ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 3, pp. 270–286.

Optimization Statements of Signal Detection Problems

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]*babikov@ipu.ru*

Received July 7, 2025
Revised September 1, 2025
Accepted September 24, 2025

Abstract—The article develops the authors' previous work and examines the sequence of problems that form an approach to solving the problem of detecting a signal in noise, including cases with a low signal-to-noise ratio. The sequence of mathematical problems needed to be solved is determined. For the case of proximity of two hypotheses, analytical constructions are used to obtain test statistics, observational statistics, and decision rules based on frequency, time, and time-frequency distributions. In this paper, the possibility of increasing the signal-to-noise ratio is established when distinguishing between two hypotheses for d-signals. The theoretical results are established by the proof of the corresponding lemmas.

Keywords: detection criteria, signal processing, Fourier transform, signal detection in noise

DOI: 10.7868/S1608303226030073

1. INTRODUCTION

At first glance, the problem of detecting a deterministic signal in the presence of environmental noise appears to be a binary classification problem. However, the deeper a researcher dives into formalizing the problem, the more questions arise regarding what information they possess about the properties of both the signal and the noise and about how to correctly interpret the solution to this problem. In the general case, the detection problem is a binary classification problem that requires formulating a decision rule regarding the presence or absence of a useful signal in a signal-noise mixture. In this context, features derived from the input time series may belong to a set of different classes which include both the useful signal and the noise. The more classes available for decision-making, the more precisely one needs to know the frequency-time properties of the signal and noise [1, 2].

In binary classification, the discrimination between two hypotheses relies on the discrimination function [3], which indicates the limit of the methods ability to detect a deterministic signal in noise. This can be viewed as an analogue of the uncertainty principle in quantum mechanics or digital signal processing (the Kotelnikov theorem) [2]: even by increasing the number of features, it is impossible to solve the problem of signal detection in noise at a low signal-to-noise ratio. It follows that the key physical characteristics of a signal-noise mixture that are known or can be measured are the type and length of the time window, the sampling rate, the signal-to-noise ratio, the statistical properties of the noise, and the number of harmonics [4]. The remaining characteristics can be classified as local or integral information characteristics, which are functions of the input time series [5–8]. Therefore, it is necessary to formulate optimal rules for synthesizing an information system that solves the detection problem based on limited prior information [9, 10].

A distinctive feature of such optimal processing is the instability to variations in the statistical properties of noise and the prior uncertainty regarding the signal properties [11, 12]. It turns out that when solving the problem of detecting a deterministic signal under a low signal-to-interference ratio, the discriminatory power of the hypotheses serves as additional information [3, 13]. Overall, practically relevant problems include the detection of deterministic signals under a low signal-to-noise ratio and the discrimination between chaotic and random signals. Mathematical formulations of such problems reasonably rest on on additional assumptions concerning the properties of the signal and noise. In practice, however, many mathematical constructs prove to be either unimplementable or implementable only with significant errors.

The present work is devoted to establishing the properties of the detection problem and developing an approach to improve the quality of the solution. Several successive stages of formalizing and solving the detection problem will be considered, building upon known formulations of the binary hypothesis testing problem and the Neyman–Pearson lemma. Subsequently, it is necessary to identify the criteria, properties, and features of a deterministic signal, to apply optimal signal processing methods with the computation of probabilistic time, frequency, and time-frequency distributions, and, at the final stage, to construct statistics for model experiments.

2. NEUMANN-PEARSON LEMMAS, OPTIMAL TESTS, AND THE FORM OF THE ERROR FUNCTION

2.1. Detection Problem

The problem of detecting the signal $s(n)$ is traditionally reduced to the problem of distinguishing between two hypotheses

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

Hypothesis Γ_0 corresponds to the decision that only noise is received, while hypothesis Γ_1 corresponds to the reception of a mixture of the signal of interest and noise, where the sequences $x(n)$, $n = 1, \ldots, N$ represent a time series of the received data, $s(n)$ is the signal of interest, $w(n)$ is additive random noise, and N is the length of the data time series.

The random variables of the time series $(x(1),\ldots,x(n),\ldots,x(N))$ take values $(x_1,\ldots,x_n,\ldots,x_N)\in\mathbb{R}^N$.

2.2. Classical Neyman–Pearson Lemma

To obtain an analytical expression for estimating the probability of hypothesis discrimination error, one can apply a variant of the Neyman–Pearson lemma, which serves as a condition for optimality [14, 15].

Lemma 1 [Neumann–Pearson (classical)]. *Let there be an arbitrary measurable function of several variables $(x_1, \ldots, x_N) \in \mathbb{R}^N$, called a decision rule or test, such that*

$$\varphi(x_1, \ldots, x_N) = \begin{cases} 1, \ \textit{hypothesis } \Gamma_0 \textit{ is true}, \\ 0, \ \textit{hypothesis } \Gamma_1 \textit{ is true}, \end{cases}$$

by which we find

$$\alpha(\varphi) = \textit{probability } (\textit{of accepting } \Gamma_1 | \Gamma_0 \textit{ is true}),$$
$$\beta(\varphi) = \textit{probability } (\textit{of accepting } \Gamma_0 | \Gamma_1 \textit{ is true}).$$

Then the decision rule φ^* is optimal if

$$1 - \beta(\varphi^*) = \sup_{\varphi}[1 - \beta(\varphi)] \ \textit{if } \alpha(\varphi^*) = \alpha_0, \tag{1}$$

where the supremum is taken over all possible tests.

Here, $\alpha(\cdot)$ is the false alarm probability, and $\beta(\cdot)$ is the probability of missing a useful signal.

2.3. Unconditional Neyman–Pearson Lemma

Lemma 2 [Neumann–Pearson (unconditional)]. *Let there be an arbitrary measurable function of several variables* $(x_1, \ldots, x_N) \in \mathbb{R}^N$*, called a decision rule or test, such that*

$$\varphi(x_1, \ldots, x_N) = \begin{cases} 1, & \textit{hypothesis } \Gamma_0 \textit{ is true}, \\ 0, & \textit{hypothesis } \Gamma_1 \textit{ is true}, \end{cases}$$

from which we find

$$\begin{aligned} \alpha(\varphi) &= \textit{probability } (\textit{of accepting } \Gamma_1 | \Gamma_0 \textit{ is true}), \\ \beta(\varphi) &= \textit{probability } (\textit{of accepting } \Gamma_0 | \Gamma_1 \textit{ is true}). \end{aligned}$$

Then the decision rule φ^* *is optimal if*

$$\alpha(\varphi^*) + \beta(\varphi^*) = \inf_{\varphi} [\alpha(\varphi) + \beta(\varphi)] = \mathcal{E}r(N; \Gamma_0, \Gamma_1) \text{ — } \textit{error function}, \tag{2}$$

where the infimum is taken over all possible tests.

The peculiarity of applying this variant of the Neyman–Pearson lemma lies in the requirement to avoid errors when accepting one hypothesis in place of the other.

