_1^{(N)}$ are close, then $\|P_0^{(N)} - P_1^{(N)}\| \approx 0$ and, in this case, $\mathcal{E}r(N; \Gamma_0, \Gamma_1) \approx 1$.

The peculiarity of the problem is to detect a deterministic signal under small SNR values, i.e., $\mathcal{E}r(N; \Gamma_0, \Gamma_1) \approx 1$, by analyzing the spectral properties of the received signal-noise mixture and information criteria. Due to this peculiarity, the following problem was formulated in the paper [16].

Problem 1. Consider a given realization $\{x_1, \ldots, x_{2N+2}\}$ for a sequence of independent random variables $\{\xi_1, \ldots, \xi_{2N+2}\}$ with zero mean. Let the discrete Fourier transform (DFT)

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

be applied to this sequence to obtain the random variable

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

where $k = 0, \dots, N$. (Due to the symmetry of the DFT of a real signal, the second half of $N + 1, \dots,$ $2N + 1$ complex amplitudes of the spectral samples is conjugate to the first.)

It is required to find the discrete probability function of the normalized ordered spectral distribution $n_k(N)$ as the normalized mean for each kth value of the random variable

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

where $I_k = \Xi_k \Xi_k^*$ (the square of the amplitude modulus or the energy of the spectral sample), E_X is half of the signal energy, and $\mathbf{T}$ is the non-descending ordering operator of the series, and to investigate the properties of the resulting distribution on different information measures.

To solve Problem 1 in the general case, one needs to calculate

$$\mathbb{E}[\eta_k(N)] = \mathbb{E}\left[\frac{(\mathbf{T}I)_k}{E_X}\right]. \tag{5}$$

In 1898 A. Schuster established that the distributions of the random variables I_k, $k = 1, \dots, N$, are exponential, and the random variable I_0 obeys the χ^2 distribution with one degree of freedom. This fact holds if the random variables ξ_n, $n = 1, \dots, 2N + 2$, are independent Gaussian ones with zero mean and variance σ_0^2, as demonstrated in [17].

Remark 1. The total number of signal samples in the time domain is set equal to $2N + 2$ for the convenience of analyzing the energies I_k of half of the spectral samples, only N of which obey the exponential distribution.

In this paper, the detection problem is solved with a criterion defined by an information characteristic, i.e., complexity represented as the product of entropy and the $\mathcal{L}_1$ norm of two distributions. Being only a function of the discrete spectral distribution, entropy possesses insensitivity to the permutation of its samples; being a function of two distributions, complexity is sensitive to permutation when calculating a relatively nonuniform distribution. Therefore, when postulating the order of samples in a discrete distribution, a rule is established for calculating the information characteristics. As such a rule, we will use ascending or descending ordering for spectrum samples.

3. MAIN RESULTS. CALCULATING THE MEAN OF NORMALIZED ORDER STATISTICS

The mean of an order statistic normalized by the sum of samples possesses the following property.

Lemma 1. *Let $z_1, \dots, z_N$ be the observations of a random variable Z obeying the exponential distribution $F(z) = 1 - \exp(-z)$ with the density function $f(z) = \exp(-z)$. Consider the values of the sequence $z_{(1)}, \dots, z_{(N)}$ of the same results rearranged in ascending order, where the random variable $Z_{(k)}$ is a nondecreasing kth order statistic.*

Then the mean of such a random variable normalized by the sum of all elements of the sample has the form

$$\mathbb{E}\left[\frac{Z_{(k)}}{\sum_{i=1}^N Z_{(i)}}\right] = \frac{\int\limits_0^\infty \int\limits_0^{z_N} \dots \int\limits_0^{z_3} \int\limits_0^{z_2} \frac{z_k}{\sum_{i=1}^N z_i} \exp\left(-\sum\limits_{i=1}^N z_i\right) dz_1 dz_2 \dots dz_{N-1} dz_N}{\int\limits_0^\infty \int\limits_0^{z_N} \dots \int\limits_0^{z_3} \int\limits_0^{z_2} \exp\left(-\sum\limits_{i=1}^N z_i\right) dz_1 dz_2 \dots dz_{N-1} dz_N}. \tag{6}$$

Proof. The lemma is directly verified by writing the mean for the order statistic.

Formula (6) differs from the known ones in the calculation of the mean of $\frac{z_k}{\sum_{i=1}^N z_i}$ by the joint distribution of the order statistics $Z_{(1)}, \dots, Z_{(N)}$. A difficulty arises when calculating the integral in the numerator of (6), which is overcome using the relation between the means $\mathbb{E}\left[\frac{Z_{(k)}}{\sum_{i=1}^{N} Z_{(i)}}\right]$ and $\mathbb{E}[Z_{(k)}]$.

The density function of the distribution for the kth order statistic $Z_{(k)}$ is defined as follows:

$$f_{Z_{(k)}}(z) = \frac{N!}{(k-1)!(N-k)!} f(z)[F(z)]^{k-1}[1-F(z)]^{N-k}.$$

When substituting the exponential law $F(z)$ this density takes the explicit form

$$f_{Z_{(k)}}(z) = \frac{N!}{(k-1)!(N-k)!} \exp(-z(1+N-k))(1-\exp(-z))^{k-1}. \tag{7}$$

The next formula is a well-known result for finding the mean of order statistics of the standard exponential law [7]:

$$\mathbb{E}\left[Z_{(k)}\right] = \int\limits_0^\infty z f_{Z_{(k)}}(z)dz = \sum_{i=N-k+1}^{N} \frac{1}{i}. \tag{8}$$

On the other hand, this mean can be obtained by the formula of Lemma 1, i.e.,

$$\mathbb{E}\left[Z_{(k)}\right] = \frac{\int\limits_0^\infty \int\limits_0^{z_N} \ldots \int\limits_0^{z_3} \int\limits_0^{z_2} z_k \exp\left(-\sum\limits_{i=1}^{N} z_i\right) dz_1 dz_2 \ldots dz_{N-1} dz_N}{\int\limits_0^\infty \int\limits_0^{z_N} \ldots \int\limits_0^{z_3} \int\limits_0^{z_2} \exp\left(-\sum\limits_{i=1}^{N} z_i\right) dz_1 dz_2 \ldots dz_{N-1} dz_N} = \sum_{i=1}^{N} \frac{1}{i} - \sum_{i=1}^{N-k} \frac{1}{i} = \sum_{i=N-k+1}^{N} \frac{1}{i}. \tag{9}$$

Proposition 1. *The mean of the normalized order statistic* (6) *is given by*

$$\mathbb{E}\left[\frac{Z_{(k)}}{\sum_{i=1}^{N} Z_{(i)}}\right] = \frac{1}{N} \sum_{i=N-k+1}^{N} \frac{1}{i}, \tag{10}$$

which coincides, up to the factor $\frac{1}{N}$, *with the value of* $\mathbb{E}[Z_{(k)}]$.