The following exact formula holds for the error function:

$$\mathcal{E}r(N; \Gamma_0, \Gamma_1) = 1 - \frac{1}{2} \|P_0^{(N)} - P_1^{(N)}\| = 1 - TV(P_0, P_1), \tag{3}$$

where $P_0^{(N)}$ is the multivariate distribution function of the observation statistics under hypothesis Γ_0, $P_1^{(N)}$ is the multivariate distribution function of the observation statistics under hypothesis Γ_1, and $TV(P_0, P_1)$ is the total variation of the signed measure. Here, $\|Q\| = 2 \sup_A |Q(A)|$. Thus, if the support sets of the measures P_0 and P_1 do not intersect, then the hypotheses can be distinguished without error. If, however, the measures $P_0^{(N)}$ and $P_1^{(N)}$ are close, then $\|P_0^{(N)} - P_1^{(N)}\| \approx 0$, and then $\mathcal{E}r(N; \Gamma_0, \Gamma_1) \approx 1$.

Remark 1. For the problem of detecting a deterministic useful signal, for example, under low signal-to-noise ratio conditions, the case where $\|P_0^{(N)} - P_1^{(N)}\| = 2TV(P_0, P_1) \approx 0$ is of interest, as is the possibility of reasonably estimating this quantity. Therefore, when the probability of a total error in distinguishing between two hypotheses is close to one, it becomes possible to use the analytical expression $TV(P_0, P_1)$ to construct a criterion for the problem of detecting a useful signal in a signal-noise mixture.

Let us emphasize once again that in the classical Neyman–Pearson lemma, the probability of signal detection is maximized for a given false alarm probability, whereas in the unconditional Neyman–Pearson lemma, the total error of false alarms and missed signals is minimized. Due to the analytical representation of the error function, the unconditional lemma is preferable when working with low signal-to-noise ratios and unknown noise conditions. The following problem arises.

Problem 1. Determine the rules for selecting a criterion and constructing distributions P_0, P_1 that minimize the hypothesis discrimination error in a signal detection problem under low signal-to-noise ratio.

Here and in the following, we consider discrete probability distributions $P = \{p = (p_1, \ldots, p_i, \ldots, p_N)\}$, which, by definition, possess the following properties:

$$\forall\ p_i \in [0, 1], \quad \sum_{i=1}^{N} p_i = 1. \tag{4}$$

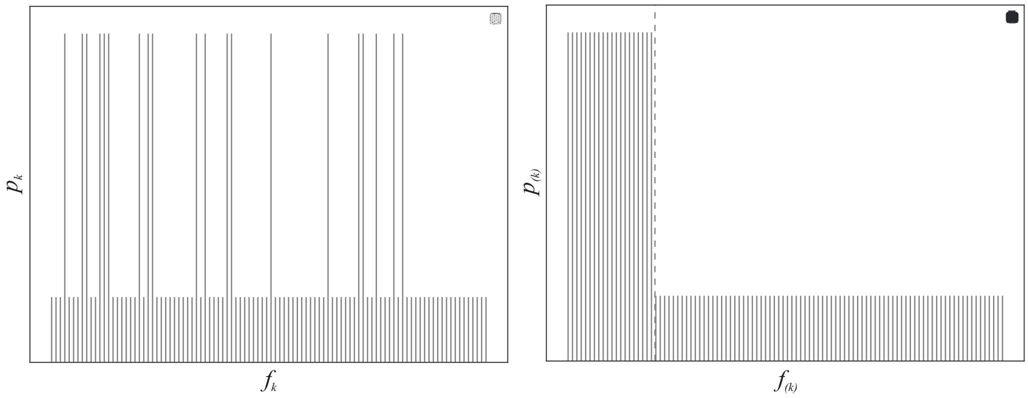

Fig. 1. The energy disordered and ordered spectra of a d-signal in uniform noise ($d = 20, N = 100$).

We now give the definition of a d-signal, which will be used in this work. To this end, we consider a distribution of the following form, where each probability p_k can take one of two distinct values:

$$\begin{cases} p_k = \dfrac{1 - p_{\max}}{N - d}, & \forall\, k = 1, \ldots, N - d, \\ p_k = p_N = \dfrac{p_{\max}}{d}, & \forall\, k = N - d + 1, \ldots, N. \end{cases} \tag{5}$$

Definition 1. A signal $s(t)$ for which the time, spectral, or frequency-time representations can be expressed as a distribution (5) is called a d-signal.

In [16], the following definition of a d-signal is given.

Definition 2. A signal consisting of d sinusoidal signals of equal amplitude will be referred to as a d-signal, and the individual frequencies of this signal will be referred to as harmonics. Then the spectrum of this signal will take the form

$$A_d = \sum_{i=1}^{d} A_0 \exp(-j2\pi f_i \Delta T), \tag{6}$$

where A_0 is the amplitude of the unit spectrum sample, f_i is the frequency of the ith sample (Hz), ΔT is the signal time interval in seconds.

An example of the spectrum and the ordered spectrum of a d-signal is shown in Fig. 1.

Remark 2. The frequencies of the d-harmonics f_i and the window size ΔT, where a window is defined as the time interval of the signal over which the discrete Fourier transform (DFT) is performed, are such that the following condition holds: $\forall f_i : f_i \Delta T = r_i$, $r_i \in \mathbb{N}$, $r_i < N/2$.

The meaning of the expression $r_i \in \mathbb{N}$ is that the integer number of periods r_i of the corresponding ith d-harmonic spectrum must fit within the time window ΔT of the signal-noise mixture.

Remark 3. Depending on the values of $p_{\max}$ and d, different d-signals are obtained.

3. OPTIMAL DISCRIMINATION OF TWO HYPOTHESES

It follows from the classical Neyman–Pearson lemma that the best method for distinguishing hypothesis is based on the likelihood ratio for discrete time of the following form:

$$l_N(x(t)) = \ln \frac{P_0(x(1), \ldots, x(N))}{P_1(x(1), \ldots, x(N))}, \tag{7}$$

where P_0 and P_1 are the joint distributions of the random process ξ_t, which takes values $x(1), \ldots, x(N)$ at times $t = \Delta t, \ldots, t = N\Delta t$. For a given time window width $T = N\Delta t$, there exists a strict functional relationship between α (false alarm probability) and β (probability of missing a useful signal) under optimal reception conditions

$$\alpha = \alpha(\beta, T) \text{ or } \beta = \beta(\alpha, T) \text{ or } T = T(\alpha, \beta). \tag{8}$$

It follows that $T(\alpha, \beta)$ represents the minimum observation time required to distinguish two hypotheses with given false alarm and missed detection rates [3].

Remark 4. Since the number of samples N can be increased in the discrete Fourier transform (DFT), each observation can be considered sufficient for investigating patterns

$$\alpha = \alpha(\beta) \text{ or } \beta = \beta(\alpha). \tag{9}$$

Definition 3. Let us define the mutual entropy of hypothesis Γ_0 relative to Γ_1 as the Kullback–Leibler divergence

$$H_{0/1}(P_0, P_1) = D_{KL}(P_0, P_1). \tag{10}$$

Proposition 1. *If the probability of error β is not too small, specifically $-\ln\beta \ll N$, then*

$$\beta(\alpha) = \exp(-H_{0/1}(P_0, P_1))(1 - \alpha) + o\left(\frac{\ln\beta}{N}\right). \tag{11}$$

Definition 4. We will define the discriminatory ability of hypotheses as the symmetrized Kullback–Leibler distance, also known as the information distance,

$$I_D = D_{KL}(P_0, P_1) + D_{KL}(P_1, P_0). \tag{12}$$

Discriminatory ability is used when the results of experiments under hypotheses Γ_0 and Γ_1 are close to each other, i.e., when the task is to detect a weak signal. In general, discriminatory ability depends on the signal-to-noise ratio. The following examples illustrate this feature.

Example 1. Let there be two hypotheses Γ_0 and Γ_1, and let the false alarm probability α and the missed detection probability β be calculated; then, provided the conditions of Proposition 1 are satisfied, it follows that

$$I_D = -\ln\frac{\beta}{(1-\alpha)} - \ln\frac{\alpha}{(1-\beta)} = \ln\frac{(1-\alpha)(1-\beta)}{\alpha\beta} \approx -(\ln\alpha + \ln\beta). \tag{13}$$

Example 2. Let there be two hypotheses Γ_0 and Γ_1 for distinguishing a deterministic signal from Gaussian noise with spectral density σ^2, and let us represent the deterministic signal as

$$s(t) = s_0 + \sum_{i=1}^{\infty} a_i \cos(\omega_i t + \varphi_i), \tag{14}$$

where $\{\omega_i\}$ and $\{\varphi_i\}$ are constants. Then

$$I_D = \frac{\int_0^T s^2(t)dt}{\sigma^2} = \frac{s_0^2}{2\pi\sigma^2} + \frac{1}{4\pi}\sum_{i=1}^{\infty}\frac{a_i^2}{\sigma^2}. \tag{15}$$

This means that in separating signal and noise, the discriminatory ability depends on the signal-to-noise ratio [3].

Example 3. Let there be two hypotheses Γ_0 and Γ_1 for distinguishing a deterministic d-signal from Gaussian noise with spectral density σ^2, where $d \ll N$, and the spectrum of this signal takes the form

$$A_d = \sum_{i=1}^{d} A_0 \exp(-j\omega_i + \varphi_i), \tag{16}$$

where A_0 is the amplitude of the unit spectral bin, ω_i is the frequency of the ith bin, and φ_i is the phase of the ith bin. Then, using spectral complexity and order statistics to construct the ordered spectral density of the given signal-to-noise ratio [17], the following is obtained

$$I_D \approx \ln\left(\frac{N}{d}\right)\left(\frac{\int_0^T s^2(t)dt}{\sigma^2}\right). \tag{17}$$

This means that the discriminatory ability depends on the number of signal harmonics and the signal-to-noise ratio.

The examples demonstrate the limits of hypothesis differentiation under given approaches to their comparison. Examples 1 and 2 show that the statistical and energy-based approaches yield the same results under otherwise identical conditions. Example 3 illustrates the difference in the ability to distinguish between hypotheses using the energy-based approach, provided that the spectral bandwidth of the signal is known at a given sampling rate. Example 1, adapted to the conditions of Example 3, establishes that under its conditions, the probability of signal rejection decreases for a given false alarm rate.

The functions of information divergences presented above can be unified under the general concept of *f-divergence* [18]:

$$D_f(p||q) = \sum_{x \in \mathbb{R}^N} q(x) f\left(\frac{p(x)}{q(x)}\right). \tag{18}$$

The choice of the function f gives rise to a whole family of different divergences:

- The Kullback–Leibler divergence $D_{KL}(p,q)$ is obtained from (18) by choosing $f(x) = x\log_2(x)$, $x > 0$.
- The Jensen–Shannon divergence is obtained from (18) by choosing

$$f(x) = x\log_2\frac{2x}{x+1} + \log_2\frac{2}{x+1}, \; x \; > \; 0. \tag{19}$$

- The total variation is obtained for $f(x) = \frac{1}{2}|1-x|$:

$$TV(p,q) = \frac{1}{2}\sum_{x \in \mathbb{R}^N} |p(x) - q(x)|. \tag{20}$$

Remark 5. For discrete distributions of the form (4), the expression for the total variation of a measure denoted by TV, written as

$$TV(P_0, P_1) = \frac{1}{2}\sum_{i=1}^{N} |p_i^0 - p_i^1|. \tag{21}$$

Furthermore, $TV(p,q)$ is also a metric on the space of probability distributions and provides an upper bound on the Jensen–Shannon divergence:

$$JSD(p||q) \leq TV(p,q). \tag{22}$$

Each of the functions discussed in this section can be used to construct observation statistics during the secondary processing of a sequence of measurement sets, each containing $i = 1, \ldots, N$ samples.

To formalize criteria that take into account not only the random component of signals but also the deterministic component, let us introduce the concepts of the disequilibrium function D and the statistical complexity C of a distribution

$$C(p) = H(p)D(p,q),$$

where $H(p)$ is the entropy of the distribution p, and the disequilibrium $D(p,q)$ defines the distance between the signal and noise distributions. The simplest example of disequilibrium is the Euclidean distance in the space of discrete probability distributions [19], which is convenient to use when evaluating and comparing signals with a spectral distribution q that is close to uniform.

4. CONSTRUCTING THE DISTRIBUTIONS P_0 AND P_1

Let us describe the approaches to constructing the distributions P_0 and P_1 for a single observation. Three methods for constructing these distributions are proposed: based on the time series $\{x_i\}$, based on its discrete Fourier transform, and based on a frequency-time transform, specifically the classical or robust spectrogram.

4.1. Optimization of Frequency-Time Distributions and Types of Spectrograms

It turns out that the short-time Fourier transform can be obtained as the solution to an optimization problem [20].

Problem 2. It is necessary to find

$$F_s(t,\omega) = m^* = \arg\min_m I(t,\omega,m), \tag{23}$$

where

$$I(t,\omega,m) = \sum_{n=-N/2}^{N/2-1} \mathrm{w}(n\Delta t)\mathbf{F}(e(t,\omega,n,m)). \tag{24}$$

Here, $\mathbf{F}(e)$ is the loss function, $\mathrm{w}(n\Delta t)$ is the window function, Δt is the sampling interval, and the error function takes the form

$$e(t,\omega,n,m) = s(t+n\Delta t)\exp(-i\omega n\Delta t) - m.$$

The error function determines the residual value that expresses the similarity between the signal and the harmonic $\exp(i\omega n\Delta t)$. Different types of spectrograms can be obtained using various functions $\mathbf{F}(e)$.