Proof. The repeated integral in the denominator is analytically calculated for an arbitrary upper limit of the last integral:

$$\int\limits_0^a \int\limits_0^{z_N} \ldots \int\limits_0^{z_3} \int\limits_0^{z_2} \exp\left(-\sum_{i=1}^{N} z_i\right) dz_1 dz_2 \ldots dz_{N-1} dz_N = \frac{1}{N!} + O(a^{N-1}\exp(-a)).$$

Passing to the limit as $a \to \infty$ yields

$$\int\limits_0^\infty \int\limits_0^{z_N} \ldots \int\limits_0^{z_3} \int\limits_0^{z_2} \exp\left(-\sum_{i=1}^{N} z_i\right) dz_1 dz_2 \ldots dz_{N-1} dz_N = \frac{1}{N!}. \tag{11}$$

This result can be alternatively established by observing that

$$\int\limits_0^\infty \int\limits_0^\infty \ldots \int\limits_0^\infty \int\limits_0^\infty \exp\left(-\sum_{i=1}^{N} z_i\right) dz_1 dz_2 \ldots dz_{N-1} dz_N = 1.$$

At the same time, this integral consists of the sum of $N!$ integrals of the form (11) (by the number of all permutations of z_i, $i = 1, \ldots, N$).

Now we study the integral in the numerator of (10). The repeated integral in the numerator is analytically calculated for an arbitrary upper limit of the last integral:

$$\int_0^a \int_0^{z_N} \dots \int_0^{z_3} \int_0^{z_2} \frac{z_k}{\sum_{i=1}^N z_i} \exp\left(-\sum_{i=1}^N z_i\right) dz_1 dz_2 \dots dz_{N-1} dz_N$$
$$= \frac{1}{N!}\frac{1}{N}\left(\sum_{i=1}^N \frac{1}{i} - \sum_{i=1}^{N-k} \frac{1}{i}\right) + O(a^N \exp(-a)) + O(a^{N+1}\mathrm{Ei}(1,a)),$$

where $\mathrm{Ei}(1,a) = \int\limits_1^\infty t^{-1}\exp(-ta)dt$.

Passing to the limit as $a \to \infty$ gives the desired result.

Section 4 proposes a method for calculating the variance of a normalized order random variable, which can also be used to prove Proposition 1.

Simultaneously, several interesting results have been obtained concerning repeated improper integrals:

$$\int_0^\infty \int_0^{z_N} \dots \int_0^{z_3} \int_0^{z_2} \exp\left(-\sum_{i=1}^N z_i\right) dz_1 dz_2 \dots dz_N = \frac{1}{N!}, \tag{12}$$

$$\int_0^\infty \int_0^{z_N} \dots \int_0^{z_3} \int_0^{z_2} \frac{1}{\sum\limits_{i=1}^N z_i} \exp\left(-\sum_{i=1}^N z_i\right) dz_1 dz_2 \dots dz_N = \frac{1}{N-1}\frac{1}{N!}, \tag{13}$$

$$\int_0^\infty \int_0^{z_N} \dots \int_0^{z_3} \int_0^{z_2} \frac{1}{\left(\sum\limits_{i=1}^N z_i\right)^2} \exp\left(-\sum_{i=1}^N z_i\right) dz_1 dz_2 \dots dz_N = \frac{1}{N-2}\frac{1}{N-1}\frac{1}{N!}. \tag{14}$$

They can be calculated using the following property.

Proposition 2. *For an arbitrary natural degree* p, $N - p \geqslant 1$, *wee have*

$$M(p,N) = \int_0^\infty \int_0^\infty \dots \int_0^\infty \int_0^\infty \frac{1}{\left(\sum\limits_{i=1}^N z_i\right)^p} \exp\left(-\sum_{i=1}^N z_i\right) dz_1 dz_2 \dots dz_N = \frac{(N-p-1)!}{(N-1)!}. \tag{15}$$

For $p \geqslant N$, *this integral becomes meaningless due to its divergence. Moreover, for an arbitrary natural degree* p,

$$\int_0^\infty \int_0^\infty \dots \int_0^\infty \int_0^\infty \left(\sum_{i=1}^N z_i\right)^p \exp\left(-\sum_{i=1}^N z_i\right) dz_1 dz_2 \dots dz_N = \frac{(N+p-1)!}{(N-1)!}. \tag{16}$$

Proof. The change of variables $\{z_i\}_{i=1}^N \to \{x_i\}_{i=1}^N$ in the integrals (15) and (16) is performed by the following rule [18]:

$$x_1 = \frac{z_1}{x_N}, \dots, x_{N-1} = \frac{z_{N-1}}{x_N},\ x_N = \sum_{i=1}^N z_i. \tag{17}$$

To find the Jacobian $J(x_1, \ldots, x_N)$ of the mapping (17), it is necessary to express the variables $\{z_i\}_{i=1}^N$ through $\{x_i\}_{i=1}^N$:

$$z_1 = x_1 x_N, \ldots, z_{N-1} = x_{N-1} x_N,$$
$$z_N = \sum_{i=1}^{N} z_i - \sum_{i=1}^{N-1} z_i = x_N - \sum_{i=1}^{N-1} x_N x_i = x_N \left(1 - \sum_{i=1}^{N-1} x_i\right).$$

After elementary transformations, the Jacobian of this mapping is reduced to the determinant of the upper-triangular matrix and equals $J(x_1, \ldots, x_N) = x_N^{N-1}$. Therefore, the integral (15) is calculated as follows:

$$M(p,N) = \int_0^\infty \int_0^\infty \ldots \int_0^\infty \int_0^\infty \left(\sum_{i=1}^{N} z_i\right)^{-p} \exp\left(-\sum_{i=1}^{N} z_i\right) dz_1 dz_2 \ldots dz_N$$
$$= \int_0^\infty x_N^{-p} x_N^{N-1} \exp(-x_N)\, dx_N \left(\int_0^1 \int_0^{x_{N-1}} \ldots \int_0^{x2} dx_1 \ldots dx_{N-2} dx_{N-1}\right)$$
$$= \int_0^\infty x_N^{-p} \frac{x_N^{N-1}}{\Gamma(N)} \exp(-x_N)\, dx_N \left(\Gamma(N) \int_0^1 \int_0^{x_{N-1}} \ldots \int_0^{x2} dx_1 \ldots dx_{N-2} dx_{N-1}\right)$$
$$= \int_0^\infty x_N^{-p} \frac{x_N^{N-1}}{\Gamma(N)} \exp(-x_N)\, dx_N \int_0^1 \ldots \int_0^1 dx_1 \ldots dx_{N-1} = \frac{\Gamma(N-p)}{\Gamma(N)} = \frac{(N-p-1)!}{(N-1)!}.$$

The integral (16) is found by analogy. Finally, note that the integral (15) consists of $N!$ identical integrals of the form (13) for $p = 1$.