Remark 6. If $\mathbf{F}(e) = |e|^2$, hen the necessary extremum conditions for the Problem 2

$$\frac{\partial I(t,\omega,m)}{\partial m} = 0$$

yield its solution in the form

$$F_s(t,\omega) = \frac{1}{a_{\mathrm{w}}} \sum_{n=-N/2}^{N/2-1} \mathrm{w}(n\Delta t)s(t+n\Delta t)\exp\left(-i\omega n\Delta t\right), \quad a_{\mathrm{w}} = \sum_{n=-N/2}^{N/2-1} \mathrm{w}(n\Delta t). \tag{25}$$

It follows that the short-time Fourier transform $S_t(\omega)$ of the signal $s(t)$ is equal to

$$S_t(\omega) = F_s(t,\omega),$$

and the corresponding spectrogram is given by

$$P_S(t,\omega) = |F_s(t,\omega)|^2. \tag{26}$$

Remark 7. Since (t,ω) are discrete variables in discrete transforms, we can define

$$P_0 = \frac{P_S(t,\omega)}{\sum_{t,\omega} P_S(t,\omega)}.$$

Remark 8. The maximum likelihood estimation method is applicable when the noise distribution density $p(e)$ is known. MLE assumes a loss function $\mathbf{F}(e) = -\ln p(e)$. Thus, the MLE estimate of signal spectra distorted by Gaussian noise $p(e) \sim \exp(-|e|^2)$, with the corresponding loss function $\mathbf{F}(e) = |e|^2$, is the short-time Fourier transform. Furthermore, in many cases, MLE estimates are highly sensitive to deviations from the parametric model and the assumed distribution. Even a small deviation from the assumption can lead to a significant deterioration of such an estimate.

Following the Remarks 6 and 8, a minimax robust approach [8] was developed in statistics as an alternative to the traditional MLE to reduce the sensitivity of MLE estimate. The loss function in this case takes the form $\mathbf{F}(e) = |e|$. The robust minimax short-time Fourier transform (M-STFT) is based on the solution of the nonlinear system of equations (27)–(29)

$$F_s(t,\omega) = \frac{1}{a_{\mathrm{w}}(t,\omega)} \sum_{n=-N/2}^{N/2-1} d(t,\omega,n)s(t+n\Delta t)\exp\left(-i\omega n\Delta t\right), \tag{27}$$

$$d(t,\omega,n) = \frac{\mathrm{w}(n\Delta t)}{|s(t+n\Delta t)\exp\left(-i\omega n\Delta t\right) - F_s(t,\omega)|}, \tag{28}$$

$$a_{\mathrm{w}}(t,\omega) = \sum_{n=-N/2}^{N/2-1} d(t,\omega,n). \tag{29}$$

Definition 5. Let $\mathbf{F}(e) = |e|$ and suppose that (27)–(29) are given in Problem 2; then the robust spectrogram is defined as

$$P_s(t,\omega) = I(t,\omega,0) - I(t,\omega,F_s(t,\omega)). \tag{30}$$

Note that for a quadratic loss function, the expressions (30) and (25) are identical.

Another way to construct a robust spectrogram is to construct a windowed discrete Wigner–Ville distribution. The distribution is defined as

$$W_s(t,\omega) = \frac{1}{a_{\mathrm{w}}} \sum_{n=-N/2}^{N/2} \mathrm{w}(n\Delta t)s(t+n\Delta t)s^*(t-n\Delta t)\exp\left(-i2\omega n\Delta t\right), \tag{31}$$

where

$$a_{\mathrm{w}} = \sum_{n=-N/2}^{N/2-1} \mathrm{w}(n\Delta t). \tag{32}$$

It can be found by solving the following optimization problem.

Problem 3. It is necessary to find

$$W_s(t,\omega) = m^* = \arg\min_m J(t,\omega,m), \tag{33}$$

where

$$J(t,\omega,m) = \sum_{n=-N/2}^{N/2-1} \mathrm{w}(n\Delta t)\mathbf{F}(|e|). \tag{34}$$

The error function is defined as follows:

$$e(t,\omega,n,m) = s(t+n\Delta t)s^*(t-n\Delta t)\exp\left(-i2\omega n\Delta t\right) - m,$$

and $\mathbf{F}(|e|) = |e|^2$.

4.2. Discrete Fourier Transform

Problem 4. Let $\{x_1,\ldots,x_{2N+2}\}$ be a realization of the sequence of independent random variables $\{\xi_1,\ldots,\xi_{2N+2}\}$ with zero expected value, to which the DFT is applied

$$X_k = \sum_{n=1}^{2N+2} x_n e^{-2\mathbf{i}\pi k(n-1)/(2N+2)}, \tag{35}$$

which defines the random variable

$$\Xi_k = \sum_{n=1}^{2N+2} \xi_n e^{-2\mathbf{i}\pi k(n-1)/(2N+2)}, \tag{36}$$

where $k = 0,\ldots,N$, since, by the symmetry property of the DFT of a real signal, the second half of the $N+1,\ldots,2N+1$ complex amplitudes of the spectral samples is complex conjugate to the first.

It is necessary 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

$$\eta_k(N) = \frac{(\mathbf{T}I)_k}{E_X}, \tag{37}$$

where $I_k = \Xi_k\Xi_k^*$ (the square of the amplitude modulus or the energy of the spectral bin), E_X is half the signal energy, and $\mathbf{T}$ is the operator that orders the series in non-increasing order.

The solution of Problem 4 in the case of observing independent normally distributed random variables $\xi_n,\ n = 1,\ldots,2N+2$ with mean zero and variance σ_0^2 is given in [17] and [13]. The estimates of the probability function $\widetilde{n}_k(N)$ of the normalised ordered discrete spectrum, generated by the corresponding exponential random variable I_k, are as follows:

$$\widetilde{n}_k(N) = \frac{1}{N}\left(\sum_{i=1}^{N}\frac{1}{i} - \sum_{i=1}^{k-1}\frac{1}{i}\right) = \frac{1}{N}\sum_{i=k}^{N}\frac{1}{i}. \tag{38}$$

To obtain an approximate value of $n_k(N)$, let us choose the following formula

$$n_k(N) = -\frac{1}{NK_N}\ln\frac{2k-1}{2N+1}, \quad K_N = -\frac{1}{N}\sum_{k=1}^{N}\ln\frac{2k-1}{2N+1}. \tag{39}$$

Remark 9. Assuming that the normalised ordered statistics $(\eta_1, \dots, \eta_N)$ take the values $P_0 = \{p = (\widetilde{n}_1, \dots, \widetilde{n}_N)\}$ or $P_0 = \{p = (n_1, \dots, n_N)\}$, and the observation statistics $P_1 = \{p = (p_1, \dots, p_N)\}$, the spectral complexity can also be defined on the basis of the distribution (38) using the formula

$$C_{SS}(p) = -\frac{1}{4\log_2 N}\left(\sum_{k=1}^{N} p_k \log_2 p_k\right)\left(\sum_{k=1}^{N} |p_k - \widetilde{n}_k(N)|\right)^2 \tag{40}$$

or its approximate equivalent (39) using the formula

$$C_S(p) = -\frac{1}{4\log_2 N}\left(\sum_{k=1}^{N} p_k \log_2 p_k\right)\left(\sum_{k=1}^{N} |p_k - n_k(N)|\right)^2. \tag{41}$$