Here is another interesting mathematical result on the calculation of the integral (13).

Lemma 2. *For any $N \in \mathbb{N}$, the integral (15) with $p = 1$ can be somehow represented by dividing into two parts containing k and $N - k$ elements:*

$$M(1,N) = \frac{1}{\Gamma(N-k)\Gamma(k)} \int_0^\infty \int_0^\infty \frac{\zeta^{N-k-1}\eta^{k-1}}{\zeta + \eta} \exp(-(\zeta + \eta))\, d\zeta d\eta. \tag{18}$$

Proof. This lemma can be verified using two changes of variables in the integral, similar to the proof of Proposition 2. Note only that in this case, as before by Proposition 2, $M(1,N) = \frac{1}{N-1}$.

4. MAIN RESULTS. CALCULATING THE VARIANCE OF NORMALIZED ORDER STATISTICS

Now we have the problem of calculating the variance $\mathbb{D}\left[\frac{Z_{(k)}}{\sum_{i=1}^N Z_{(i)}}\right]$; to this end, it is necessary to find $\mathbb{E}\left[\frac{Z_{(k)}^2}{\left(\sum_{i=1}^N Z_{(i)}\right)^2}\right]$.

Proposition 3. *The desired mean can be obtained as follows:*

$$\mathbb{E}\left[\frac{Z_{(k)}^2}{\left(\sum_{i=1}^N Z_{(i)}\right)^2}\right] = \frac{1}{N(N+1)} \mathbb{E}\left[Z_{(k)}^2\right]. \tag{19}$$

Proof. The means in both parts of the equality are written by definition considering the order of integration:

$$\mathbb{E}\left[Z_{(k)}^2\right] = N!\int\limits_0^\infty\int\limits_0^{z_N}\dots\int\limits_0^{z_3}\int\limits_0^{z_2} z_k^2\exp\left(-\sum_{i=1}^N z_i\right)dz_1dz_2\dots dz_{N-1}dz_N,$$

$$\mathbb{E}\left[\frac{Z_{(k)}^2}{\left(\sum\limits_{i=1}^N Z_{(i)}\right)^2}\right] = N!\int\limits_0^\infty\int\limits_0^{z_N}\dots\int\limits_0^{z_3}\int\limits_0^{z_2}\frac{z_k^2}{\left(\sum\limits_{i=1}^N z_i\right)^2}\exp\left(-\sum_{i=1}^N z_i\right)dz_1dz_2\dots dz_{N-1}dz_N.$$

Similar to the proof of Proposition 2, we apply the change of variables (17) in the integrals to continue the corresponding equalities:

$$\mathbb{E}\left[Z_{(k)}^2\right] = N!\int\limits_0^\infty\int\cdots\int\limits_\Omega x_N^{N-1}x_k^2x_N^2\exp\left(-x_N\right)dx_1dx_2\dots dx_{N-1}dx_N$$

$$= N!\int\limits_0^\infty x_N^{N+1}\exp\left(-x_N\right)dx_N\int\cdots\int\limits_\Omega x_k^2dx_1dx_2\dots dx_{N-1}$$

$$= N(N+1)N!\int\limits_0^\infty x_N^{N-1}\exp\left(-x_N\right)dx_N\int\cdots\int\limits_\Omega x_k^2dx_1dx_2\dots dx_{N-1},$$

$$\mathbb{E}\left[\frac{Z_{(k)}^2}{\left(\sum\limits_{i=1}^N Z_{(i)}\right)^2}\right] = N!\int\limits_0^\infty\int\cdots\int\limits_\Omega x_N^{N-1}x_k^2\exp\left(-x_N\right)dx_1dx_2\dots dx_{N-1}dx_N$$

$$= N!\int\limits_0^\infty x_N^{N-1}\exp\left(-x_N\right)dx_N\int\cdots\int\limits_\Omega x_k^2dx_1dx_2\dots dx_{N-1},$$

where Ω denotes the domain of integration for the new variables $x_1,\dots,x_{N-1}$ under the change. In the case under consideration, this domain is part of the simplex that arises under a similar change when deriving the Dirichlet distribution from the generating gamma distributions [19, 20].

Comparing the expressions for $\mathbb{E}\left[\frac{Z_{(k)}^2}{\left(\sum_{i=1}^N Z_{(i)}\right)^2}\right]$ and $\mathbb{E}\left[Z_{(k)}^2\right]$, we arrive at the desired result.

The method above is also suitable for proving Proposition 1.

Corollary 1. *Due to Proposition 3 and formula* (10)*, the variance is given by*

$$\mathbb{D}\left[\frac{Z_{(k)}}{\sum_{i=1}^N Z_{(i)}}\right] = \frac{1}{N(N+1)}\mathbb{E}\left[Z_{(k)}^2\right] - \left(\frac{1}{N}\mathbb{E}\left[Z_{(k)}\right]\right)^2. \tag{20}$$

Now we can estimate its value.

Corollary 2. *For all N the variance* (20) *satisfies the upper bound*

$$\mathbb{D}\left[\frac{Z_{(k)}}{\sum_{i=1}^N Z_{(i)}}\right] \leqslant \frac{1}{N^2}\mathbb{D}\left[Z_{(k)}\right].$$

Proof. It follows from Corollary 1 that

$$\mathbb{D}\left[\frac{Z_{(k)}}{\sum_{i=1}^{N} Z_{(i)}}\right] = \frac{1}{N(N+1)}\mathbb{E}\left[Z_{(k)}^2\right] - \left(\frac{1}{N}\mathbb{E}\left[Z_{(k)}\right]\right)^2 = \frac{1}{N^2}\mathbb{D}\left[Z_{(k)}\right] - \frac{1}{N^2(N+1)}\mathbb{E}\left[Z_{(k)}^2\right].$$

The proof of this corollary is complete.

Proposition 4. *The second moment and variance of a non-decreasing order statistic are given by*

$$\mathbb{E}\left[Z_{(k)}^2\right] = \left(\sum_{i=N-k+1}^{N} \frac{1}{i}\right)^2 + \sum_{i=N-k+1}^{N} \frac{1}{i^2}, \tag{21}$$

$$\mathbb{D}\left[Z_{(k)}\right] = \sum_{i=N-k+1}^{N} \frac{1}{i^2}. \tag{22}$$

Proof. For $\mathbb{E}\left[Z_{(k)}^2\right]$, we have the formula [7]

$$\mathbb{E}\left[Z_{(k)}^2\right] = \int_0^\infty z^2 f_{Z_{(k)}}(z)dz = \int_0^\infty \frac{z^2 N!}{(k-1)!(N-k)!}\exp(-z(1+N-k))(1-\exp(-z))^{(k-1)}dz$$
$$= (H_N - H_{N-k})^2 + \psi^{(1)}(N-k+1) - \psi^{(1)}(N+1),$$

which contains the polygamma function $\psi^{(m)}(n) = (-1)^{(m+1)} m! \sum_{k=n}^{\infty} \frac{1}{k^{(m+1)}}$ and the harmonic series $H_N = \sum_{i=1}^{N} \frac{1}{i}$.