Now let us assume that $N = N_1N_2$, where N_1 is the number of samples in a single window and N_2 is the number of windows. Since the entropy in a single window does not depend on the order of the spectral samples, it can be defined as

$$H(n_k(N_1)) = -\frac{1}{\log_2 N_1}\sum_{k=1}^{N_1} n_k(N_1)\log_2 n_k(N_1).$$

The entropy in N_2 statistically identical windows is equal to

$$H(P_S(N_2, N_1)) = -\frac{1}{\log_2 N_1 + \log_2 N_2}\left(\log_2 N_2 + \sum_{k=1}^{N_1} p_k(N_1)\log_2 p_k(N_1)\right).$$

5. OPTIMISATION OF INFORMATION CRITERIA AND d-SIGNAL DETECTION

Now let us turn to some well-known information criteria for discriminating between hypotheses and establish their properties. Since Lemma 2 establishes that the discrimination error function for two hypotheses depends on the total variance $TV(p, q)$, one more concept of disequilibrium and statistical complexity based on it is introduced [10]. It is also shown there that such a criterion is one of the best for solving the problem of distinguishing between two hypotheses and indicating the presence of a deterministic component of a weak useful signal in white noise:

$$C_{TV}(p) = H(p)D_{TV}(p), \tag{42}$$

where

$$D_{TV}(p) = \frac{1}{4}\left(\sum_{i=1}^{N}\left|p_i - \frac{1}{N}\right|\right)^2, \tag{43}$$

$$H(p) = \frac{1}{\log_2 N}\left(-\sum_{i=1}^{N} p_i \log_2 p_i\right) \tag{44}$$

— normalised Shannon entropy [21]. When calculating the sum in (44), it is assumed that $\frac{0}{\log_2 0} = 0$ by continuity. It was also shown there that the maximum value of C_{TV} is attained on a distribution consisting of only two values, with an optimal number of samples for each of these values. The following holds

Lemma 3. *The maximum statistical complexity* (42) *is attained on the class of distributions* (5). It is necessary to determine the optimal values for d and $p_{\max}$. To do this, let us compute the value of the disequilibrium D_{TV} on the distribution (5):

$$D_{TV}(\omega, p_{\max}) = (p_{\max} - \omega)^2, \quad \omega = \frac{d}{N}. \tag{45}$$

In turn, the entropy is given by

$$H(\omega, p_{\max}) = 1 - \frac{1}{\log_2 N}\left((1 - p_{\max})\log_2 \frac{1 - p_{\max}}{1 - \omega} + p_{\max}\log_2 \frac{p_{\max}}{\omega}\right). \tag{46}$$

The table shows the optimal values of the parameters for the formulas (45) and (46), which yield the maximum statistical complexity $C_{TV}(\omega, p_{\max})$.

Optimal parameters $C_{TV}(\omega, p_{\max})$ for various values of N

N	$C_{TV}(\omega^*, p^*_{\max})$	$p^*_{\max}$	$1 - \omega^*$	d^*
3	0.1289	0.8241	0.6751	1 or 2
256	0.4789	0.9976	0.8752	32
512	0.5120	0.9991	0.8901	56
1024	0.5410	0.9997	0.9022	100
2048	0.5667	0.9999	0.9122	180

From Lemma 3 and the table, it is clear that d-signals are important for solving the detection problem. This leads to the following

Problem 5. Detection of a d-signal in a signal-noise mixture of length N, $N > d$.

To solve Problem 5 on d-signal detection, we apply the Renyi entropy H_γ with parameter γ ($\gamma > 0$, $\gamma \neq 1$), the detailed properties of which are presented in [22].

$$H_\gamma = \frac{1}{1 - \gamma}\left(\log_2 \sum_{k=1}^{N} p_k^\gamma\right). \tag{47}$$

It turns out that the Renyi entropy for a d-signal is calculated as

$$H_\gamma^{(d)} = \frac{1}{1 - \gamma}\log_2\left(d\left(\frac{p_{\max}}{d}\right)^\gamma + (N - d)\left(\frac{1 - p_{\max}}{N - d}\right)^\gamma\right). \tag{48}$$

For large N, it can be estimated as follows:

$$H_\gamma^{(d)} \approx \widetilde{H}_\gamma^{(d)} = \begin{cases} \dfrac{1}{1 - \gamma}\log_2(d^{1-\gamma}p_{\max}^\gamma) = \log_2 d + \dfrac{\gamma}{1 - \gamma}\log_2(p_{\max}), & \gamma > 1, \\ \log_2 N + \dfrac{\gamma}{1 - \gamma}\log_2(1 - p_{\max}), & 0 < \gamma < 1, \end{cases} \tag{49}$$

which shows an approximate linear relationship between the logarithms of of the number of harmonics in the signal, $p_{\max}$ and the Renyi entropy.

From the table and (49), for $\gamma > 1$ the following equality holds:

$$\widetilde{H}_\gamma^{(d^*)} = \log_2 d^*. \tag{50}$$

Similar relations are known for the frequency-time processing of known signals [20]. The Renyi entropy on white noise takes its maximum value $\widetilde{H}_\gamma^{(N)} = \log_2 N$.

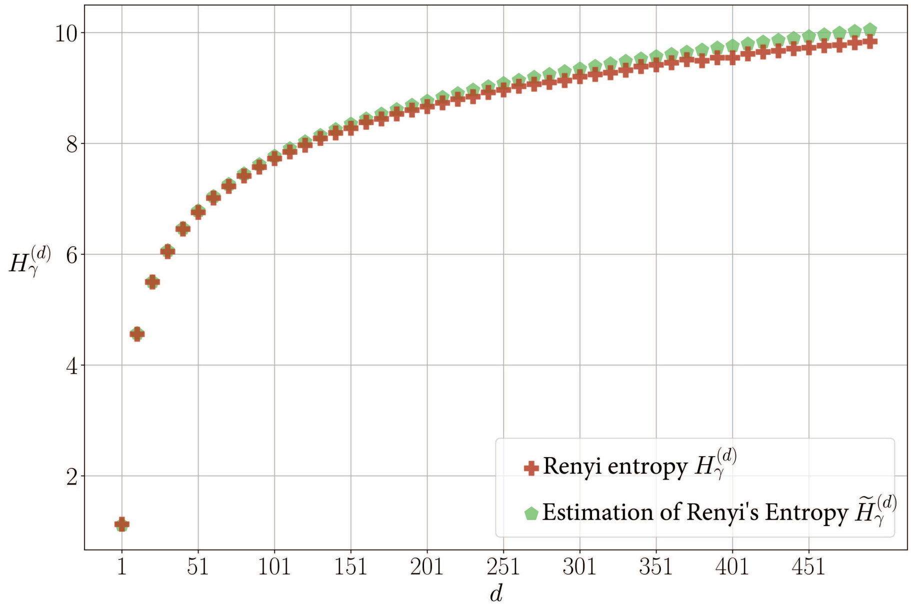

Fig. 2. A comparison of the Renyi entropy ($\gamma = 10$), calculated using formula (47), and its estimate (53).

Definition 6. Let us define the signal-to-noise ratio for a d-signal in the case where the system (5) corresponds to the power spectrum of a signal-noise mixture as:

$$SNR = 10\log_{10}\delta = 10\log_{10}\left(d\left(\frac{p_{\max}}{d}\right)\bigg/(N-d)\frac{(1-p_{\max})}{(N-d)}\right) = 10\log_{10}\left(\frac{p_{\max}}{1-p_{\max}}\right). \quad (51)$$

In this case, $p_{\max}$ can be expressed in terms of SNR in order to be used in the calculation of Renyi entropy

$$p_{\max} = \frac{10^{\frac{SNR}{10}}}{1+10^{\frac{SNR}{10}}}. \quad (52)$$

Then $\widetilde{H}_\gamma^{(d)}$ can be written for $\gamma > 1$ according to (49) in the following form:

$$\widetilde{H}_\gamma^{(d)} = \log_2 d + \frac{\gamma}{1-\gamma}\log_2\left(\frac{10^{\frac{SNR}{10}}}{1+10^{\frac{SNR}{10}}}\right) = \log_2 d + \frac{\gamma}{1-\gamma}\left[\frac{SNR}{10}\log_2 10 - \log_2\left(1+10^{\frac{SNR}{10}}\right)\right]. \quad (53)$$

Figure 2 shows a comparison of the Renyi entropy, calculated using formula (47), and its estimate (53) as a function of the number d of harmonics in the generated signal, mixed with white noise, at $SNR = 0\ dB$. It can be seen that the estimate is close to the calculated values.

Corollary 1. *With a low signal-to-noise ratio, the Renyi entropy takes the form*

$$\widetilde{H}_\gamma^{(d)} \approx \log_2 d + \frac{\gamma\log_2 10}{10(1-\gamma)}SNR. \quad (54)$$

Thus, when the maximum Renyi entropy is known, the minimum SNR can be estimated as

$$SNR_{\min} = \frac{10(1-\gamma)}{\gamma\log_2 10}\left(\log_2 N - \log_2 d\right). \quad (55)$$

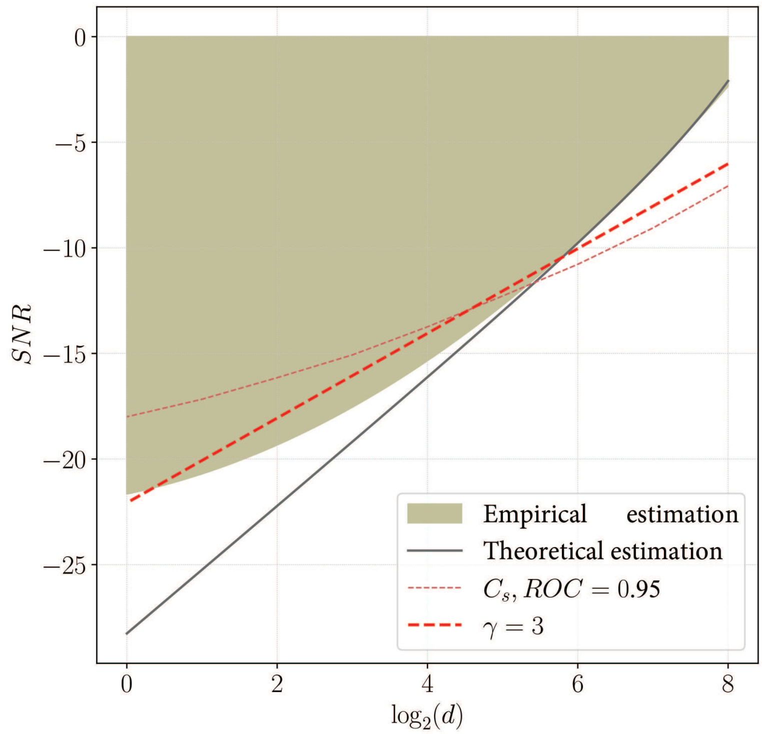

Fig. 3. Limit values of SNR for signal detection in white noise as a function of the number d ($N = 2048$).

According to [4], the one-to-one correspondence between the set H (entropy), C_{TV} (statistical complexity) and the set SNR (signal-to-noise ratio), $\log d$ (discreteness) makes it possible to use metric (analytical)[1] grids in signal detection and classification tasks in noise.

A comparative analysis of SNR estimates as a function of the number d of harmonics in the spectrum is shown in Fig. 3 ($N = 2048$). The shaded area illustrates the results of an empirical study of the limits of signal classification in white noise using a statistical complexity diagram. The solid black line represents the theoretical estimate obtained using the formula from [16]:

$$SNR > 10 \lg \left(\frac{3d}{N - 3d} \right). \tag{56}$$

The thin dotted line in Fig. 3 shows the limiting capabilities of signal detection using the C_s criterion (spectral complexity). The result of the dependence of the limiting SNR on the number of harmonics in the spectrum was obtained through a numerical experiment (ROC $= 0.95$)[2]. The thick dotted line represents the estimate according to formula (55) for $\gamma = 3$.

The formula (53) obtained for $\widetilde{H}_{\gamma}^{(d)}$ allows us to formulate the following lemma for the estimate $\hat{d}$ of the number of components of a d-signal.

[1] The concept of a metric grid is explained in detail in [4]. In the context of the complexity diagram, a metric grid is an approximate one-to-one mapping between a set of information characteristics of a signal-noise mixture, such as information entropy and statistical (spectral) complexity, and a set of characteristics such as discreteness (the number of harmonics in the d-signal spectrum) and the signal-to-noise ratio.

[2] The area under the ROC (Receiver Operating Characteristic) curve is the AUC. In practice, for machine learning model performance metrics, an ROC-AUC value of 0.5 indicates an inability to detect a signal in the noise based on a given criterion; a value between 0.7 and 0.8 is considered acceptable; between 0.8 and 0.9 is considered high; and a value above 0.9 is considered excellent.