Substituting the expression $\psi^{(m)}(n)$ into the last equality yields (21). The variance (22) is obtained from (21) by considering formula (8) for $\mathbb{E}\left[Z_{(k)}\right]$.

Proposition 5. *The variance* (20) *is calculated using the constructive formula*

$$\mathbb{D}\left[\frac{Z_{(k)}}{\sum_{i=1}^{N} Z_{(i)}}\right] = \frac{1}{N(N+1)}\left(\sum_{i=N-k+1}^{N} \frac{1}{i^2} - \frac{1}{N}\left(\sum_{i=N-k+1}^{N} \frac{1}{i}\right)^2\right). \tag{23}$$

Proof. Let us transform the general expression (20) for the variance of a normalized order statistic using (21) and (10):

$$\mathbb{D}\left[\frac{Z_{(k)}}{\sum_{i=1}^{N} Z_{(i)}}\right] = \frac{1}{N(N+1)}\mathbb{E}\left[Z_{(k)}^2\right] - \left(\frac{1}{N}\mathbb{E}\left[Z_{(k)}\right]\right)^2$$
$$= \frac{1}{N(N+1)}\left(\left(\sum_{i=N-k+1}^{N} \frac{1}{i}\right)^2 + \sum_{i=N-k+1}^{N} \frac{1}{i^2}\right) - \left(\frac{1}{N}\sum_{i=N-k+1}^{N} \frac{1}{i}\right)^2$$
$$= \left(\sum_{i=N-k+1}^{N} \frac{1}{i}\right)^2 \left(\frac{1}{N(N+1)} - \frac{1}{N^2}\right) + \frac{1}{N(N+1)}\sum_{i=N-k+1}^{N} \frac{1}{i^2} \tag{24}$$
$$= \left(\sum_{i=N-k+1}^{N} \frac{1}{i}\right)^2 \left(\frac{-1}{N^2(N+1)}\right) + \frac{1}{N(N+1)}\sum_{i=N-k+1}^{N} \frac{1}{i^2}$$
$$= \frac{1}{N(N+1)}\left(\sum_{i=N-k+1}^{N} \frac{1}{i^2} - \frac{1}{N}\left(\sum_{i=N-k+1}^{N} \frac{1}{i}\right)^2\right).$$

Here are a few more interesting facts from the results above.

Corollary 3. *For $k = N$, formula* (22) *yields*

$$\lim_{N\to\infty} \mathbb{D}\left[Z_{(N)}\right] = \lim_{N\to\infty} \sum_{i=1}^{N} \frac{1}{i^2} = \frac{\pi^2}{6}.$$

Corollary 4. *For all N, we have*

$$\mathbb{E}\left[Z_{(1)}^2\right] = \frac{1}{N+1}\frac{2}{N^2}.$$

5. THE MEAN VALUE OF THE NORMALIZED ENERGY SPECTRUM DENSITY, ITS ESTIMATES AND SPECTRAL COMPLEXITY

Let us return to the original Problem 1. After finding the energy spectral samples of one observation window, $y_1, \ldots, y_N$, the results of observations of the exponentially distributed random variable Y generating $\eta_k(N)$, are introduced [16]. Then the values of the sequence $y_{(1)}, \ldots, y_{(N)}$ (the same results rearranged in descending order) are an inverse variational series, where the random variable $Y_{(k)}$ is a nonincreasing kth order statistic.

In this case, we have

$$\mathbb{E}\left[\frac{Z_{(k)}}{\sum_{i=1}^{N} Z_{(i)}}\right] = \mathbb{E}\left[\frac{Y_{(N-k+1)}}{\sum_{i=1}^{N} Y_{(i)}}\right],$$

or

$$\widetilde{n}_k(N) = \mathbb{E}\left[\frac{Y_{(k)}}{\sum_{i=1}^{N} Y_{(i)}}\right] = \mathbb{E}\left[\frac{Z_{(N-k+1)}}{\sum_{i=1}^{N} Z_{(i)}}\right] = \frac{1}{N}\left(\sum_{i=1}^{N}\frac{1}{i} - \sum_{i=1}^{k-1}\frac{1}{i}\right) = \frac{1}{N}\sum_{i=k}^{N}\frac{1}{i}. \tag{25}$$

Based on (25), it follows that

$$\sum_{k=1}^{N} \mathbb{E}\left[\frac{Y_{(k)}}{\sum_{i=1}^{N} Y_{(i)}}\right] = N\frac{1}{N} = 1,$$

and the normalization condition is obviously valid.

In turn, the variance is given by

$$\mathbb{D}\left[\frac{Z_{(k)}}{\sum_{i=1}^{N} Z_{(i)}}\right] = \mathbb{D}\left[\frac{Y_{(N-k+1)}}{\sum_{i=1}^{N} Y_{(i)}}\right],$$

or

$$\mathbb{D}\left[\frac{Y_{(k)}}{\sum_{i=1}^{N} Y_{(i)}}\right] = \frac{1}{N(N+1)}\left(\sum_{i=k}^{N}\frac{1}{i^2} - \frac{1}{N}\left(\sum_{i=k}^{N}\frac{1}{i}\right)^2\right). \tag{26}$$

The probability function of the normalized ordered discrete spectrum is approximately calculated using the formula [16]

$$n_k(N) = -\frac{1}{NK_N}\ln\frac{k}{N+1}, \quad \text{where} \quad K_N = -\frac{1}{N}\sum_{k=1}^{N}\ln\frac{k}{N+1}. \tag{27}$$

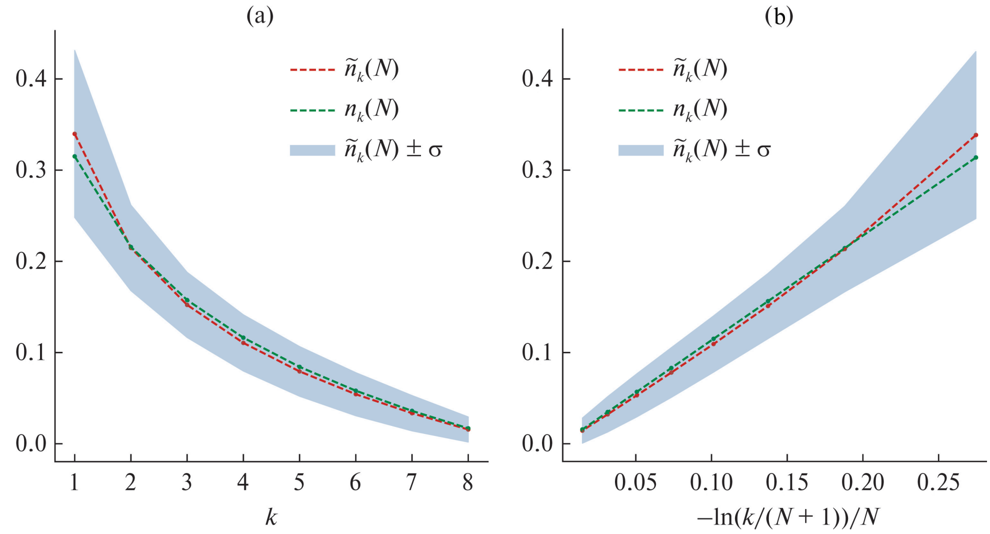

Fig. 1. Discrete distributions $\widetilde{n}_k(N)$ and $n_k(N)$ for the series of size $N = 8$: the horizontal axis in (a) linear and (b) logarithmic scale.

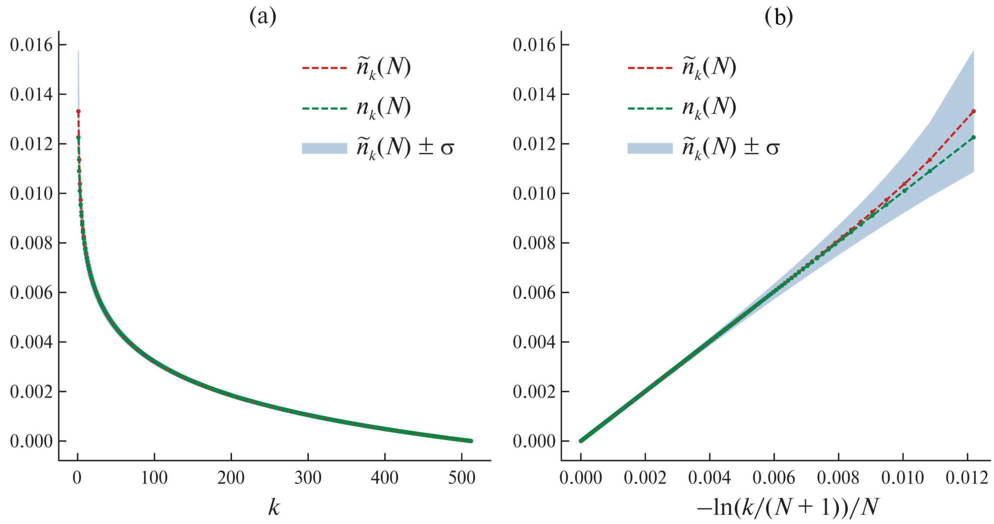

Fig. 2. Discrete distributions $\widetilde{n}_k(N)$ and $n_k(N)$ for the series of size $N = 512$: the horizontal axis in (a) linear and (b) logarithmic scale.

Now we compare the discrete distributions: the exact (25) and approximate (27) ones.

For large values of N and k, the expression (25), which is the difference of harmonic series, can be approximated in various ways. Let us choose the estimate

$$\sum_{i=1}^{N} \frac{1}{i} - \sum_{i=1}^{k-1} \frac{1}{i} \approx -\ln \frac{k}{N+1},$$

which is well-defined and makes sense for all N and $k \in [1, \ldots, N]$ and also appears in (27).

Figures 1 and 2 present the plots of the distributions (25) and (27) depending on the sample number. The symbol σ indicates the standard deviation of the distribution of the order statistic given by (26).

According to Figs. 1 and 2, the exact values of the mean of (25) slightly differ from those of the distribution (27) for a small number of points (members of the spectral series), almost coinciding for the other samples. As N grows, the share of points deviating from the estimate rapidly decreases.

Lemma 3. *In the general case, a more accurate estimate is*

$$\sum_{i=1}^{N}\frac{1}{i}-\sum_{i=1}^{k-1}\frac{1}{i}\approx -\ln\frac{k-0.5}{N+0.5}, \tag{28}$$

which is also well-defined and makes sense for all N and $k\in[2,\ldots,N]$.

Proof. To show this fact, we address the theory of harmonic series.

The partial sum of the first N terms of the harmonic series is called the Nth harmonic number:

$$H_N=\sum_{i=1}^{N}\frac{1}{i}=1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{N}. \tag{29}$$

In 1740, Euler derived an asymptotic expression for H_N, called the Euler–Maclaren formula:

$$H_N=\ln N+\gamma+\frac{1}{2N}-\varepsilon_N, \tag{30}$$

where $\gamma=0.5772\ldots$ is the Euler–Mascheroni constant and $0\leqslant\varepsilon_N\leqslant 1/8N^2$. Hence, $\varepsilon_N\to 0$ as $N\to\infty$, and for large N we have

$$H_N=\ln N+\gamma+O(N^{-1})\approx\ln N+\gamma. \tag{31}$$

This expression is called the Euler formula for the sum of the first n members of the harmonic series. Returning to (30), note that

$$\ln\left(N+\frac{1}{2}\right)=\ln N+\frac{1}{2N}+O\left(\frac{1}{N^2}\right). \tag{32}$$

Substituting the right-hand side into (30) yields

$$H_N=\ln\left(N+\frac{1}{2}\right)+\gamma+O\left(\frac{1}{N^2}\right), \tag{33}$$

since the numbers ε_N and ε_{k+1} have a close order of smallness with respect to $\frac{1}{N^2}$ and $\frac{1}{k^2}$, which rapidly vanish with increasing N and k.

Therefore, (28) takes the form

$$\sum_{i=1}^{N}\frac{1}{i}-\sum_{i=1}^{k-1}\frac{1}{i}=H_N-H_{k-1}=\ln\left(N+\frac{1}{2}\right)+O\left(\frac{1}{N^2}\right)-\ln\left(k-\frac{1}{2}\right)+O\left(\frac{1}{(k-1)^2}\right). \tag{34}$$

Thus, for relatively large N and k, we arrive at the estimate

$$\sum_{i=1}^{N}\frac{1}{i}-\sum_{i=1}^{k-1}\frac{1}{i}\approx -\ln\frac{k-0.5}{N+0.5}, \tag{35}$$

and the proof of this lemma is complete.