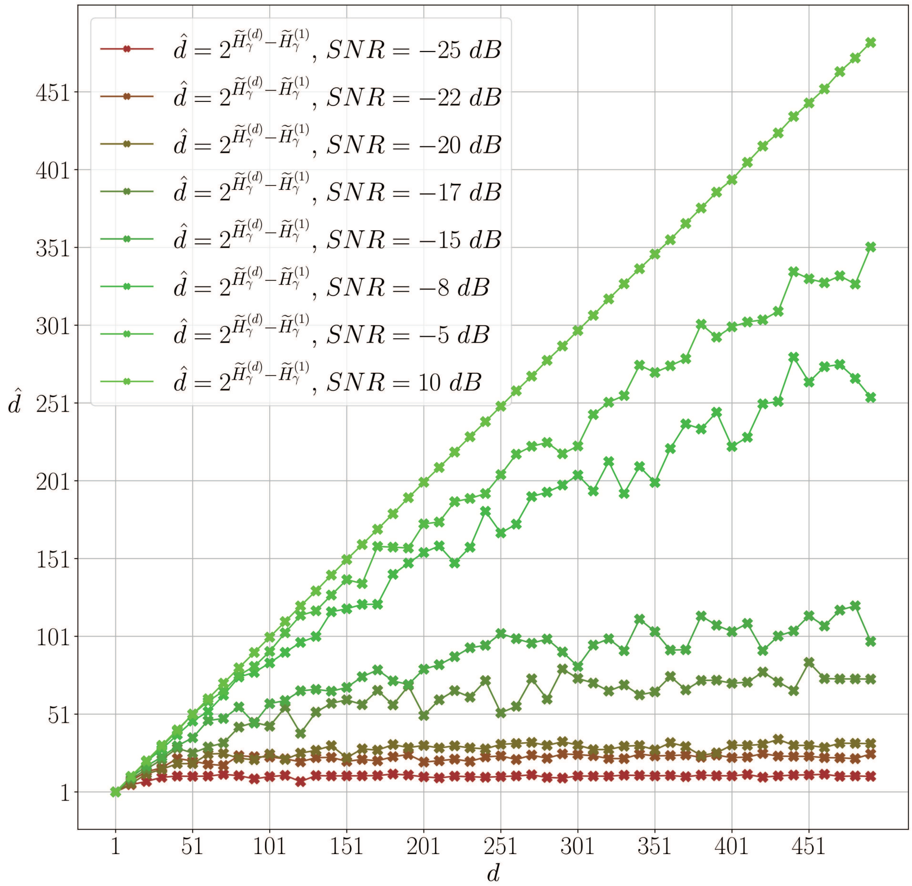

Fig. 4. Estimation of the number of components in a d-signal of length $N = 65536$ in white noise using Renyi entropy ($\gamma = 10$) for different SNR values.

Remark 10. The estimate $\hat{d}$ for the number of d-signal harmonics, using Renyi entropy, is written as follows:

$$\hat{d} = 2^{\widetilde{H}_\gamma^{(\hat{d})} - \widetilde{H}_\gamma^{(1)}}. \tag{57}$$

Proof. Let us write out equation (53) for $d = \hat{d}$ and $d = 1$:

$$\begin{cases} \log_2 \hat{d} = \widetilde{H}_\gamma^{(\hat{d})} - \dfrac{\gamma}{1-\gamma}\left[\dfrac{SNR}{10}\log_2 10 - \log_2\left(1 + 10^{\frac{SNR}{10}}\right)\right], \\ \log_2 1 = 0 = \widetilde{H}_\gamma^{(1)} - \dfrac{\gamma}{1-\gamma}\left[\dfrac{SNR}{10}\log_2 10 - \log_2\left(1 + 10^{\frac{SNR}{10}}\right)\right]. \end{cases} \tag{58}$$

By expressing in this way the term dependent on SNR, which is identical in both cases, in the system (58) via $\widetilde{H}_\gamma^{(1)}$, we obtain the statement of the lemma.

Figure 4 presents a numerical validation of the remark 10 in the form of a dependence of the component number estimate, calculated using formula (57), for generated signals consisting of d harmonic components mixed with white noise at various SNR values. It can be seen from Fig. 4 that the estimate is reasonably stable for up to 100 components, provided that the white noise level is up to $-10\ dB$.

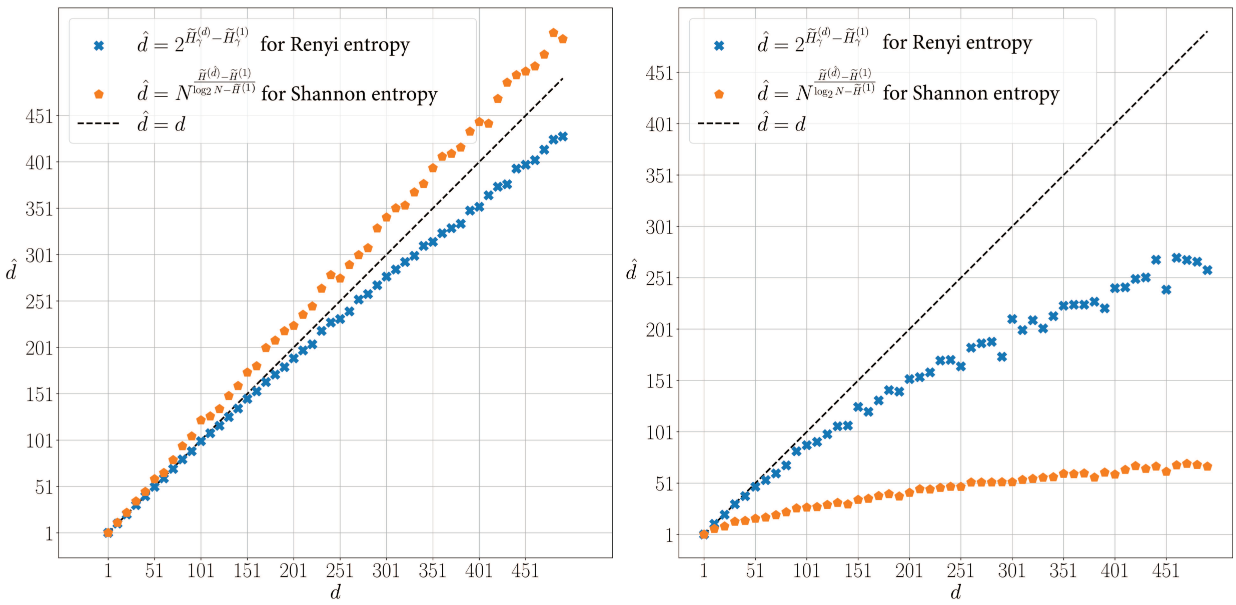

Fig. 5. Estimation of the number of components in a d-signal of length $N = 65536$ in white noise using Renyi entropy ($\gamma = 10$) and Shannon entropy for $SNR = 0\ dB$ and $SNR = -8\ dB$.