To proceed, we approximately calculate the probability function of the normalized ordered discrete spectrum by the formula

$$n_k(N)=-\frac{1}{NK_N}\ln\frac{2k-1}{2N+1},\quad\text{where}\quad K_N=-\frac{1}{N}\sum_{k=1}^{N}\ln\frac{2k-1}{2N+1}. \tag{36}$$

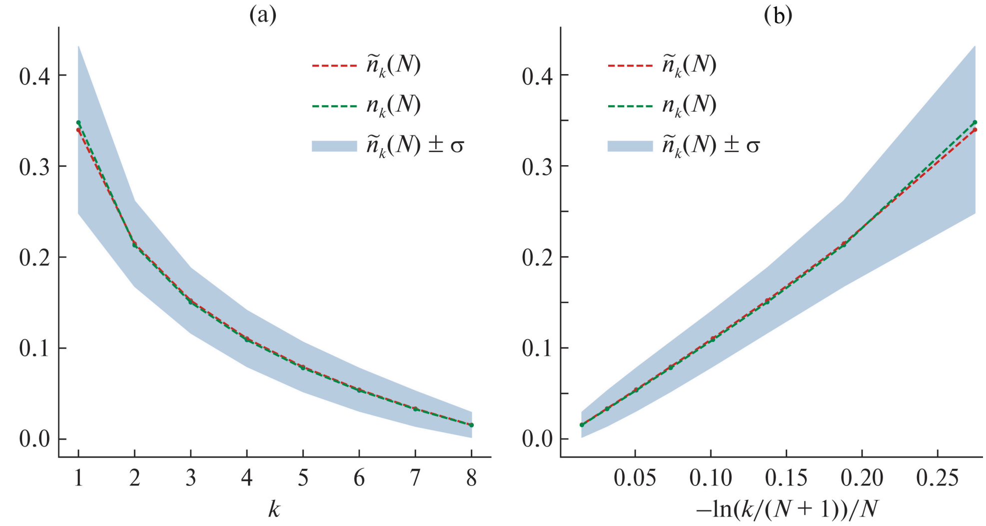

Fig. 3. Discrete distributions $\widetilde{n}_k(N)$ and $n_k(N)$ for the series of size $N = 8$: the horizontal axis in (a) linear and (b) logarithmic scale.

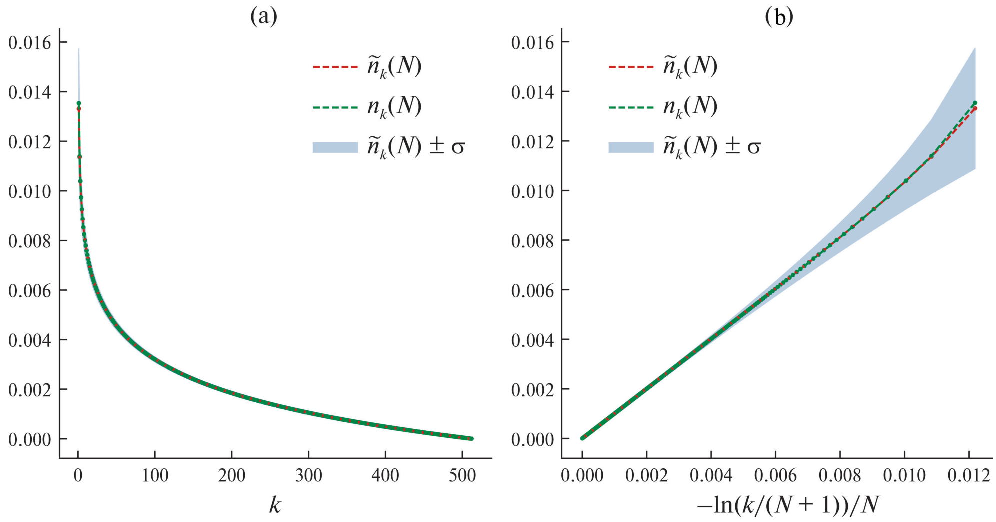

Fig. 4. Discrete distributions $\widetilde{n}_k(N)$ and $n_k(N)$ for the series of size $N = 512$: the horizontal axis in (a) linear and (b) logarithmic scale.

The result of this lemma is illustrated in Figs. 3 and 4.

Obviously, the two distributions almost coincide even for small N. Now, assuming that the normalized order statistics $(\eta_1, \ldots, \eta_N)$ take values $p = (p_1, \ldots, p_N)$, the spectral complexity can also be set based on the distribution (25) by the formula

$$C_{SS}(p) = -\frac{1}{4\log_2 N}\left(\sum_{k=1}^{N} p_k \log_2 p_k\right)\left(\sum_{k=1}^{N} |p_k - \widetilde{n}_k(N)|\right)^2 \qquad (37)$$

or its approximate analog (36) by the formula

$$C_S(p) = -\frac{1}{4\log_2 N}\left(\sum_{k=1}^{N} p_k \log_2 p_k\right)\left(\sum_{k=1}^{N} |p_k - n_k(N)|\right)^2; \qquad (38)$$

this will be demonstrated in Section 6.

6. STATISTICAL SIMULATION OF DETERMINISTIC SIGNAL DETECTION

To illustrate the analytical results of this paper, we use the statistical simulation procedure based on the analysis of the generated numerical data. It was described in detail in the authors' previous paper [16]. As before, all numerical results were obtained using the Python language and the Numpy and Scipy libraries.

Consider pairs of data sequences corresponding to two hypotheses of signal receipt:

$$\begin{cases} \Gamma_0 : x_n = w_n \\ \Gamma_1 : x_n = s_n + w_n, n = 1, \ldots, N. \end{cases} \tag{39}$$

The hypothesis Γ_0 is associated with the decision to receive only noise while the hypothesis Γ_1 with the decision to receive a mixture of a useful signal and noise, where the sequences $\{x(n)\}$, $\{s(n)\}$, and $\{w(n)\}$, $n = 1, \ldots, N$, denote the time series of the received data, useful signal, and additive white Gaussian random noise, respectively, and N is the time series (frame) length.

To test the quality of distinguishing between the useful deterministic signal and noise, statistics were collected on $Q = 50\,000$ numerically generated frames $\{x_n\}$ of the signal-noise mixture of length $2N = 16\,384$ with spectra size $N = 8192$, respectively. In all realizations, the signal $\{s_n\}$ remained the same, i.e., a fixed (in number and amplitude) set of $K = 30$ sinusoids evenly spaced along the spectrum with random phases. The additive white Gaussian noise $\{w_n\}$ was obtained by a Gaussian sequence generator with the mean $\mu = 0$ and variance Σ (within one set of Q frames).

Figure 5 shows the ordered discrete spectral distributions p_k for particular signal realizations and the theoretical means of $\widetilde{n}_k(N)$ described by (25). Obviously, in each realization, the real spectral distribution of white noise may appreciably differ from the theoretical one $\widetilde{n}_k(N)$. However, all such realizations lie within the confidence interval defined by the standard deviation σ of the distribution of the order statistic (26). The blue color in Fig. 5 indicates the domain bounding the confidence interval for each sample number k with the upper and lower bounds $\widetilde{n}_k(N) \pm 3\sigma$.

For each resulting sequence $\{x_n\}$, the order statistic of the discrete normalized spectral distribution p_k was calculated; for details, see [9, 15]. Next, based on p_k, the values of C_{SS} (37) and C_S (38) were obtained for the noise and noise-signal mixture satisfying the two hypotheses from the

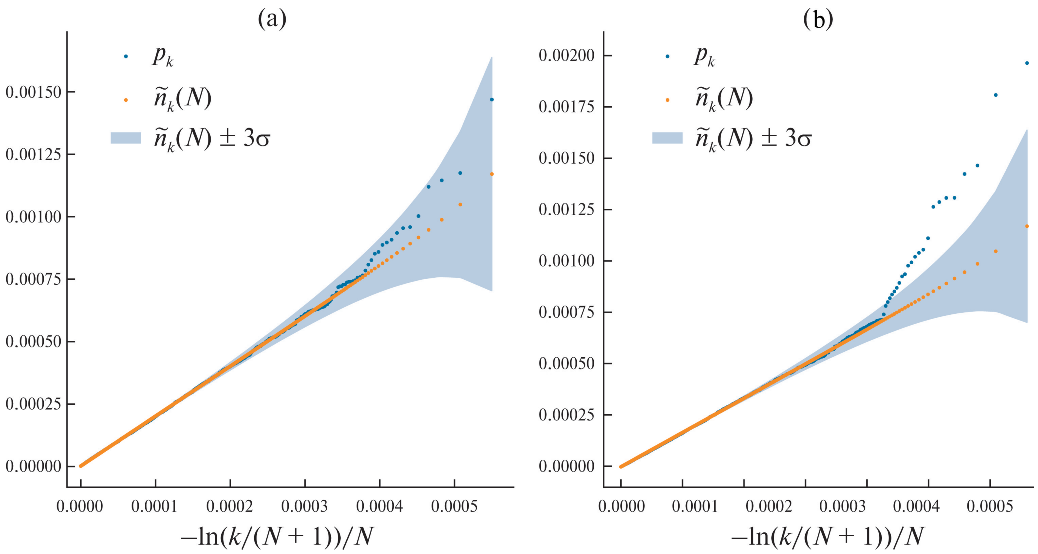

Fig. 5. Discrete distributions $\widetilde{n}_k(N)$ and p_k for the series of size $N = 8192$ and $SNR = -15.22$ dB: the time sequence of (a) noise data w_n and (b) signal and noise data $s_n + w_n$. The horizontal axis in logarithmic scale.

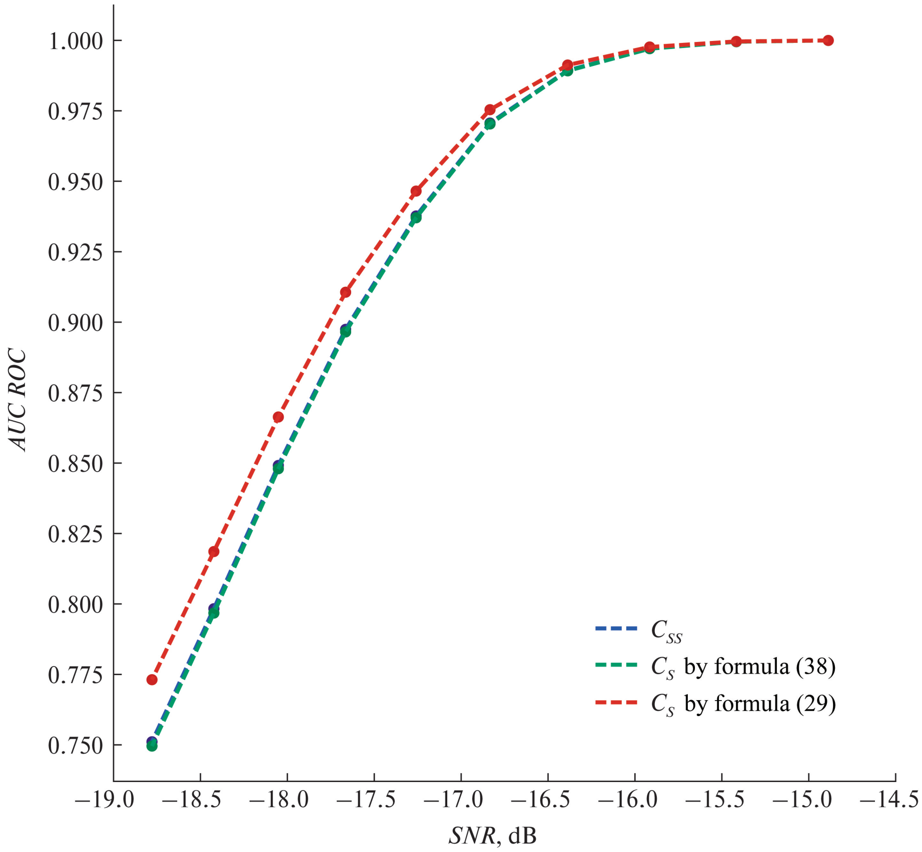

Fig. 6. The quality of binary classification in the *AUC ROC* metric depending on SNR for the spectral complexity functions under consideration.

expressions (39). The value of C_S was found for the two different approximations (27) and (36) of the order statistics of the spectral distribution of the white noise: it seems interesting to compare them as well.

The final result of simulating and comparing the calculated information metrics is the quality of binary classification in the *AUC ROC* (Area Under the Receiver Operating Characteristic Curve) metric depending on SNR, defined by

$$SNR = 10\log_{10}\left(\frac{E_{signal}}{E_{noise}}\right), \tag{40}$$

where E_{signal} and E_{noise} denote the total signal and noise energies, respectively, calculated as the sum of the powers of the spectral decomposition of the sequences $\{s_n\}$ and $\{w_n\}$.

To obtain this dependence, the statistic Q of frames described above was collected for several values of the noise variance Σ, the histograms of the distributions of C_{SS} and C_S were built, and the values of *AUC ROC* were finally calculated [16].

Figure 6 shows the comparison of the quality of binary classification of the signal and noise for the information metrics C_{SS} and C_S. Obviously, the indicators for the functions C_{SS} and C_S, calculated using (36) with theoretical justification by Lemma 3, coincide almost completely. In addition, the information characteristic C_S calculated using formula (36) from the authors' previous paper [16] has a slight improvement in the quality of detection. All the analytical information criteria introduced in this paper demonstrate a high quality of detection of the deterministic signal in noise under small SNR values.