Further, let us estimate the number of components using unnormalised Shannon entropy. To do this, let us consider the expression $\widetilde{H}^{(d)}$, which estimates the Shannon entropy for a d-signal when N is large.

$$\begin{aligned} H^{(d)} &= -\left(d\frac{p_{\max}}{d}\log_2\frac{p_{\max}}{d} + (N-d)\frac{1-p_{\max}}{N-d}\log_2\frac{1-p_{\max}}{N-d}\right) \\ &= -\left(p_{\max}\log_2\frac{p_{\max}}{d} + (1-p_{\max})\log_2\frac{1-p_{\max}}{N-d}\right) \end{aligned} \tag{59}$$
$$= -\left(p_{\max}\log_2 p_{\max} + (1-p_{\max})\log_2(1-p_{\max})\right) + \left(p_{\max}\log_2 d + (1-p_{\max})\log_2(N-d)\right).$$

The terms in the left-hand side are constrained by the constraint $0 < p_{\max} < 1$, so they will not appear in the expression for $\widetilde{H}^{(d)}$.

$$H^{(d)} \approx \widetilde{H}^{(d)} = p_{\max}\log_2 d + (1-p_{\max})\log_2 N. \tag{60}$$

To obtain the estimate $\hat{d}$, by analogy with the proof of remark 10, let us consider Shannon entropy for $d = \hat{d}$ and $d = 1$.

$$\widetilde{H}^{(1)} = (1-p_{\max})\log_2 N \Longrightarrow p_{\max} = 1 - \frac{\widetilde{H}^{(1)}}{\log_2 N}. \tag{61}$$

Let us use the expression $p_{\max}$ obtained for $\widetilde{H}^{(\hat{d})}$.

$$\widetilde{H}^{(\hat{d})} = \left(1 - \frac{\widetilde{H}^{(1)}}{\log_2 N}\right)\log_2 d + \log_2 N\frac{\widetilde{H}^{(1)}}{\log_2 N}. \tag{62}$$

Hence

$$\log_2\hat{d} = \frac{(\widetilde{H}^{(\hat{d})} - \widetilde{H}^{(1)})\log_2 N}{(\log_2 N - \widetilde{H}^{(1)})} \Longrightarrow \hat{d} = N^{\frac{\widetilde{H}^{(\hat{d})} - \widetilde{H}^{(1)}}{\log_2 N - \widetilde{H}^{(1)}}}. \tag{63}$$

Consequently, the following remark is valid.

Remark 11. The estimate $\hat{d}$ of the component number of a d-signal, using Shannon entropy, is expressed as follows:

$$\hat{d} = N^{\dfrac{\widetilde{H}^{(\hat{d})} - \widetilde{H}^{(1)}}{\log_2 N - \widetilde{H}^{(1)}}}. \tag{64}$$

Figure 5 shows a comparison of the performance of component number estimates based on the formulas (57) and (64), which are based on the Renyi and Shannon entropies respectively. It can be seen that the Renyi entropy provides a more robust estimate.

6. CONCLUSION

This paper establishes a sequence of optimisation sub-problems that need to be solved in order to detect a signal in noise, including cases with a low signal-to-noise ratio. In the general case, it is necessary to determine the optimal statistical test for the problem of distinguishing between two hypotheses, to establish optimal rules for assigning test statistic observations based on discrete spectral or frequency-time distributions, and to specify the form of the optimal discrimination criterion. Analytical constructions of criteria, the concept of d-signals, and information-based detection criteria form the basis for solving this binary classification problem: whether or not a signal is present. Further improvement in the quality of the solution to this problem is possible by refining the statistical properties of the noise and the signal in the noise.

FUNDING

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

REFERENCES

1. Rihaczek, A., Signal energy distribution in time and frequency, *IEEE Transactions on Information Theory*, 1968, vol. 14, no. 3, pp. 369–374.
2. Cohen, L., *Time-Frequency Analysis*, New Jersey: Prentice Hall, 1995.
3. Dobrushin, R., Pinsker, M., and Shiryaev, A., An application of the concept of entropy to problems of detecting a signal in the noise, *Lithuanian Mathematical Journal*, 1963, vol. 3, no. 1, pp. 107–122.
4. Babikov, V. and Galyaev, A., Analytical representation of complexity diagrams, *Problems of Information Transmission*, 2025, vol. 61, no. 1, pp. 31–46.
5. Horie, T. et al., Sample entropy in electrocardiogram during atrial fibrillation, *Yonago Acta Med.*, 2018, vol. 61, no. 1, pp. 49–57.
6. Ribeiro, M. et al., The entropy universe, *Entropy*, 2021, vol. 23, no. 2, pp. 1–35.
7. Zunino, L., Soriano, M., and Rosso, O., Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 2012, vol. 86, no. 4, pp. 1–10.
8. Shen, J., Hung, J., and Lee, L., Robust entropy-based endpoint detection for speech recognition in noisy environments, in *5th International Conference on Spoken Language Processing (ICSLP 1998)*, ISCA: ISCA, 1998, pp. 1–4.
9. Allen, R. and Mills, D., *Signal Analysis: Time, Frequency, Scale, and Structure*, New Jersey: John Wiley & Sons, 2004.
10. Berlin, L., Galyaev, A., and Lysenko, P., Comparison of information criteria for detection of useful signals in noisy environments, *Sensors*, 2023, vol. 23, no. 4, pp. 1–17.

11. Ramirez, J. et al., A new Kullback–Leibler VAD for speech recognition in noise, *IEEE Signal Process. Lett.*, 2004, vol. 11, no. 2, pp. 266–269.

12. Mehrotra, K., Mohan, C., and Huang, H., *Anomaly detection principles and algorithms*, Terrorism, Security, and Computation, Cham, Switzerland: Springer International Publishing, 2018.

13. Galyaev, A. et al., Order statistics of the normalized spectral distribution for detecting weak signals in white noise, *Autom. Remote Control.*, 2024, vol. 12, no. 1, pp. 1–16.

14. Shiryaev, A., *Veroyatnostno-statisticheskie metody v teorii prinyatiya reshenij*, (Probabilistic and Statistical Methods in Decision-Making Theory), M.: MCNMO: NMU, 2020.

15. Johnson, P., Moriarty, J., and Peskir, G., Detecting changes in real-time data: a user's guide to optimal detection, *Philos. Trans. A Math. Phys. Eng. Sci.*, 2017, vol. 375, no. 2100, pp. 1–16.

16. Babikov, V. and Galyaev, A., Information diagrams and their capabilities for classifying weak signals, *Problems of Information Transmission*, 2024, vol. 60, no. 2, pp. 127–140.

17. Galyaev, A. et al., A new spectral measure of complexity and its capabilities for detecting signals in noise, *Dokl. Math.*, 2024, vol. 110, no. 1, pp. 361–368.

18. Sason, I., On f-divergences: Integral representations, local behavior, and inequalities, *Entropy*, 2018, vol. 20, no. 5, pp. 1–32.

19. López-Ruiz, R., Shannon information, LMC complexity and rényi entropies: a straightforward approach, *Biophys. Chem.*, 2005, vol. 115, no. 2–3, pp. 215–218.

20. Boashash, B., *Time-Frequency Signal Analysis and Processing. A Comprehensive Reference*, London: Academic Press, 2016.

21. Shannon, C., A mathematical theory of communication, *Bell Syst. Tech. J.*, 1948, vol. 27, pp. 379–423.

22. Rényi, A., On measures of entropy and information, in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, vol. 4, University of California Press, 1961, pp. 547–562.

This paper was recommended for publication by A.V. Nazin, a member of the Editorial Board