7. CONCLUSIONS

This paper has established that the order statistic of a discrete normalized spectral distribution is a powerful tool for detecting a deterministic signal under small SNR values in single-window observations. The spectral complexity calculated on a particular realization of the signal-noise mixture has been used as a detection criterion.

An attempt to separate the two problems (the detection/pre-detection of deterministic signals in white noise under small SNR values and classification) has demonstrated that increasing the sensitivity of the detection method causes the loss of all physical properties of the signal: further work involves only the informational ones. A simple analogy with quantum systems is suggested here, in which there exists an uncertainty relation when the researcher cannot accurately measure the speed and coordinate simultaneously. In fact, the reduced informativeness of particular signal frequencies makes it impossible to classify the signal, let alone recover the signal from a measurement together with noise. However, the ordering of the noise spectrum has a particular rigid structure allowing one to judge the presence of a signal even beyond the sensitivity of classical energy receivers.

Moreover, it seems promising to apply the proposed detection method for other types of noise as well as to construct classification grids for signals based on the information characteristics of the spectra of signal-noise mixtures. These will be the subjects of the next research by the authors.

FUNDING

This work was supported in part by the Russian Science Foundation, project no. 23-19-00134.

REFERENCES

1. Amigó, J.M. and Rosso, O.A., Ordinal Methods: Concepts, Applications, New Developments, and Challenges – in Memory of Karsten Keller (1961–2022), *Chaos*, 2023, vol. 33, no. 8, art. no. 080401. https://pubs.aip.org/cha/article/33/8/080401/2905538/Ordinal-methods-Concepts-applications-new

2. Rosso, O.A., Larrondo, H.A., Martin, M.T., Plastino, A., and Fuentes, M.A., Distinguishing Noise from Chaos, *Phys. Rev. Lett.*, 2007, vol. 99, no. 15, art. no. 154102. https://doi.org/10.1103/PhysRevLett.99.154102

3. Terebizh, V.Yu., *Analiz vremennykh ryadov v astrofizike* (Analysis of Time Series in Astrophysics), Moscow: Nauka, 1992.

4. Dobrushin, R.L., Pinsker, M.S., and Shiryaev, A.N., Application of the Notion of Entropy in the Problems of Detecting a Signal in Noise, *Lithuanian Math. J.*, 1963, vol. 3, no. 1, pp. 107–122.

5. Chen, Z., Wu, C., Wang, J., and Qiu, H., Tsallis Entropy-Based Complexity-IPE Casualty Plane: A Novel Method for Complex Time Series Analysis, *Entropy*, 2024, vol. 26, no. 6, art. no. 521. https://www.mdpi.com/1099-4300/26/6/521

6. Hoeffding, W., On the Distribution of the Expected Values of the Order Statistics, *Ann. Math. Stat.*, 1953, vol. 24, no. 1, pp. 93–100. https://doi.org/10.1214/aoms/1177729086

7. Balakrishnan, N. and Cohen, A.C., *Order Statistics & Inference: Estimation Methods*, Academic Press, 1991. https://doi.org/10.1016/C2009-0-22411-1

8. Vecchio, F., Miraglia, F., Pappalettera, C., Orticoni, A., Alù, F., Judica, E., Cotelli, M., and Rossini, P.M., Entropy as Measure of Brain Networks' Complexity in Eyes Open and Closed Conditions, *Symmetry*, 2021, vol. 13, no. 11, art. no. 2178. https://www.mdpi.com/2073-8994/13/11/2178

9. Berlin, L.M., Galyaev, A.A., and Lysenko, P.V., Comparison of Information Criteria for Detection of Useful Signals in Noisy Environments, *Sensors*, 2023, vol. 23, no. 4, art. no. 2133. https://www.mdpi.com/1424-8220/23/4/2133

10. Zhang, W., Huang, D., Zhou, M., Lin, J., and Wang, X., Open-Set Signal Recognition Based on Transformer and Wasserstein Distance, *Appl. Sci.*, 2023, vol. 13, no. 4, art. no. 2151. https://www.mdpi.com/2076-3417/13/4/2151

11. Cazelles, E., Robert, A., and Tobar, F., The Wasserstein–Fourier Distance for Stationary Time Series, *IEEE Transact. Signal Proc.*, 2021, vol. 69, pp. 709–721. https://ieeexplore.ieee.org/document/9303405

12. Ferracuti, F., Freddi, A., Monteriu, A., and Romeo, L., Fault Diagnosis of Rotating Machinery Based on Wasserstein Distance and Feature Selection, *IEEE Transact. Autom. Sci. Engin.*, 2022, vol. 19, no. 3, pp. 1997–2007. https://doi.org/10.1109/TASE.2021.3069109

13. Quan, Z., Zhang, W., Shellhammer, S.J., and Sayed, A.H., Optimal Spectral Feature Detection for Spectrum Sensing at Very Low SNR, *IEEE Transact. Commun.*, 2011, vol. 59, no. 1, pp. 201–212. https://doi.org/10.1109/TCOMM.2010.112310.090306

14. Bardenet, R., Flamant, J., and Chainais, P., On the Zeros of the Spectrogram of White Noise, *Appl. Comput. Harmon. Anal.*, 2020, vol. 48, no. 2, pp. 682–705. https://doi.org/10.1016/j.acha.2018.09.002

15. Galyaev, A.A., Lysenko, P.V., and Berlin, L.M., Statistical Complexity as a Criterion for the Useful Signal Detection Problem, *Autom. Remote Control*, 2023, vol. 84, no. 7, pp. 852–871. https://doi.org/10.25728/arcRAS.2023.27.32.001

16. Galyaev, A.A., Babikov, V.G., Lysenko, P.V., and Berlin, L.M., A New Spectral Measure of Complexity and Its Capabilities for Detecting Signals in Noise, *Dokl. Math.*, 2024, vol. 110. https://doi.org/10.1134/S1064562424702235

17. Richards, M.A., The Discrete-Time Fourier Transform and Discrete Fourier Transform of Windowed Stationary White Noise, *Technical Memorandum*, 2013, pp. 1–24.

18. DasGupta, A., *Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics*, New York: Springer, 2011. https://doi.org/10.1007/978-1-4419-9634-3

19. Frigyik, B., Kapila, A., and Gupta, M., Introduction to the Dirichlet Distribution and Related Processes, UWEE Technical Report no. UWEETR-2010-0006, 2010, pp. 1–28.

20. Hafeez, M., Kamal, S., and Shahbaz, M., The Multivariate Order Statistics for Exponential and Weibull Distributions, *Pak. J. Stat. Oper. Res.*, 2014, vol. 10, pp. 361–368. https://doi.org/10.18187/pjsor.v10i3.825

This paper was recommended for publication by B.M. Miller, a member of the Editorial Board